Am. J. Hum. Genet. 67:1515–1525, 2000

Family-Based Tests of Association in the Presence of Linkage

Stephen L. Lake,[1] Deborah Blacker,[2,3] and Nan M. Laird[1]

Departments of [1]Biostatistics and [2]Epidemiology, Harvard School of Public Health, and [3]Department of Psychiatry, Massachusetts General Hospital and Harvard Medical School, Harvard University, Boston

Linkage analysis may not provide the necessary resolution for identification of the genes underlying phenotypic variation. This is especially true for gene-mapping studies that focus on complex diseases that do not exhibit Mendelian inheritance patterns. One positional genomic strategy involves application of association methodology to areas of identified linkage. Detection of association in the presence of linkage localizes the gene(s) of interest to more-refined regions in the genome than is possible through linkage analysis alone. This strategy introduces a statistical complexity when family-based association tests are used: the marker genotypes among siblings are correlated in linked regions. Ignoring this correlation will compromise the size of the statistical hypothesis test, thus clouding the interpretation of test results. We present a method for computing the expectation of a wide range of association test statistics under the null hypothesis that there is linkage but no association. To standardize the test statistic, an empirical variance-covariance estimator that is robust to the sibling marker-genotype correlation is used. This method is widely applicable: any type of phenotypic measure or family configuration can be used. For example, we analyze a deletion in the *A2M* gene at the 5′ splice site of "exon II" of the bait region in Alzheimer disease (AD) discordant sibships. Since the *A2M* gene lies in a chromosomal region (chromosome 12p) that consistently has been linked to AD, association tests should be conducted under the null hypothesis that there is linkage but no association.

Introduction

Although linkage analysis has been applied successfully to the mapping of genes involved in the pathogenesis of diseases exhibiting Mendelian inheritance, its application in the setting of genetically complex diseases has been less fruitful (Risch and Merikangas 1996). With complex diseases, the resolution from linkage analysis is reduced, and extended segments of the genome containing large numbers of genes may be implicated in disease etiology (Hauser and Boehnke 1997; Roberts et al. 1999). Fine mapping of these linked regions may be accomplished through the use of allelic-association methods that are designed to jointly detect linkage and gametic-phase disequilibrium. Detecting association significantly refines the search for disease susceptibility genes, because linkage disequilibrium between a genetic marker and disease susceptibility polymorphisms is expected to exist only over relatively small genetic distances in most populations. The sequential approach of linkage-based genomic screening followed by dissection of linked regions with association methodology recently has been used to identify a susceptibility locus for human hypertension (Bray et al. 2000).

Allelic association can be detected through traditional contingency-table analysis using cases and controls (Woolf 1955). Although straightforward to implement, tests based on this approach are sensitive to spurious association caused by population admixture (Ott 1989). Family-based association tests (FBATs) are a class of tests that utilize within- and between-family marker-inheritance patterns to test for association and that are safeguarded, by design, from confounding caused by admixture (Ewens and Spielman 1995). A widely used FBAT is the transmission/disequilibrium test (TDT; Terwilliger and Ott 1992; Spielman et al. 1993), which uses the marker genotypes of an affected child and those of his/her parents to test for association. FBATs have received much attention lately, with numerous extensions and generalizations of the TDT being proposed in the literature. Recently, Rabinowitz and Laird (2000) developed a unified approach to family-based association tests that puts tests of different genetic models, tests of different sampling designs, tests involving different disease phenotypes, tests with missing parents, and tests of different null hypotheses, all in the same framework. Algorithms for calculating the distribution of association test statistics for these many settings are also presented.

A distinction must be made between tests for linkage

Received July 28, 2000; accepted for publication September 21, 2000; electronically published October 31, 2000.

Address for correspondence and reprints: Mr. Stephen Lake, Department of Biostatistics, Harvard School of Public Health, 655 Huntington Avenue, Boston, MA 02115. E-mail: slake@hsph.harvard.edu

0002-9297/2000/6706-0017$02.00

that use association methods and tests for association in the presence of linkage. Letting θ be the recombination parameter and δ be a measure of allelic association, the tests for linkage that use association methods have a composite null hypothesis (type I H_0) that can be expressed as H_0:$\delta = 0$ or $\theta = 1/2$. The null hypothesis for testing association in the presence of linkage (type II H_0) is H_0:$\delta = 0$ and $\theta < 1/2$. Both settings have the same alternative hypothesis, H_a:$\delta > 0$ and $\theta < 1/2$. Complications arise in tests addressing the type II H_0 setting, because sibling marker genotypes are correlated under H_0 (Martin et al. 1997; Lazzeroni and Lange 1998). Ignoring the correlation in the type II H_0 setting compromises the α level of the tests. In this article, we show that valid tests for association in the presence of linkage may be performed using the mean of the test statistic computed via the Rabinowitz-Laird (RL) algorithm for the type I H_0 setting and an empirical variance-covariance estimator that adjusts for the correlation among sibling marker genotypes. This provides a convenient means for testing allelic association in the presence of linkage that can be used with a wide range of test statistics and any pedigree configuration. For example, the nine strategies for testing the type I H_0 advocated by S. Horvath, X. Xu, and N. Laird (unpublished data), which include applications to binary, quantitative and time-to-onset phenotypes, can all be adapted to the type II H_0 setting with the method presented here. We note that in the biallelic setting and with a qualitative trait, the pedigree disequilibrium test (PDT; Martin et al. 2000*c*) is similar to the approach developed here.

As an illustration, we focus on the reported association between alleles of the *A2M* gene and late-onset Alzheimer disease. Blacker et al. (1998) reported a strong association between a deletion near the 5′ splice site of exon 18 of the *A2M* gene (*A2M*-18i) and AD in a sample of sibships from the National Institute of Mental Health (NIMH) Genetics Initiative (Blacker et al. 1997). During the course of the *A2M* association study, linkage to a nearby region on chromosome 12 was reported as part of a genome screen (Pericak-Vance et al. 1997). Subsequent linkage analyses revealed linkage peaks at or near the *A2M* gene (Rimmler et al. 1997; Rogaeva et al. 1998; Wu et al. 1998; Kehoe et al. 1999; Scott et al. 1999). The reported *A2M* association has been controversial, with further findings both confirmatory and nonconfirmatory (Dow et al. 1999; Rogaeva et al. 1999; Rudrasingham et al. 1999; Romas et al. 2000). In any case, *A2M* is useful as an illustration of association tests conducted in the presence of linkage. We use the NIMH data set, in which a strong *A2M*/AD association has been reported (Blacker et al. 1999), to illustrate our method.

FBATs

We assume that there are N nuclear families, with n_i children in each family. Let m_{ij} be the marker genotype for the jth child in the ith family and $\boldsymbol{m}_i$ be the vector of marker genotypes for the n_i children in the ith family. In addition, the vector of parental marker genotypes will be denoted by $\boldsymbol{M}_i$. Let $\boldsymbol{X}(m_{ij})$ be an $h \times 1$ vector that codes for marker genotype. Depending on the coding scheme, $\boldsymbol{X}(m_{ij})$ may be a scalar or a vector (see Schaid 1996; Laird et al. 2000; S. Horvath, X. Xu, and N. Laird [unpublished data]). Last, let y_{ij} be the phenotype of the jth child in the ith family and $T(y_{ij})$ be some function of the phenotype. In what follows we will often abbreviate $\boldsymbol{X}(m_{ij})$ with $\boldsymbol{X}_{ij}$ and $T(y_{ij})$ with T_{ij} and drop the subscript indicating family when dealing with data from only one family.

Association test statistics are constructed to detect correlation between genotype and phenotype. In this article, we restrict attention to the class of test statistics that can be expressed as

$$\boldsymbol{S} = \sum_i \boldsymbol{S}_i = \sum_i \sum_j T_{ij}\boldsymbol{X}_{ij} , \qquad (1)$$

where the summation is over all children in all families and $\boldsymbol{S}_i$ is the contribution from the ith nuclear family, $i = 1, \ldots, N$. Test statistics in this general class constitute the majority of family-based association test statistics proposed in the literature, including tests in the multiallelic setting, tests using quantitative phenotypes, and tests that allow missing parental marker information (Laird et al. 2000; Rabinowitz and Laird 2000). For example, with simplex families, letting T_{ij} be an indicator function for child disease status and $\boldsymbol{X}_{ij}$ be the count of a particular marker allele, $\boldsymbol{S}_i$ counts the total number of alleles in the affected child and $\boldsymbol{S}$ is the same test statistic used in the TDT. Other types of test statistics are discussed in S. Horvath, X. Xu, and N. Laird (unpublished data).

Under the assumption that the N families are unrelated, the distribution of the test statistic $\boldsymbol{S}$ under H_0 depends on the distributions of the independent $\boldsymbol{S}_i$, $i = 1, \ldots, N$. For the ith family, the general distribution of $\boldsymbol{S}_i$ depends on the joint distribution of the observed children's marker genotypes, children's phenotypes, and parental marker genotypes $p(\boldsymbol{m}_i, \boldsymbol{M}_i, \boldsymbol{y}_i)$. Under the type I H_0, $p(\boldsymbol{m}_i, \boldsymbol{M}_i, \boldsymbol{y}_i)$ depends on allele frequencies and the genetic model; conditioning on the phenotypes and the parental genotypes eliminates these unknown nuisance parameters and makes the distribution of $\boldsymbol{S}_i$ dependent only on the conditional distribution of the children's marker genotypes (Lazzeroni and Lange 1998). When parental genotypes are unknown, the nuisance parameters can be elim-

inated by conditioning on the sufficient statistic for the parental genotypes $S(\boldsymbol{M})$, which is composed of the observed parental genotypes (when available) $\boldsymbol{M}_{\text{obs}}$ and the children's genotype configuration $\boldsymbol{C}_m$ (Rabinowitz and Laird 2000). The distribution under the type II H_0 is discussed in the next section.

Using the conditional distribution of the children's marker genotypes, we take the approach of standardizing $\boldsymbol{S}$ and using the large sample normal or χ^2 approximation. In this case, the mean and variance of the $\boldsymbol{S}_i$ are required. For the type I H_0, letting $\boldsymbol{\Phi}_I = [S(\boldsymbol{M}),\boldsymbol{y}]$, S. Horvath, X. Xu, and N. Laird (unpublished data) show that $E(\boldsymbol{S}_i|\boldsymbol{\Phi}_I)$ can be computed with the univariate conditional distribution of the children's marker genotype, and $\text{Var}(\boldsymbol{S}_i|\boldsymbol{\Phi}_I)$ can be computed with the univariate and bivariate conditional distributions of the children's marker genotypes, where $\text{Var}(\cdot)$ refers to the variance-covariance matrix. That is, by using just the joint distributions of (m_{ij},m_{ik}) (which, under the type I H_0, do not depend on j and k), we can compute $\text{Var}(\boldsymbol{S}_i|\boldsymbol{\Phi}_I)$. These distributions can be computed using the RL algorithm for the type I H_0.

Tests of Association in the Presence of Linkage

As discussed above, association tests performed in areas of known linkage may significantly refine gene-mapping studies. The challenge is that, among siblings, genetic markers that reside within linked regions are correlated even in the absence of association and after conditioning on $\boldsymbol{\Phi}_I = [\boldsymbol{y},S(\boldsymbol{M})]$. The dependence exists because siblings with similar phenotypes are more likely to share the putative disease genes, even in the absence of allelic association. Linkage between a marker and the putative disease gene, therefore, induces positive correlation between the genetic markers of siblings with similar phenotypes. The opposite holds for siblings with disparate phenotypes. The correlation makes $p(\boldsymbol{m}|\boldsymbol{\Phi}_I)$ dependent on the recombination parameter and the genetic model for the phenotype.

Conditioning on the minimal sufficient statistic for θ and the phenotypes removes the dependence of the marker genotypes on θ and $\boldsymbol{y}$ under the type II H_0. When the patterns of allele sharing among siblings can be unambiguously determined, they serve as the minimal sufficient statistic for θ (Rabinowitz and Laird 2000). With incomplete identification of the allele sharing patterns, the outcome space of the children's marker genotypes given the minimal sufficient statistic under the type II H_0 may be computed using the RL algorithm (type II H_0 case). Therefore, under the type II H_0, the minimal sufficient statistic $\boldsymbol{\Phi}_{II}$ consists of the minimal sufficient statistic for the recombination parameter $S(\theta)$, the minimal sufficient statistic for the parental marker genotypes $S(\boldsymbol{M})$, and the observed phenotypes $\boldsymbol{y}$.

Since patterns of allele sharing are defined by the joint realization of sibling marker genotypes, the conditional outcome space consists of the various joint outcomes of sibling marker genotypes satisfying the constraints of the minimal sufficient statistic for the type II H_0 (Martin et al. 1997; Rabinowitz and Laird 2000). Therefore, after conditioning on $\boldsymbol{\Phi}_{II}$, the convenient expression of $E(\boldsymbol{S}_i|\boldsymbol{\Phi}_{II})$ and $\text{Var}(\boldsymbol{S}_i|\boldsymbol{\Phi}_{II})$, in terms of the univariate and bivariate conditional distribution of marker genotypes under the type I H_0, cannot be paralleled. Rather, under the type II H_0, expressions for $E(\boldsymbol{S}_i|\boldsymbol{\Phi}_{II})$ and $\text{Var}(\boldsymbol{S}_i|\boldsymbol{\Phi}_{II})$ using the RL algorithm can be found with the multinomial distribution.

For a given family, assume that there are p compatible realizations of the sibling marker genotypes, and let $\boldsymbol{r}$ be a $p \times 1$ random vector, with the kth element being an indicator function that assumes the value 1, when the realization of the sibling marker genotypes corresponds to the kth element of the conditional outcome space, and 0 otherwise. The set of possible outcomes is given in tables 4–7 in Rabinowitz and Laird (2000) for nuclear families. Because, under the type II H_0 and conditional on $\boldsymbol{\Phi}_{II}$, all outcomes are equally likely, with probability $1/p$, $\boldsymbol{r}$ follows a multinomial distribution, with mean and variance given by

$$\boldsymbol{\mu}_r = E(\boldsymbol{r}|\boldsymbol{\Phi}_{II}) = \frac{1}{p}1_p$$

and

$$\boldsymbol{\Sigma}_r = \text{Var}(\boldsymbol{r}|\boldsymbol{\Phi}_{II}) = \frac{1}{p}\left(\boldsymbol{I}_p - \frac{1}{p}1_p 1_p'\right),$$

where 1_p is a $p \times 1$ vector of 1s and $\boldsymbol{I}_p$ is a $p \times p$ dimensional identity matrix.

The moments of $\boldsymbol{S}_i$ can be derived using the moments of $\boldsymbol{r}$. Let $\boldsymbol{S}_i^r$ be an $h \times p$ matrix with the kth column equal to $\sum_j T_{ij}\boldsymbol{X}(m_{ij}^{(k)})$ where $\boldsymbol{m}^{(k)} = (m_{i1}^{(k)},\ldots,m_{im_i}^{(k)})$ is the vector of sibling marker genotypes corresponding to the kth element of the conditional outcome space and h is the length of the marker genotype coding vector $\boldsymbol{X}$. The conditional mean and variance of $\boldsymbol{S}_i$ are

$$\boldsymbol{\mu}_{S_i} = E(\boldsymbol{S}_i|\boldsymbol{\Phi}_{II}) = \boldsymbol{S}_i^r m_r$$

and

$$\boldsymbol{\Sigma}_{S_i} = \text{Var}(\boldsymbol{S}_i|\boldsymbol{\Phi}_{II}) = \boldsymbol{S}_i^r\boldsymbol{\Sigma}_r(\boldsymbol{S}_i^r)' .$$

Under the type II H_0, the approximate distribution of $\boldsymbol{S} - E(\boldsymbol{S}|\boldsymbol{\Phi}_{II})$ is $N_b(0,\sum_i \boldsymbol{\Sigma}_{S_i})$.

The last column of table 1 indicates which combinations of parental marker genotypes and children marker configurations are potentially informative in the biallelic setting with the RL algorithm applied to the type II H_0 setting. When parental data are missing (as is often the case for late-onset diseases), sibships with more than two sibs and $C_m = \{AA,AB\}$ or $C_m = \{BB,AB\}$ are not informative, because allele sharing cannot be discerned. The removal of these types of sibships may cause a substantial loss in the effective sample size, especially when one of the alleles is rare, because homozygotes of the rare allele will be infrequent. An alternative to conditioning on the allele sharing is to take advantage of the linear form of the test statistic (eq. [1]) and to use the RL algorithm for the type I H_0 to calculate the expectation, in conjunction with a robust variance-covariance estimator. The development of this approach follows.

Factorization of $p(\boldsymbol{m}|\boldsymbol{\Phi}_I)$ under Type II H_0

In view of the potentially severe loss of information caused by conditioning on sibling identical-by-descent (IBD) patterns, we here develop a method that employs the type I H_0 RL algorithm to compute $\sum_{i=1}^{N} \boldsymbol{S}_i - E(\boldsymbol{S}_i|\boldsymbol{\Phi}_I)$ and an empirical variance-covariance estimator that is robust to the correlation among the sibling marker genotypes. To show that $\sum_{i=1}^{N} \boldsymbol{S}_i - E(\boldsymbol{S}_i|\boldsymbol{\Phi}_I)$ is a valid measure of association in the presence of linkage, we derive the marginal conditional distribution for the kth sibling marker genotype $p(m_k|\boldsymbol{\Phi}_I)$ and show that this marginal distribution is the same under both the type I H_0 and the type II H_0 and does not depend on the recombination parameter θ or on the observed phenotypes $\boldsymbol{y}$ for $k = 1, \ldots, n$ (see Appendix). Since the linear form of the test statistic (eq. [1]) permits its expectation to be found using $p(m_k|\boldsymbol{\Phi}_I)$, the RL algorithm for the type I H_0 can be used to compute $E(\boldsymbol{S}_i|\boldsymbol{\Phi}_I)$. Therefore, without specification or estimation of θ and without parameterization of the phenotype distribution, $\boldsymbol{S} - E(\boldsymbol{S}|\boldsymbol{\Phi}_I)$ can be used to construct an unbiased test for association in the presence of linkage. Since family-specific contributions comprise $\boldsymbol{S} - E(\boldsymbol{S}|\boldsymbol{\Phi}_I)$, only the variances of these contributions are needed to compute $\mathrm{Var}[\boldsymbol{S} - E(\boldsymbol{S}|\boldsymbol{\Phi}_I)]$; the correlation among children need not be addressed when finding $\mathrm{Var}[\boldsymbol{S}_i - E(\boldsymbol{S}_i|\boldsymbol{\Phi}_I)]$.

The derivation in the Appendix employs an ordered notation similar to that of Thomson (1995), where m_k^* is the marker genotype of the kth child, expressed in terms of the parental derived haplotypes (see Appendix). In particular, it is shown that under both the type I H_0 and the type II H_0, the joint conditional probability for a family can be factored into

$$\Pr(\boldsymbol{m}|\boldsymbol{\Phi}_I) = \sum_{\boldsymbol{M}_u \in \mathscr{A}} \Pr(\boldsymbol{m}_{-k}|m_k,\boldsymbol{M},\boldsymbol{y}) \times \left[\frac{\sum_{m_k^* \in \mathscr{B}} \Pr(m_k^*,\boldsymbol{M})}{S(\boldsymbol{M})}\right],$$

where $\boldsymbol{m}_{-k}$ is the vector of sibling marker alleles with the kth sibling information omitted, $\boldsymbol{M}_u$ is the unobserved parental marker genotypes, $\mathscr{A}$ is the set of unobserved parental maker genotypes that coincide with $S(\boldsymbol{M})$ and $\mathscr{B}$ corresponds to the set of paternal and maternal derived markers for parents with marker genotypes $\boldsymbol{M}$ that result in the kth sibling's observed marker genotype m_k. Marginalization of $\Pr(\boldsymbol{m}|\boldsymbol{\Phi}_I)$ with respect to $\boldsymbol{m}_{-k}$ results in the marginal conditional probability for the kth sibling marker genotype with $\Pr(m_k|\boldsymbol{\Phi}_I) = \Pr[m_k|S(\boldsymbol{M})]$. In addition, we show that $\Pr[m_k|S(\boldsymbol{M})]$ is not a function of θ and can be computed using the RL algorithm for the type I H_0. Although the factorization can be used to find the correct conditional expectation of the test statistic, it cannot be used to derive expressions for the covariance between sibling marker genotypes, because it marginalizes over the IBD relationships.

Since $\boldsymbol{S}_i - E(\boldsymbol{S}_i|\boldsymbol{\Phi}_I)$ are independent mean 0 random vectors with unspecified variance-covariance matrices, we can apply the results of White (1980) to construct a robust variance-covariance estimator of $\boldsymbol{S} - E(\boldsymbol{S}|\boldsymbol{\Phi}_I)$. Specifically, White (1980) addresses estimation of the variance-covariance matrix for estimated regression pa-

Table 1

Nuclear Family Informativeness for Both Conditioning Approaches

Parental Genotypes[a]	Children Configuration[b]	Family Informativeness	
		EV-FBAT	RL Algorithm Type II H_0
AA,AA	NA	No	No
AA,AB	NA	Yes	Yes
AA,BB	NA	No	No
AB,AB	NA	Yes	Yes
AA,−	{*AA*}	No	No
AA,−	{*AA*,*AB*}	Yes	Yes
AB,−	{*AA*}	No	No
AB,−	{*AB*}	No	No
AB,−	{*AA*,*AB*}	Yes	No when $n > 2$
AB,−	{*AA*,*BB*}	Yes	Yes
AB,−	{*AA*,*AB*,*BB*}	Yes	Yes
−,−	{*AA*}	No	No
−,−	{*AB*}	No	No
−,−	{*AA*,*AB*}	Yes	No when $n > 2$
−,−	{*AA*,*BB*}	Yes	Yes
−,−	{*AA*,*AB*,*BB*}	Yes	Yes

[a] − = Not genotyped.
[b] NA = not applicable.

rameters in linear models with heteroscedastic errors. The test statistic $S - E(S_i|\Phi_I)$ can be couched as proportional to a vector of parameter estimates from a linear model and, therefore, the White empirical variance-covariance estimator, given by

$$\hat{\Sigma}_W = \hat{\mathrm{Var}}\left\{\sum_{i=1}^{N}[S_i - E(S_i|\Phi_I)]\right\} = \sum_{i=1}^{N}[S_i - E(S_i|\Phi_I)][S_i - E(S_i|\Phi_I)]' , \quad (2)$$

provides a consistent estimate of the variance-covariance matrix of $S - E(S|\Phi_I)$. Alternatively, $\hat{\Sigma}_W$ can be derived using the results of Liang and Zeger (1986) on generalized estimating equations. When S is vector-valued, $\hat{\Sigma}$ may not be full rank. In this case, the test statistic for the type II H_0 is $[S - E(S|\Phi_I)]'\hat{\Sigma}_W^-[S - E(S|\Phi_I)]$, where $\hat{\Sigma}_W^-$ is the generalized inverse of $\hat{\Sigma}_W$. It should be noted that the empirical variance-covariance estimator (2) reduces to a simple sum of squares for the biallelic case.

Extensions to more-complex pedigrees are straightforward. Assume that the ith pedigree can be split into q_i nuclear families, for $i = 1, \ldots, F$, and let

$$S - E(S|\Phi_I) = \sum_{i=1}^{F}\sum_{j=1}^{q_i}[S_{ij} - E(S_{ij}|\Phi_I)] ,$$

where S_{ij} is the test-statistic contribution from the jth nuclear family in the ith pedigree and $E(S_{ij}|\Phi_I)$ is computed using formulas by S. Horvath, X. Xu, and N. Laird (unpublished data). Although the contributions from nuclear families in the same pedigree are not independent, we can again appeal to White (1980) to construct a consistent estimate of the variance-covariance matrix of $S - E(S|\Phi_I)$:

$$\hat{\Sigma}_W = \sum_{i=1}^{F}\left[\sum_{j=1}^{q_i}S_{ij} - E(S_{ij}|\Phi_I)\right]\left[\sum_{j=1}^{q_i}S_{ij} - E(S_{ij}|\Phi_I)\right]' .$$

The advantage of the empirical variance-covariance approach is that more nuclear-family marker configurations are informative than is the case with the type II conditioning method. Table 1 indicates which nuclear family configurations are informative for the two approaches in the setting of a biallelic marker. In addition, since the conditioning is different for the two approaches, the expected values and variance-covariance terms are also not the same. We will refer to the empirical variance-covariance approach as "EV-FBAT."

Example: Testing for Association in the *A2M* Gene

As an example, we tested for association between the *A2M*-18i deletion and AD in a set of sibships from the National Institute of Mental Health (NIMH) Genetics Initiative AD Sample. The ascertainment and assessment of the AD families collected have been discussed elsewhere (Blacker et al. 1997). The sample we used is composed of 437 individuals in 120 sibships and is identical to the sample analyzed by Blacker et al. (1999); 246 of the siblings met the NINCDS/ADRDA criteria for AD and/or had autopsy confirmation of the diagnosis.

Table 2 contains the results for testing the *A2M*-18i/AD association. The test statistic used in the applications of the RL algorithm is the sum of the *A2M*-1 alleles in AD-affected siblings. This corresponds to the following coding schemes:

$$T_{ij} = \begin{cases} 1 & \text{if sibling } j \text{ in } i\text{th sibship is affected} \\ 0 & \text{otherwise} \end{cases}$$

and

$$X_{ij} = \begin{cases} 2 & \text{if } m_{ij} = A2M\text{-}1/A2M\text{-}1 \\ 1 & \text{if } m_{ij} = A2M\text{-}1/A2M\text{-}2 \\ 0 & \text{otherwise} \end{cases} .$$

Implementation of the RL algorithm consists of finding the expected value of X_{ij} conditional on the minimal sufficient statistic corresponding to the null hypothesis. Variance estimation is accomplished through the procedures described above.

Application of the RL algorithm to test for linkage and association (type I H_0) results in 51 informative sibships and a significant finding. As discussed above, the type I H_0 may not be appropriate in view of the reported linkage evidence in the region spanning the *A2M* gene. Conditioning on the type II H_0 minimal sufficient statistic results in a dramatic decrease in the effective sample size. With only 10 informative sibships, the test statistic is only marginally significant, and its large sample χ^2 approximation may not be reliable (ta-

Table 2

***A2M*/Alzheimer Disease Association Test Results for Various Methods**

Method	No. of Informative Sibships	χ^2	P
Type I RL algorithm	51	8.599	.0034
Type II RL algorithm	10	6.125	.0133
EV-FBAT	44	8.631	.0033
Siegmund et al. (2000)	51	6.916	.0085
PDT	50	8.387	.0038
SDT	46	...	.0016

ble 2). With EV-FBAT, 44 sibships were informative resulting in a highly significant result ($\chi^2 = 2.94$, $P = .0033$).

The discrepancy in the number of informative families is a consequence of the absence of parental genotype data and the distribution of genotypes among the siblings [$p(A2M\text{-}1/A2M\text{-}1) = .732$, and $p(A2M\text{-}1/A2M\text{-}2) = .231$, $p(A2M\text{-}2/A2M\text{-}2) = .037$]. The 34 families that are informative for EV-FBAT but not informative for the type II H_0 conditioning approach have more than two siblings and $\mathbf{C}_m = \{A2M\text{-}1/A2M\text{-}1, A2M\text{-}1/A2M\text{-}2\}$ or $\mathbf{C}_m = \{A2M\text{-}2/A2M\text{-}2, A2M\text{-}1/A2M\text{-}2\}$ as the sibling marker configuration. As indicated by table 1, these sibships are not informative for the type II H_0 RL approach because no definite allele sharing can be discerned. Because it does not condition on the allele sharing, the empirical variance approach is not subject to these constraints. The difference between the number of informative families for the type I H_0 RL test and for EV-FBAT is a result of the definition of the empirical variance (2). Families with $\mathbf{S}_i = E(\mathbf{S}_i|\mathbf{\Phi}_I)$ do not contribute to the test statistic or the empirical variance-covariance estimate.

To justify the EV-FBAT χ^2 approximation with 44 informative sibships, we empirically estimated the significance level under the type II H_0 for various numbers of informative sibships. We simulated sibships that were similar to the NIMH sibships in that the size distribution of the sibships was maintained, the biallelic marker had population allele frequencies of 0.20 and 0.80, and the baseline prevalence was fixed at 0.30. Because simulated data with the same number of sibships will have different numbers of informative families, we report the mean number of informative families. For each number of sibships we simulated 10,000 data sets. In figure 1, the circles represent the empirical significance levels for the mean number of informative families. The dashed lines are the pointwise 95% Monte Carlo sampling-error levels (0.0457, 0.0543). Figure 1 shows that the empirical significance level is within Monte Carlo sampling error for a large range of informative sibships. Indeed, the χ^2 approximation appears to hold even for samples with only 20 informative sibships. With <20 informative sibships, the test appears to become conservative.

Robust variance-covariance estimation has been implemented in the context of a TDT extension (TRANSMIT; Clayton 1999), conditional logistic regression (Siegmund et al. 2000), and the PDT (Martin et al. 2000*c*). All three procedures are limited to qualitative traits, whereas the application of Siegmund et al. (2000) is further restricted to discordant sibships. When applied to the *A2M* data set, the Wald statistic from conditional logistic regression with robust variance estimation produces a test statistic that is not as pronounced as that of EV-FBAT but is still significant (table 2). The PDT produces a test statistic that is essentially equivalent to the test statistic of EV-FBAT in these data.

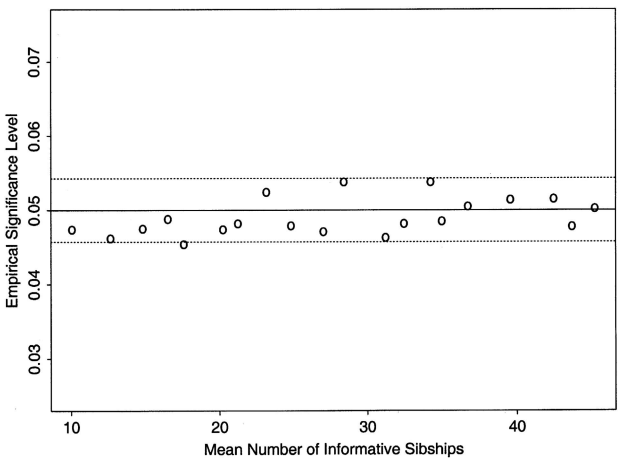

Figure 1 Empirical significance levels under the type II H_0 for average number of informative sibships. The dashed lines are the pointwise 95% Monte Carlo sampling error levels (0.0457, 0.0543).

Another alternative is to use the sibship disequilibrium test (SDT; Horvath and Laird 1998). As shown in table 2, the SDT provides the strongest evidence for linkage disequilibrium. The SDT is well suited to the discordant sibships setting of the NIMH data, but it is restricted to qualitative phenotypes and cannot efficiently handle families with genotype-known parents.

Discussion

One strategy for positional genomic analysis is to focus allelic-association testing on regions that have been identified through linkage analysis as putatively containing a gene or genes influencing phenotypic variation. Supplementing linkage results with association methodology is needed because, with complex diseases, linkage peaks may span regions of ⩾10–20 cM that cover a large number of genes and are beyond the reach of positional cloning (Hauser and Boehnke 1997). A significant association finding may greatly refine the search for the underlying trait gene, since linkage disequilibrium will not generally extend over regions >1 cM in outbred populations (Pericak-Vance 1998). Although the utility of association methodology in this setting has been questioned (Terwilliger and Weiss 1998), the use of association methodology in the dissection of a region linked to human hypertension has recently yielded a susceptibility locus (Bray et al. 2000).

Candidates for the association tests within regions

identified by linkage may be chosen via database searches using knowledge of biological pathways (Brookes et al. 2000). In addition, as dense maps of single-nucleotide polymorphisms (SNPs) become available and costs of genotyping decline, the dissection of linked regions may be accomplished by saturating the linked regions with SNPs and performing association tests on them. Martin et al. (2000*a*, 2000*b*) have used the *APOE* gene to illustrate the potential for using SNPs in mapping studies of complex traits.

With these strategies in mind, we have presented a method for evaluating the mean and variance-covariance of a wide range of test statistics computed under the null hypothesis that there is linkage but no association (type II H_0). The method, EV-FBAT, determines the expected value of an association test statistic by conditioning on the minimal sufficient statistic under the null hypothesis of no linkage and no association (type I H_0) and uses an empirical variance-covariance estimator that is consistent even when the sibling marker genotypes are correlated. As discussed above, the expectation of the test statistic is computed via the RL algorithm, and the resulting standardized test statistic is unbiased as a test for association in the presence of linkage. In addition, while retaining the robust properties of family based association tests, EV-FBAT does not suffer from the costly reduction in sample size caused by missing parental data that is inherent with approaches that condition on sibling IBD patterns.

The results of the *A2M*/AD example strongly suggest that the *A2M*-18i deletion is in linkage disequilibrium with a polymorphism that contributes to AD development. Whether or not the *A2M*-18i polymorphism is the polymorphism of interest (in which case the linkage disequilibrium is complete) cannot be deduced by association tests. In light of the evidence for linkage, relying on the type I H_0 test alone would leave open the interpretation of the *P* value. Here, the *P* values of the type I RL approach and EV-FBAT agree; in general, we expect the type II H_0 *P* values to be larger if H_0 is true. Additional work will investigate the power of EV-FBAT and various proposed methods under H_a.

For qualitative traits and biallelic markers, EV-FBAT is similar to the PDT (Martin et al. 2000*c*). In the PDT, pedigrees are broken into nuclear families and discordant sibships. Let *A* and *B* be the two alleles of the marker. The contribution to the test statistic of a particular pedigree consists of weighted sums of the number of *A* alleles for each affected child minus an "expected" number of *A* alleles. This expectation is computed from unaffected siblings when the affected child belongs to a discordant sibship and is computed using a pseudo-control (as defined by Falk and Rubinstein 1987) when the affected child belongs to a nuclear family. If a child belongs to a nuclear family and a discordant sibship, both differences are computed. Under the type II H_0, the sum of the pedigree contributions has expectation 0 and is standardized with an empirical estimator of the variance.

In this setting, the difference between the PDT and EV-FBAT is in the derivation of the expected number of *A* alleles under the type II H_0. In using the RL algorithm for the type I H_0, EV-FBAT conditions on the minimal sufficient statistic and, by definition, makes the most efficient use of the observed data in constructing the control genotype (see Cox and Hinkley [1974] or Rabinowitz and Laird [2000]). Further, the PDT can not use concordant sibships with missing parental marker information and is also limited to the dichotomous-phenotype case.

EV-FBAT uses a robust variance-covariance estimation to take into account the correlation among sibling marker genotypes under the type II H_0. In addition to the PDT and EV-FBAT, a robust variance-covariance estimation for the qualitative setting has been implemented in the context of a TDT extension (TRANSMIT; Clayton 1999) and conditional logistic regression (Siegmund et al. 2000). The method of Clayton (1999) uses the EM algorithm (Dempster et al. 1977) to impute the likelihood contribution from family trios in which there is missing parental information and/or ambiguous genetic transmissions. Such imputation requires a full specification of the family-trio likelihood that depends on estimates of allele frequencies and population genetic assumptions that are difficult to justify. A score test based on these likelihood contributions is used to test for association with a robust variance-covariance estimator when multiple siblings are allowed.

The merits of association tests based on conditional logistic regression have been discussed (Witte et al. 1998; Kraft and Thomas 2000). Siegmund et al. (2000) recommend generalized estimating equations applied to the conditional logistic likelihood when the type II H_0 is used. Unlike EV-FBAT, this method does not make any use of available parental data and is restricted to discordant sibships. As with the PDT, both TRANSMIT and the Siegmund et al. (2000) procedure are limited to qualitative traits.

In summary, EV-FBAT provides a flexible framework for association testing in the presence of linkage because it can be used with any type of phenotype and with any pedigree configuration. Therefore, the researcher is not restricted to particular sampling designs and is free to test for associations with quantitative or time-to-onset traits. Indeed, with EV-FBAT, the approaches to association testing with binary, quantitative, and time-to-onset phenotypes for the type I H_0 advocated by S. Horvath, X. Xu, and N. Laird (unpublished data) can all be adapted to the type II H_0. Application of EV-FBAT is limited to the class of test statistics that can be ex-

pressed in a linear form (eq. [1]), but, as discussed in Laird et al. (2000), a number of family-based association-test statistics are of this form. Furthermore, Clayton and Jones (1999) and Lunetta et al. (2000) have shown that the score statistics from generalized linear models in which the coded marker genotype is the covariate can be expressed in the form of equation (1). The case when the test statistic may depend on unknown nuisance parameters is discussed in Lunetta et al. (2000). The method is also valid as a test of the type I H_0 of no linkage or no association, since the empirical variance-covariance estimator is a consistent estimator under both types of null hypotheses.

The empirical variance approach for testing association in the presence of linkage has been implemented in a program called FBAT. It is invoked with the `-e` (for empirical variance) option for the `fbat` command. The program and its documentation are available free of charge from our Web site. There are different versions of the program for different operating systems: MAC, Solaris/Sparc, and Windows. If you encounter problems, please e-mail fbat@hsph.harvard.edu.

Acknowledgements

We thank Dr. Steve Horvath for valuable conversations and Dr. John Rogus for helpful comments on the manuscript. Support for this research was provided by National Institutes of Health (NIH) grant MH 59532. We are indebted to two anonymous referees for their helpful suggestions. The genotypes of the sibships were generated in the laboratory of Dr. Rudy Tanzi, with support from NIH grant R01 MH60009. Data and biomaterials were collected in three projects that participated in the NIMH Alzheimer Disease Genetics Initiative. From 1991 to 1998, the principal investigators and coinvestigators were: Marilyn S. Albert, Ph.D., and Deborah Blacker, M.D., Sc.D., Massachusetts General Hospital, Boston, grant U01 MH46281; Susan S. Bassett, Ph.D., Gary A. Chase, Ph.D., and Marshal F. Folstein, M.D., Johns Hopkins University, Baltimore, grant U01 MH46290; and Rodney C. P. Go, Ph.D., and Lindy E. Harrell, M.D., University of Alabama, Birmingham, grant U01 MH46373.

Appendix A

Proof

We show that, under the type II H_0, the joint conditional distribution of the sibling marker genotypes $\boldsymbol{m}$ given the sufficient statistic for the parental marker genotypes $S(\boldsymbol{M})$ and the observed phenotypes $\boldsymbol{y}$ can be factored into a form amenable to the approach discussed above. The key point is that the marginal conditional distribution of a child's marker genotype is not a function of the recombination parameter θ or of the observed phenotypes $\boldsymbol{y}$. Therefore, under the type II H_0, the expectation of the test statistic conditional on the minimal sufficient statistic for the type I H_0 can be found using the type I H_0 RL algorithm, without modeling the correlation between the children's marker genotypes.

Since $S(\boldsymbol{M}) = (\boldsymbol{C}_m, \boldsymbol{M}_{\text{obs}})$, where $\boldsymbol{C}_m$ is the configuration of sibling marker genotypes and $\boldsymbol{M}_{\text{obs}}$ is any observed parental marker genotype, the joint conditional distribution can be expressed as

$$\begin{aligned}
\Pr[\boldsymbol{m}|S(\boldsymbol{M}),\boldsymbol{y}] &= \Pr[S(\boldsymbol{M}),\boldsymbol{y}]^{-1} \Pr[\boldsymbol{m},S(\boldsymbol{M}),\boldsymbol{y}] \\
&= \Pr[S(\boldsymbol{M}),\boldsymbol{y}]^{-1} \Pr(\boldsymbol{m},\boldsymbol{C}_m,\boldsymbol{M}_{\text{obs}},\boldsymbol{y}) \\
&= \Pr[S(\boldsymbol{M}),\boldsymbol{y}]^{-1} \Pr(\boldsymbol{m},\boldsymbol{M}_{\text{obs}},\boldsymbol{y}) \\
&= \Pr[S(\boldsymbol{M}),\boldsymbol{y}]^{-1} \sum_{\boldsymbol{M}_u \in \mathcal{M}_{S(\boldsymbol{M})}} \Pr(\boldsymbol{m},\boldsymbol{M},\boldsymbol{y}) \ ,
\end{aligned}$$

where $\mathcal{M}_{S(\boldsymbol{M})}$ is the set of possible unobserved parental marker genotypes with elements $\boldsymbol{M}_u$ that correspond to $S(\boldsymbol{M})$ and where $\boldsymbol{M} = (\boldsymbol{M}_{\text{obs}}, \boldsymbol{M}_u)$.

To derive the marginal conditional distribution of a child's marker genotype we arbitrarily select the kth sibling (referred to as the reference sibling) and let $\boldsymbol{m}_{-k}$ be the vector of sibling marker alleles with the kth sibling information omitted. For all $k = 1, \ldots, n$ we have that

$$\sum_{\boldsymbol{M}_u \in \mathcal{M}_{S(\boldsymbol{M})}} \Pr(\boldsymbol{m},\boldsymbol{M},\boldsymbol{y}) = \sum_{\boldsymbol{M}_u \in \mathcal{M}_{S(\boldsymbol{M})}} \Pr(\boldsymbol{m}_{-k}|m_k,\boldsymbol{M},\boldsymbol{y}) \Pr(m_k,\boldsymbol{M},\boldsymbol{y}) \ .$$

We next show that $\Pr(m_k,\boldsymbol{M},\boldsymbol{y}) = \Pr(m_k,\boldsymbol{M}) f(\boldsymbol{y})$, where $f(\boldsymbol{y})$ is the joint distribution of the sibling phenotypes. To do this, we adopt a notation similar to the ordered notation of Thomson (1995), which identifies the paternally

and maternally derived haplotypes that comprise the marker genotypes of the children. This is accomplished by expanding the parental marker genotypes into specific haplotypes, $\boldsymbol{M}_i^* = [m_{i1}^{(p)}/m_{i2}^{(p)}, m_{i1}^{(m)}/m_{i2}^{(m)}]$, and letting m_{ij}^* be the marker genotype of the jth child expressed in terms of the parental-derived haplotypes. That is, $m_{ij}^* = [m_{id_j}^{(p)}/m_{id_j'}^{(m)}]$, where d_j, $d_j' = 1,2$ indicate inheritance from each parent. Furthermore, let $\mathcal{B}_{m_k^*,M}$ correspond to the set of paternally and maternally derived markers from parents with marker genotypes $\boldsymbol{M}$ that result in the kth sibling's observed marker genotype m_k, and let $\boldsymbol{G} = [g_1^{(p)}/g_2^{(p)}, g_1^{(m)}/g_2^{(m)}]$ be the unobserved disease genotypes for the parents and $\boldsymbol{g}$ be the vector of unobserved disease genotypes for the children. The joint probability, $\Pr(m_k,\boldsymbol{M},\boldsymbol{y})$, thus can be expressed as the summation

$$\begin{aligned}\Pr(m_k,\boldsymbol{M},\boldsymbol{y}) &= \sum_{m_k^* \in \mathcal{B}} \Pr(m_k^*,\boldsymbol{M},\boldsymbol{y}) \\ &= \sum_{m_k^* \in \mathcal{B}} \sum_G \sum_{g \in \mathcal{G}} \Pr(\boldsymbol{y},\boldsymbol{g},m_k^*,\boldsymbol{H}) \ , \end{aligned} \tag{A1}$$

where the additional summations in (A1) are with respect to the set of possible parental disease genotype combinations and the set of siblings' disease genotypes conditional on parental disease genotypes $\mathcal{G}$ and where $\boldsymbol{H} = [m_1^{(p)}g_1^{(p)}/m_2^{(p)}g_2^{(p)}, m_1^{(m)}g_1^{(m)}/m_2^{(m)}g_2^{(p)}]$ describes the parental haplotypes.

Under the assumption that sibling disease genotypes are conditionally independent given parental haplotypes, equation (A1) can be expressed as

$$\sum_{m_k^* \in \mathcal{B}} \sum_G \sum_{g \in \mathcal{G}} f(\boldsymbol{y}|\boldsymbol{g}) \left[\prod_{i \neq k} \Pr(g_i|\boldsymbol{H}) \right] \Pr(g_k|m_k^*,\boldsymbol{H}) \Pr(m_k^*,\boldsymbol{H}) \ . \tag{A2}$$

Under the type II null hypothesis of no association, we have that $\Pr(g_i|\boldsymbol{H}) = \frac{1}{4}$ for $i = 1,\ldots,n;\ i \neq k$, and $\Pr(m_k^*,\boldsymbol{H}) = \Pr(m_k^*,\boldsymbol{M}) \Pr(\boldsymbol{G})$. Therefore, (A2) can be simplified to

$$\sum_{m_k^* \in \mathcal{B}} \Pr(m_k^*,\boldsymbol{M}) \left\{ \sum_G \left[\left(\frac{1}{4}\right)^{n-1} \sum_{g \in \mathcal{G}} f(\boldsymbol{y}|\boldsymbol{g}) \Pr(g_k|m_k^*,\boldsymbol{H}) \right] \Pr(\boldsymbol{G}) \right\} \ . \tag{A3}$$

Let F_G denote the expression within square brackets in equation (A3). There are 4^n terms in F_G, corresponding to all the combinations of disease genotypes in the n children. The summation over all combinations of parental disease genotypes makes the terms in F_G with the same parental disease allele sharing patterns equivalent. For example, in the case of two children with the first child being the reference sibling,

$$\begin{aligned}\sum_G f\{\boldsymbol{y}|g_1 = [g_1^{(p)},g_1^{(m)}],g_2 = [g_1^{(p)},g_2^{(m)}]\} \Pr(\boldsymbol{G}) &= \sum_G f\{\boldsymbol{y}|g_1 = [g_2^{(p)},g_1^{(m)}],g_2 = [g_2^{(p)},g_2^{(m)}]\} \Pr(\boldsymbol{G}) \\ &= \sum_G f\{\boldsymbol{y}|g_1 = [g_1^{(p)},g_2^{(m)}],g_2 = [g_1^{(p)},g_1^{(m)}]\} \Pr(\boldsymbol{G}) \\ &= \sum_G f\{\boldsymbol{y}|g_1 = [g_2^{(p)},g_2^{(m)}],g_2 = [g_2^{(p)},g_1^{(m)}]\} \Pr(\boldsymbol{G}) \ . \end{aligned}$$

Furthermore, if we assume $m_1^* = (m_1^{(p)},m_1^{(m)})$, then we have that

$$\begin{aligned}&\sum_G \sum_{g \in \mathcal{G}_{IBD=1_p}} f(\boldsymbol{y}|g_1,g_2) \Pr(g_1|m_1^*,\boldsymbol{H}) \Pr(\boldsymbol{G}) \\ &= \sum_G f\{\boldsymbol{y}|g_1 = [g_1^{(p)},g_1^{(m)}],g_2 = [g_1^{(p)},g_2^{(m)}]\} \Pr(\boldsymbol{G}) \ , \end{aligned}$$

where $\mathcal{G}_{IBD=1_p}$ is the set of disease allele–sharing patterns, between the two siblings, that result in them sharing the paternally but not the maternally derived disease allele. Because of the ordered notation, $\Pr(g_1|m_1^*,\boldsymbol{H})$ is a simple function of the recombination parameter θ, which cancels in the summation.

The same logic can be applied to any disease allele–sharing patterns for any number of children, making it straightforward to show that $\sum_G F_G \Pr(G) = f(\boldsymbol{y})$. Therefore, $\Pr(m_k,\boldsymbol{M},\boldsymbol{y}) = \Pr(m_k,\boldsymbol{M}) f(\boldsymbol{y})$, where $\Pr(m_k,\boldsymbol{M})$ is not a function of θ or of $\boldsymbol{y}$, and we have the following factorization of the joint conditional distribution:

$$\Pr[\boldsymbol{m}|S(\boldsymbol{M}),\boldsymbol{y}] = \sum_{M_u \in \mathcal{M}} \Pr(\boldsymbol{m}_{-k}|m_k,\boldsymbol{M},\boldsymbol{y}) \left[\frac{\sum_{m_k^* \in \mathcal{B}} \Pr(m_k^*,\boldsymbol{M})}{S(\boldsymbol{M})} \right],$$

where we have used the fact that, under the type II H_0, $\Pr[S(\boldsymbol{M}),\boldsymbol{y}] = \Pr[S(\boldsymbol{M})]\Pr(\boldsymbol{y})$. We can marginalize the joint distribution with respect to $\boldsymbol{m}_{-k}$ to obtain

$$\Pr[m_k|S(\boldsymbol{M}),\boldsymbol{y}] = \frac{\sum_{M_u \in \mathcal{M}, m_k^* \in \mathcal{B}} \Pr(m_k^*,\boldsymbol{M})}{S(\boldsymbol{M})} . \quad \text{(A4)}$$

The term on the right side of (A4) is the conditional distribution of marker genotypes for the kth sibling, $\Pr[m_k|S(\boldsymbol{M})]$, under the null hypothesis of no linkage and no association. It has been tabulated by Rabinowitz and Laird (2000), for arbitrary missing parental marker information, and can be used to derive $E(S_i|\boldsymbol{\Phi}_I)$ under the type II H_0. In summary, we have shown that $\sum_{i=1}^{N} S_i - E(S_i|\boldsymbol{\Phi}_I)$ is a valid measure of association in the presence of linkage.

Electronic-Database Information

The URL for data in this article is as follows:

FBAT Web page, http://www.biostat.harvard.edu/~fbat/default.html (for free FBAT program and documentation)

References

Blacker D, Haines JL, Rhodes L, Terwedow H, Go RCP, Harrell LE, Perry RT, Bassett SS, Chase G, Meyers D, Albert MS, Tanzi R (1997) ApoE-4 and age at onset of Alzheimer's disease: the NIMH genetics initiative. Neurology 48:139–147

Blacker D, Wilcox MA, Laird NM, Rhodes L, Horvath SM, Go RCP, Perry R, Watson B, Bassett SS, McInnis MG, Albert MS, Hyman BT, Tanzi RE (1998) Alpha-2 macroglobulin is genetically associated with Alzheimer disease. Nat Genet 19: 357–360

Blacker D, Crystal AS, Wilcox MA, Laird NM, Tanzi RE (1999) An alpha-2 macroglobulin insertion-deletion polymorphism in Alzheimer disease-Reply. Nat Genet 22:21–22

Bray MS, Krushkal J, Li L, Ferrell R, Kardia S, Sing CF, Turner ST, Boerwinkle E (2000) Positional genomic analysis identifies the β_2-Adrenergic receptor gene as a susceptibility locus in human hypertension. Circulation 101:2877–2882

Brookes AJ, Emahazion T, Howell WM, Jobs M, Sawyer S, Fredman D, Siegfried M, Feuk L, Prince JA (2000) Using intra-genic SNPs to study complex disease: tools, systems and practical experience. Paper presented at DNA2000: International Symposium on the State of the Art in Genetic Analysis. Boston, June 1–3

Clayton D (1999) A generalization of the transmission/disequilibrium test for uncertain haplotype transmission. Am J Hum Genet 65:1170–1177

Clayton D, Jones H (1999) Transmission/disequilibrium test for extended marker haplotypes. Am J Hum Genet 65:1161–1169

Cox DR, Hinkley DV (1974) Theoretical statistics. Halsted Press, New York

Dempster A, Laird NM, Rubin D (1977) Maximum likelihood from incomplete data via the EM algorithm. J R Stat Soc B 39:1–22

Dow DJ, Lindsey N, Cairns NJ, Brayne C, Robinson D, Huppert FA, Paykel ES, Xuereb J, Wilcock G, Whittaker JL, Rubinsztein DC (1999) α-2 macroglobulin polymorphism and Alzheimer disease risk in the UK. Nat Genet 22:16–17

Ewens W, Spielman RS (1995) The transmission/disequilibrium test: history, subdivision and admixture. Am J Hum Genet 57:455–464

Falk CT, Rubinstein P (1987) Haplotype relative risks: an easy reliable way to construct a proper control sample for risk calculations. Ann Hum Genet 51:227–233

Hauser ER, Boehnke M (1997) Confirmation of linkage results in affected-sib-pair linkage analysis for complex genetic traits. Am J Hum Genet Suppl 61:A278

Horvath S, Laird NM (1998) A discordant-sibship test for disequilibrium and linkage: no need for parental data. Am J Hum Genet 63:1886–1897

Kehoe P, Wavrant-De Vrieze F, Crook R, Wu WS, Holmans P, Fenton I, Spurlock G, Norton N, Williams H, Williams N, Lovestone S, Perez-Tur J, Hutton M, Chartier-Harlin MC, Shears S, Roehl K, Booth J, Van Voorst W, Ramic D, Williams J, Goate A, Hardy J, Owen MJ (1999) A full genome scan for late onset Alzheimer's disease. Hum Mol Genet 8: 237–245

Kraft P, Thomas DC (2000) Bias and efficiency in family based gene-characterization studies: conditional, prospective, and joint likelihoods. Am J Hum Genet 66:1119–1131

Laird NM, Horvath S, Xu X (2000) Implementing a unified approach to family based tests of association. Genet Epidemiol Suppl 19:S36–S42

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

Liang KY, Zeger SL (1986) Longitudinal data analysis using generalized estimating equations. Biometrika 73:13–22

Lunetta KL, Faraone SV, Beidermann J, Laird NM (2000) Family based tests of association that use unaffected sibs, covariates, and interactions. Am J Hum Genet 66:605–614

Martin ER, Gilbert JR, Lai EH, Riley J, Rogala AR, Slotterback BD, Sipe CA, Grubber JM, Warren LL, Conneally PM, Saunders AM, Schmechel DE, Purvis I, Pericak-Vance MA, Roses AD, Vance JM (2000*a*) Analysis of association at single nucleotide polymorphisms in the APOE region. Genomics 63:7–12

Martin ER, Kaplan NL, Weir BS (1997) Tests for linkage and association in nuclear families. Am J Hum Genet 61:439–448

Martin ER, Lai EH, Gilbert JR, Rogala AR, Afshari AJ, Riley J, Finch KL, Stevens JF, Livak KJ, Slotterbeck BD, Slifer SH, Warren LL, Conneally PM, Schmechal DE, Purvis I, Pericak-Vance MA, Roses AD, Vance JM (2000*b*) SNPing away at complex diseases: analysis of single-nucleotide polymorphisms around APOE in Alzheimer disease. Am J Hum Genet 67:383–394

Martin ER, Monks SA, Warren LL, Kaplan NL (2000*c*) A test for linkage and association in general pedigrees: the pedigree disequilibrium test. Am J Hum Genet 67:146–154

Ott J (1989) Statistical properties of the haplotype relative risk. Genet Epidemiol 6:127–130

Pericak-Vance MA (1998) Linkage disequilibrium and allelic association. In: Haines JL, Pericak-Vance MA (eds) Approaches to gene mapping in complex human diseases. Wiley-Liss, New York, pp 323–334

Pericak-Vance MA, Bass MP, Yamaoka LH, Gaskell PC, Scott WK, Terwedo HA, Menold MM, Conneally PM, Small GW, Vance JM, Saunders AM, Roses AD, Haines JL (1997) Complete genomic screen in late-onset familial Alzheimer disease—evidence for a new locus on chromosome 12. JAMA 278:1237–1241

Rabinowitz D, Laird NM (2000) A unified approach to adjusting association tests for population admixture with arbitrary pedigree structure and arbitrary missing marker information. Hum Hered 50:211–223

Rimmler JB, Gaskell PC, Aboudonia S, Welsh-Bohmer K, Jackson CE, Schamel M, Yamaoka L, Haines JL, Pericak-Vance MA (1997) A genomic screen in extended Amish families supports a locus on chromosome 12 for Alzheimer disease (AD). Am J Hum Genet Suppl 61:A1705

Risch N, Merikangas K (1996) The future of genetic studies of complex human diseases. Science 273:1516–1519

Roberts SB, MacLean CJ, Neale MC, Eaves LJ, Kendler KS (1999) Replication of linkage studies of complex traits: an examination of variation in location estimates. Am J Hum Genet 65:876–884

Rogaeva E, Premkumar S, Grubber J, Serneels L, Scott WK, Kawari T, Song Y, et al. (1999) An α-2-macroglobulin insertion-deletion polymorphism in Alzheimer disease. Nat Genet 22:19–21

Rogaeva E, Premkumar S, Song YQ, Sorbi S, Brindle N, Paterson A, Duara R, Levesque G, Yu G, Nishimura M, Ikeda M, O'Toole C, Kawarai T, Jorge R, Vilarino D, Bruni AC, Farrer LA, St. George-Hyslop PH (1998) Evidence of an Alzheimer disease susceptibility locus on chromosome 12 and for further locus heterogeneity. JAMA 280:614–618

Romas SN, Mayeux R, Rabinowitz D, Tang MX, Zadroga HR, Lantigua R, Medrano M, Tycko B, Knowles JA (2000) The deletion polymorphism and Val1000lle in $\alpha - 2$-macroglobulin and Alzheimer disease in Caribbean Hispanics. Neurosci Lett 279:133–136

Rudrasingham V, Wavrant-DeVrieze F, Lambert JC, Chakraverty S, Kehoe P, Crook R, Amouyel P, Wu W, Rice F, Perez-Tur J, Frigard B, Morris JC, Carty S, Petersen R, Cottel D, Tunstall N, Holmans P, Lovestone S, Chartier-Harlin MC, Goate A, Hardy J, Owen MJ, Williams J (1999) Alpha-2 macroglobulin gene and Alzheimer disease. Nat Genet 22: 17–19

Schaid DJ (1996) General score tests for associations of genetic markers with disease using cases and their parents. Genet Epidemiol 13:423–449

Scott WK, Grubber JM, Abou-Donia SM, Church TD, Saunders AM, Roses AD, Pericak-Vance MA, Conneally PM, Small GW, Haines JL (1999) Further evidence linking late-onset Alzheimer disease with chromosome 12. JAMA 281: 513–514

Siegmund KD, Langholz B, Kraft P, Thomas DC (2000) Testing linkage disequilibrium in sibships. Am J Hum Genet 67: 244–248

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

Terwilliger JD, Weiss KM (1998) Linkage disequilibrium mapping of complex disease: fantasy or reality? Curr Opin Biotechnol 9:578–594

Thomson G (1995) Analysis of complex human genetic traits: an ordered-notation method and new tests for mode of inheritance. Am J Hum Genet 57:474–486

White H (1980) A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica 48:817–838

Witte JS, Gauderman WJ, Thomas DC (1999) Asymptotic bias and efficiency in case-control studies of candidate genes and gene-environment interactions: basic family designs. Am J Epidemiol 149:693–705

Woolf B (1955) On estimating the relation between blood group and disease. Ann Hum Genet 19:251–253

Wu WS, Holmans P, Wavrant-DeVrieze F, Shears S, Kehoe P, Crook R, Booth J, Williams N, Perez-Tur J, Roehl K, Fenton I, Chartier Harlin MC, Lovestone S, Williams J, Hutton M, Hardy J, Owen MJ, Goate A (1998) Genetic studies on chromosome 12 in late-onset Alzheimer disease. JAMA 280: 619–622