Am. J. Hum. Genet. 70:751–757, 2002

Report

Testing for Genetic Linkage in Families by a Variance-Components Approach in the Presence of Genomic Imprinting

Sanjay Shete and Christopher I. Amos

Department of Epidemiology, The University of Texas M. D. Anderson Cancer Center, Houston

Some genes that affect development and behavior in mammals are known to be imprinted; and ⩾1% of all mammalian genes are imprinted. Hence, incorporating an imprinting parameter into linkage analysis may increase the power to detect linkage for these traits. Here we propose theoretical justifications for a recently developed model for testing of linkage, in the presence of genetic imprinting, between a quantitative-trait locus and a polymorphic marker; this is achieved in the variance-components framework. We also incorporate sex-specific recombination fractions into this model. We discuss the effects that imprinting and nonimprinting have on the power of the usual variance-components method and on the variance-components method that incorporates an imprinting parameter. We provide noncentrality parameters that can be used to determine the sample size necessary to attain a specified power for a given significance level, which is useful in the planning of a linkage study. Optimal strategies for a genome scan of potentially imprinted traits are discussed.

Some genes that affect development and behavior in mammals are known to be imprinted. Genomic imprinting affects several human genes, including those for Prader-Willi syndrome, Angelman syndrome, Wilms tumor, and Beckwith-Wiedemann syndrome (Lalande 1997; Pfeifer 2000). Imprinting results in a higher level of expression of genes inherited from only one of the two parental chromosomes. Morison et al. (2001) have made an imprinted-gene database that contains >40 imprinted genes in humans and other organisms. For excellent reviews of mechanisms of genomic imprinting, see the work of Pfeifer (2000) and Reik and Walter (2001).

Incorporation of information on imprinting into linkage analysis may result in a more powerful test for linkage. Recently, Hanson et al. (2001) introduced a method to test for linkage in the presence of imprinting. In the present report, we give a detailed theoretical justification of the model used to test for linkage, in the presence of genetic imprinting, between a quantitative-trait locus and a polymorphic marker. Because the statistical validity of any test is important, our report is complementary to the report by Hanson et al. (2001), since we derive the tests in terms of important population parameters that allow complete evaluation of the tests' properties. We also provide further details of the modeling strategy. We include imprinting within the variance-components framework. We decompose the total additive genetic variance into parent-specific additive genetic variances and the dominance variance. Testing the equality of these two parent-specific additive components is a valid test for imprinting effect. For qualitative traits, imprinting can be tested by specifying the different penetrance parameters for heterozygotes (Strauch et al. 2000). We also have extended the model to allow for sex-specific recombination fractions. This is particularly important when imprinted genes are studied because of the differences, in some regions of the genome, between the human-male and human-female recombination rates. The exact causes for the differences are not well understood. On average, the female:male map-distance ratio is 1.6:1 (Fann and Ott 1995; Broman et al. 1998), but some regions show a much larger difference between map distance in females and that in males; hence, in the imprinting-testing model, it is important to include this difference between the male and female recombination fractions.

Received September 20, 2001; accepted for publication November 29, 2001; electronically published February 8, 2002.

Address for correspondence and reprints: Dr. Sanjay Shete, Department of Epidemiology, Box 189, The University of Texas M. D. Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, TX 77030. E-mail: sshete@mdanderson.org

0002-9297/2002/7003-0018$15.00

Following Amos (1994), let X_i be the phenotypic value for the ith individual in a pedigree:

$$X_i = \mu + g_i + G_i + \sum_{k=1}^{s} \beta_k Z_{i_k} + e_i \ , \qquad (1)$$

where μ is the overall mean, g_i is the major-gene effect, G_i is the polygenic effect, β_k's are covariate effects that are assumed to be uncorrelated with genetic and environmental factors, and e_i is the environmental effect. We write

$$g_i = \begin{cases} a & \text{if individual's genotype is } BB \\ d_1 & \text{if individual's genotype is } Bb \\ d_2 & \text{if individual's genotype is } bB \\ -a & \text{if individual's genotype is } bb \end{cases} .$$

Here, without loss of generality, we assume that the first allele is derived from the father, that the second allele is derived from the mother, that d is the dominance effect, and that I is the imprinting effect. Then, $d = (d_1 + d_2)/2$ and $I = (d_1 - d_2)/2$. When $d_1 = d_2$, there is no imprinting. When there is no dominance, $|I|$ lies within $[0,a]$, with $I = a$ representing complete imprinting. The genetic variance, σ_g^2, at this locus can be decomposed into three parts: an additive component due to the paternally derived allele, $\sigma_{a_f}^2$; an additive component due to the maternally derived allele, $\sigma_{a_m}^2$; and the usual dominance component, σ_d^2. In Appendix A, we show that $\sigma_{a_f}^2 = pq[(a + I) - d(p - q)]^2$ and $\sigma_{a_m}^2 = pq[(a - I) - d(p - q)]^2$, where p and q are the frequencies of alleles B and b, respectively.

Also, $\sigma_{a_f}^2 + \sigma_{a_m}^2 = \sigma_a^2$. When $I = 0$, $\sigma_{a_f}^2$ and $\sigma_{a_m}^2$ are equal to $\frac{1}{2}\sigma_a^2$; and, when $\sigma_{a_f}^2$ and $\sigma_{a_m}^2$ are equal, $I = 0$. Hence, a test for equality of these two parent-specific additive variances can be used to test for the presence of imprinting effect.

Let π_{tfij} and π_{tmij} be the proportion of alleles (0 or 1) at the trait locus that are identical by descent (IBD) and derived from the father and mother, respectively. Then, on the basis of equation (1), the phenotypic covariance is

$$\mathrm{Cov}[X_i,X_j|(\pi_{tfij},\pi_{tmij})] = \begin{cases} \sigma_{a_f}^2 + \sigma_{a_m}^2 + \sigma_d^2 + \sigma_G^2 + \sigma_e^2 & \text{if } i = j \\ \pi_{tfij}\sigma_{a_f}^2 + \pi_{tmij}\sigma_{a_m}^2 + \Delta_{tij}\sigma_d^2 + \phi_{ij}\sigma_G^2 & \text{if } i \neq j \end{cases} . \qquad (2)$$

where Δ_{tij} is the probability that a pair of sibs share both alleles IBD, which is known as the "coefficient of fraternity" (Lynch and Walsh 1997), and ϕ_{ij} is the coefficient of relationship, which is $\frac{1}{2}$ for sib pairs. This model will be useful to test for linkage when polymorphic markers are available either within or very near the candidate gene.

Generally, we have allele IBD information only for a marker locus. We can generalize equation (2) for a linked marker. Let π_{fij} and π_{mij} be the proportions of alleles IBD at a marker locus derived from the father and mother, respectively. To obtain covariances between relatives, we need to determine the joint probability distributions $M = (\pi_{fij},\pi_{mij})$ and $T = (\pi_{tfij},\pi_{tmij})$, which denote the allele IBD–sharing information at the marker locus and the trait locus, respectively. Let θ_f and θ_m be the recombination fractions for females and males, respectively. Note that M and T take the values (0,0),(0,1),(1,0), and (1,1), respectively. The joint probability distribution of M and T for sibs is given in Appendix B. We can write

$$\mathrm{Cov}(X_i,X_j|M) = \sum_T \mathrm{Cov}(X_i,X_j|T)P(T|M) \ ,$$

which, for sib pairs, can be written as

$$\begin{aligned} \mathrm{Cov}[X_i,X_j|(\pi_{fij},\pi_{mij})] = {} & 2\theta_f(1-\theta_f)\sigma_{a_m}^2 + 2\theta_m(1-\theta_m)\sigma_{a_f}^2 \\ & + 4\theta_m\theta_f(1-\theta_m)(1-\theta_f)\sigma_d^2 \\ & + (1-2\theta_m)^2[\sigma_{a_f}^2 + 2\theta_f(1-\theta_f)\sigma_d^2]\pi_{fij} \\ & + (1-2\theta_f)^2[\sigma_{a_m}^2 + 2\theta_m(1-\theta_m)\sigma_d^2]\pi_{mij} \\ & + (1-2\theta_m)^2(1-2\theta_f)^2\sigma_d^2\pi_{mij}\pi_{fij} \ . \end{aligned} \qquad (3)$$

When $\theta_f = \theta_m$, this simplifies to

$$\begin{aligned} \mathrm{Cov}[X_i,X_j|(\pi_{fij},\pi_{mij})] = {} & 2\theta(1-\theta)\sigma_g^2 \\ & + 2\theta(\theta-1)(1-2\theta+2\theta^2)\sigma_d^2 \\ & + (1-2\theta)^2[\sigma_{a_f}^2 + 2\theta(1-\theta)\sigma_d^2]\pi_{fij} \\ & + (1-2\theta)^2[\sigma_{a_m}^2 + 2\theta(1-\theta)\sigma_d^2]\pi_{mij} \\ & + (1-2\theta)^4\sigma_d^2\pi_{fij}\pi_{mij} \ , \end{aligned} \qquad (4)$$

where θ is the recombination fraction between the trait locus and the marker locus.

A general method for estimation of parent-specific allele sharing between sib pairs is given in Appendix C. Even though the formulae in Appendix C are valid when one or both parents' genotypes are unknown, in such cases, estimation of allele IBD sharing is less accurate. From equation (4), it can be seen that the coefficients of π_{fij} and π_{mij} are equal if and only if $\sigma_{a_f}^2$ and $\sigma_{a_m}^2$ are equal, which are equal if and only if $I = 0$ (i.e., there is no parental imprinting). Hence, the null hypothesis of no imprinting can be tested by using the likelihood-ratio test for $\sigma_{a_f}^2 = \sigma_{a_m}^2$. However, if $\theta_f \neq \theta_m$, then this test is not valid, as can be seen from equation (3). Theoretical graphs (data not shown) to assess the effect that the

difference between θ_f and θ_m has on type 1 error of the test confirm simulation-based observation of Hanson et al. (2001). We found that the test is not sensitive to modest difference between θ_f and θ_m (i.e., female:male map-distance ratio $\leq$5:1). In a genome scan, one ordinarily will first test the joint hypotheses of no linkage and no imprinting, by testing $\sigma^2_{a_f} = \sigma^2_{a_m} = 0$.

All the unknown parameters can be estimated by use of the maximum-likelihood method if the data are approximately normally distributed. We implemented the maximum-likelihood–estimation method by using MAXFUN (SAGE 1997) to estimate the unknown parameters. The likelihood-ratio–test statistic can be converted to a LOD score by dividing the likelihood ratio by $2 \ln_e(10)$. The test $H_0: \sigma^2_{a_f} = \sigma^2_{a_m} = 0$ was called "LOD_{imp}"—and the usual test, $H_0: \sigma^2_a = 0$, was called "LOD_{eq}"—by Hanson et al. (2001). Our simulation results (data not shown) suggested that the LOD_{imp} test was more powerful than the LOD_{eq} test, when we had an imprinting effect. This is expected, because, in the LOD_{imp} test, we incorporated the imprinting parameter into the analysis; however, when there was no imprinting, the LOD_{eq} test was more powerful than the LOD_{imp} test. Although both tests can be converted to LOD scores, the LOD_{imp} test requires more degrees of freedom than the LOD_{eq} test does. The simulation results are verified analytically by computing the noncentrality parameter of the χ^2 test statistic and thereby obtaining the sample size required in order to attain a certain power for a given significance level.

The likelihood-ratio test is twice the log-likelihood difference between (*a*) a model in which $\sigma^2_{a_f}$ and $\sigma^2_{a_m}$ are free to vary and at least one of the two is positive and (*b*) a model in which both of these parameters are set to zero. The test statistic $2(\ln L_1 - \ln L_0)$ is asymptotically a mixture of 0, χ^2_1, and χ^2_2, in the proportions $\frac{1}{4}$, $\frac{1}{2}$, and $\frac{1}{4}$, respectively, under the null hypothesis (Self and Liang 1987). Under the alternative hypothesis, the distribution of the likelihood-ratio test is a noncentral χ^2. The noncentrality parameter and the degrees of freedom are essential ingredients for calculating both the sample size required for given expected power and the chosen critical *P* value. The noncentrality parameter is twice the difference between the expected log likelihoods under the alternate and null hypotheses, evaluated at their respective asymptotic parameter estimates (Williams and Blangero 1999; Sham et al. 2000).

From equation (2), it can be seen that, under the null hypothesis, the asymptotic parameters are

$$\left[\sum_N\right] = \begin{cases} \sigma^2_{a_f} + \sigma^2_{a_m} + \sigma^2_d + \sigma^2_G + \sigma^2_e & \text{if } i = j \\ \frac{1}{2}\sigma^2_{a_f} + \frac{1}{2}\sigma^2_{a_m} + \frac{1}{4}\sigma^2_d + \phi_{ij}\sigma^2_G & \text{if } i \neq j \end{cases},$$

and that, under the alternative hypothesis of linkage, the asymptotic parameters are given by equation (4). Let L_N and L_A be the likelihoods evaluated at the asymptotic parameter values under the null and alternative hypotheses, respectively. Then, for a sibship of size *s*, $E(2 \ln L_N) = -\ln|\Sigma_N| - s$ and $E(2 \ln L_A) = -\sum p_i \ln|\Sigma_i| - s$, where the summation is over all possible marker-genotype configurations and p_i is the probability of the *i*th configuration (Williams and Blangero 1999; Sham et al. 2000).

For sib pairs for whom there is complete linkage information, these covariance matrices have diagonal elements 1, because we fixed the total phenotypic variance to 1, and have off-diagonal elements given by

$$\sum_{M=(0,0)} = k_1 = \frac{1}{2}\sigma^2_G ,$$

$$\sum_{M=(1,0)} = k_2 = \frac{1}{2}\sigma^2_G + \sigma^2_{a_f} ,$$

$$\sum_{M=(0,1)} = k_3 = \frac{1}{2}\sigma^2_G + \sigma^2_{a_m} ,$$

$$\sum_{M=(1,1)} = k_4 = \frac{1}{2}\sigma^2_G + \sigma^2_{a_f} + \sigma^2_{a_m} + \sigma^2_d ,$$

with probability $p_i = \frac{1}{4}$ each. Under the null hypothesis, these off-diagonal elements are given by

$$k_0 = \frac{1}{2}\sigma^2_G + \frac{1}{2}\sigma^2_{a_f} + \frac{1}{2}\sigma^2_{a_m} + \frac{1}{4}\sigma^2_d .$$

Then, the noncentrality parameter per sib pair is

$$\lambda = -\frac{1}{4}\sum_{i=1}^{4} \ln(1 - k_i^2) + \ln(1 - k_0^2) .$$

We typically perform linkage analysis at a linked marker locus that is not a trait locus. In table 1, we give joint probability distribution of IBD-sharing information for the marker and the loci. The conditional sib-pair correlations in trait values can be obtained from the joint distribution in table 1, and, when $\theta_f = \theta_m$, they are given by

$$c_1 = \psi^2 k_1 + \psi(1 - \psi)(k_2 + k_3) + (1 - \psi)^2 k_4 ,$$

$$c_2 = \psi(1 - \psi)(k_1 + k_4) + \psi^2 k_2 + (1 - \psi)^2 k_3 ,$$

$$c_3 = \psi(1 - \psi)(k_1 + k_4) + (1 - \psi)^2 k_2 + \psi^2 k_3 ,$$

$$c_4 = (1 - \psi)^2 k_1 + \psi(1 - \psi)(k_2 + k_3) + \psi^2 k_4 .$$

Table 1

Joint Probability Distribution of *M* and *T*

	Joint Probability Distribution When *M* Is			
T	(0,0)	(0,1)	(1,0)	(1,1)
$\theta_f = \theta_m$:				
(0,0)	$\psi^2/4$	$\psi(1-\psi)/4$	$\psi(1-\psi)/4$	$(1-\psi)^2/4$
(0,1)	$\psi(1-\psi)/4$	$\psi^2/4$	$(1-\psi)^2/4$	$\psi(1-\psi)/4$
(1,0)	$\psi(1-\psi)/4$	$(1-\psi)^2/4$	$\psi^2/4$	$\psi(1-\psi)/4$
(1,1)	$(1-\psi)^2/4$	$\psi(1-\psi)/4$	$\psi(1-\psi)/4$	$\psi^2/4$
$\theta_m \neq \theta_f$:				
(0,0)	$\psi_m\psi_f/4$	$\psi_m(1-\psi_f)/4$	$\psi_f(1-\psi_m)/4$	$(1-\psi_m)(1-\psi_f)/4$
(0,1)	$\psi_m(1-\psi_f)/4$	$\psi_m\psi_f/4$	$(1-\psi_m)(1-\psi_f)/4$	$\psi_f(1-\psi_m)/4$
(1,0)	$\psi_f(1-\psi_m)/4$	$(1-\psi_m)(1-\psi_f)/4$	$\psi_m\psi_f/4$	$\psi_m(1-\psi_f)/4$
(1,1)	$(1-\psi_m)(1-\psi_f)/4$	$\psi_f(1-\psi_m)/4$	$\psi_m(1-\psi_f)/4$	$\psi_m\psi_f/4$

Then, the noncentrality parameter of the linkage test, per sib pair, is

$$\lambda_I = -\frac{1}{4}\sum_{i=1}^{4}\ln(1-c_i^2) + \ln(1-k_0^2) \ .$$

When $\theta_f \neq \theta_m$, these conditional sib-pair correlations in trait values are given by

$$\begin{aligned}
cs_1 &= \psi_m\psi_f k_1 + \psi_m(1-\psi_f)k_2 + \psi_f(1-\psi_m)k_3 \\
&\quad +(1-\psi_m)(1-\psi_f)k_4 \ , \\
cs_2 &= \psi_m(1-\psi_f)k_1 + \psi_m\psi_f k_2 + (1-\psi_m)(1-\psi_f)k_3 \\
&\quad +\psi_f(1-\psi_m)k_4 \ , \\
cs_3 &= \psi_f(1-\psi_m)k_1 + (1-\psi_m)(1-\psi_f)k_2 + \psi_m\psi_f k_3 \\
&\quad +\psi_m(1-\psi_f)k_4 \ , \\
cs_4 &= (1-\psi_m)(1-\psi_f)k_1 + \psi_f(1-\psi_m)k_2 \\
&\quad +\psi_m(1-\psi_f)k_3 + \psi_m\psi_f k_4 \ .
\end{aligned}$$

Then, the noncentrality parameter of the linkage test allowing for difference between θ_f and θ_m, per sib pair, is

$$\lambda_{I_s} = -\frac{1}{4}\sum_{i=1}^{4}\ln(1-cs_i^2) + \ln(1-k_0^2) \ .$$

The noncentrality parameter for the test of linkage, for the usual variance-components model (i.e., that which does not incorporate the imprinting parameter), can be obtained similarly and is

$$\begin{aligned}
\lambda_U = &-\frac{1}{4}\ln(1-r_0^2) - \frac{1}{2}\ln(1-r_1^2) \\
&-\frac{1}{4}\ln(1-r_2^2) + \ln(1-k_0^2) \ ,
\end{aligned}$$

where $r_0 = \frac{1}{2}\sigma_G^2$, $r_1 = \frac{1}{2}\sigma_G^2 + \frac{1}{2}\sigma_a^2$, and $r_2 = \frac{1}{2}\sigma_G^2 + \sigma_a^2 + \sigma_d^2$.

From these noncentrality parameters, we can obtain the sample size required in order to have a certain power and significance level. For linkage analysis, we usually use .0001 as the significance level, which is approximately equal to a LOD score of 3. The noncentrality parameter required in order to have 80% power was 20.8 for the usual variance-components model and was 23.55 for the variance-components model that incorporates the imprinting parameter. The required number of sib pairs can be obtained by dividing these required noncentrality parameters by the theoretical noncentrality parameters obtained above. In table 2, we compare the sample sizes required in order to have 80% power to detect linkage at a significance level of .0001, for various values of *I*. The numbers in parenthesis are for the usual variance-components model. The dominance variance was assumed to be zero (by choosing $d_1 = -d_2$). We chose a value of $a = 1$ and an allele frequency for p equal to half, which gave $\sigma_{a_f}^2 = \frac{1}{4}(1+I)^2$ and $\sigma_{a_m}^2 = \frac{1}{4}(1-I)^2$. It can be seen from table 2 that, when $I = 0$, the required sample size was smaller for the usual variance-components model than for the variance-components model that incorporates the imprinting parameter. The imprinting model became more powerful only when the imprinting effect was moderate to large. Because, as far as we know now, only ~1% of

Table 2

Sample Sizes for 80% Power to Detect Linkage at a LOD Score of 3, for $a = 1$, with LOD_{imp} and LOD_{eq}

	$LOD_{imp}(LOD_{eq})$ When $I =$					
θ	0	.05	.1	.2	.4	.6
.0	453 (400)	445 (398)	425 (389)	356 (358)	201 (257)	95 (146)
.01	495 (437)	487 (434)	464 (425)	389 (391)	220 (281)	105 (161)
.05	710 (627)	698 (622)	666 (610)	558 (562)	318 (409)	157 (242)
.1	1,158 (1,022)	1,139 (1,015)	1,086 (995)	910 (919)	522 (674)	263 (410)
.2	3,732 (3,296)	3,671 (3,274)	3,499 (3,211)	2,933 (2,969)	1,692 (2,200)	874 (1,370)

genes are imprinted, one should usually perform genome scans with the usual variance-components method and test for imprinting only if significant evidence for linkage is observed. However, this may reduce the power when genes are completely imprinted. As an exceptional situation, analysis of traits related to development may be more powerful when an imprinting parameter is included. The significance level should be appropriately adjusted for multiple testing. Both a version of the MULTIC program of the software ACT that incorporates imprinting and the program used to compute the sample size required in order to achieve a specific power can be obtained, on request, from the authors.

Acknowledgments

We thank the two anonymous reviewers for their helpful constructive comments. We also thank Dr. Maureen Goode for comments that led to a better presentation of the material in this report. This work is supported by National Institutes of Health grants HG 02275 and ES 09912.

Appendix A

Partitioning of Genetic Variance

Table A1

Genotype-Allele–Effect Distribution

Genotype	g	X_f	X_m	Frequency
BB	a	1	1	p^2
Bb	$d + I$	1	0	pq
bB	$d - I$	0	1	pq
bb	$-a$	0	0	$(1 - p)^2$

Let g be the genetic effect defined in the text, and let X_f and X_m be the gene-content values from the father and mother, respectively. X_f is equal to 1 if the disease allele B is inherited from the father and is equal to 0 otherwise; X_m is defined similarly. Let p and q be the frequencies of alleles B and b, respectively. g, X_f, and X_m for each genotype are shown in table A1.

For the genetic effect, consider fitting a multiple linear-regression equation onto X_f and X_m: $g = \beta_0 + \beta_1 X_f + \beta_2 X_m + e$, where e is the error term.

On the basis of regression theory and from table A1, one can show that

$$\hat{\beta}_1 = \frac{\text{Cov}(g, X_f)}{\text{Var}(X_f)} = (a + I) - d(p - q) ,$$

$$\hat{\beta}_2 = \frac{\text{Cov}(g, X_m)}{\text{Var}(X_m)} = (a - I) - d(p - q) .$$

We also know that $\sigma_g^2 = \sigma_a^2 + \sigma_d^2$, where

$$\begin{aligned}\sigma_a^2 &= \text{Var}(\hat{\beta}_0 + \hat{\beta}_1 X_f + \hat{\beta}_2 X_m) \\ &= \hat{\beta}_1^2 \text{Var}(X_f) + \hat{\beta}_2^2 \text{Var}(X_m) \\ &= \{pq[(a + I) - d(p - q)]^2\} \\ &\quad + \{pq[(a - I) - d(p - q)]^2\} \\ &= \sigma_{a_f}^2 + \sigma_{a_m}^2\end{aligned}$$

and $\sigma_d^2 = 4p^2q^2d^2$. So, we partitioned the total genetic variance into three parts: $\sigma_{a_f}^2$, the additive genetic component due to the paternally derived allele; $\sigma_{a_m}^2$, the additive genetic component due to the maternally derived allele; and σ_d^2, the dominance genetic component.

Appendix B

Joint Probability Distribution of *M* and *T*

We obtained the joint probability distribution of M and T by a method similar to that described in the work of Haseman and Elston (1972, Appendix B). Consider the mating A_1B_1/A_2B_2 and A_3B_3/A_4B_4. Let θ be the recom-

bination fraction between the *A* and *B* loci. There are 16 possible genotypes for a child with this parental mating. If two offspring of these parents have the genotypes A_1B_1/A_3B_3 and A_1B_1/A_4B_3, then, at the marker locus, *B*, sibs share both alleles IBD but, at the trait locus, *A*, share only the paternal allele IBD. The probability of such sib types is $[(1-\theta)^2/4][(1-\theta)/2](\theta/2)$. There are eight possible sib genotypes with the same probability. Also, if the sibs genotypes are A_1B_2/A_3B_3 and A_1B_2/A_4B_3, then, again, both alleles in the sibs are IBD at the marker locus, but only the paternal allele is IBD at the trait locus. The probability of such genotypes is $(\theta/2)[(1-\theta)/2](\theta^2/4)$. There are eight possible sib genotypes with the same probability. Hence,

$$P[T = (1,0), M = (1,1)] = 8\left[\frac{(1-\theta)^2}{4}\right]\left[\frac{(1-\theta)}{2}\right]\frac{\theta}{2} + 8\frac{\theta}{2}\left[\frac{(1-\theta)}{2}\right]\frac{\theta^2}{4} = \frac{\psi(1-\psi)}{4} ,$$

where $\psi = \theta^2 + (1-\theta)^2$. Other probabilities can be obtained similarly and are given in table 1. If θ_f and θ_m are the sex-specific recombination fractions for females and males, respectively, then the joint probability distribution of *M* and *T* can be obtained by allowing for the differences between θ_f and θ_m. The probability of sibs having two alleles IBD at the marker locus and having a paternal allele at the trait locus is

$$P[T = (1,0), M = (1,1)] = 8\left[\frac{(1-\theta_m)^2}{4}\right]\left[\frac{(1-\theta_f)}{2}\right]\frac{\theta_f}{2} + 8\frac{\theta_f}{2}\left[\frac{(1-\theta_f)}{2}\right]\frac{\theta_m^2}{4} = \frac{\psi_m(1-\psi_f)}{4} ,$$

where $\psi_m = \theta_m^2 + (1-\theta_m)^2$ and $\psi_f = \theta_f^2 + (1-\theta_f)^2$. The remaining probabilities can be obtained similarly and are shown in table 1.

Appendix C

Computation of Parent-Specific IBD

Let x_0 and y_0 be the marker phenotypes of parents, and let X_i, $i = 1, \ldots, p$, be the marker phenotypes of *p* sibs. Let $g_{ab}(x) = P(\text{observing phenotype } x \mid \text{genotype is } ab)$. The population frequency of genotype *ab* is denoted by ψ_{ab}. The "phenoset" corresponding to a phenotype is the set of all genotypes that could give rise to that phenotype. Let *C* and *D* be the phenosets of the first and second parents, respectively. Let *rs* be an element of *C*, and let *vw* be an element of *D*.

Then, the likelihood of a nuclear family is

$$L(\text{family data}) = \sum_{rs}\psi_{rs}g_{rs}(x_0)\sum_{vw}\psi_{vw}g_{vw}(y_0)\prod_{j=1}^{p}\frac{1}{4}[g_{rv}(x_j) + g_{rw}(x_j) + g_{sw}(x_j) + g_{sv}(x_j)] .$$

Similarly, the joint likelihood of the nuclear family and sibs 1 and 2 having one allele IBD from the father is

$$\begin{aligned} &L(\text{family data and sibs having one allele IBD from the father}) \\ &= \sum_{rs}\psi_{rs}g_{rs}(x_0)\sum_{vw}\psi_{vw}g_{vw}(y_0)\prod_{j=3}^{p}\frac{1}{4}[g_{rv}(x_j) + g_{rw}(x_j) + g_{sw}(x_j) + g_{sv}(x_j)] \\ &\times\frac{1}{16}[g_{rv}(x_1)g_{rw}(x_2) + g_{rw}(x_1)g_{rv}(x_2) + g_{sv}(x_1)g_{sw}(x_2) + g_{sw}(x_1)g_{sv}(x_2) \\ &+g_{rv}(x_1)g_{rv}(x_2) + g_{rw}(x_1)g_{rw}(x_2) + g_{sw}(x_1)g_{sw}(x_2) + g_{sv}(x_1)g_{sv}(x_2)] . \end{aligned}$$

Let π_f be the proportion of paternally derived alleles shared IBD by the sibs. Then

$$P(\pi_f = 1 \mid \text{family data}) = \frac{L\ (\text{family data and sibs sharing one allele IBD from father})}{L\ (\text{family data})} .$$

The proportion of maternal alleles shared IBD by sibs can be obtained similarly.

References

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

Broman KW, Murray JC, Sheffield VC, White RL, Weber JL (1998) Comprehensive human genetic maps: individual and sex-specific variation in recombination. Am J Hum Genet 63:861–869

Fann CS, Ott J (1995) Parsimonious estimation of sex-specific map distances by stepwise maximum likelihood regression. Genomics 29:571–575

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

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

Lalande M (1996) Parental imprinting and human disease. Annu Rev Genet 30:173–195

Lynch M, Walsh B (1997) Genetics and analysis of quantitative traits. Sinauer Associates, Sunderland, MA

Morison IM, Paton CJ, Cleverley SD (2001) The imprinted gene and parent-of-origin effect database. Nucleic Acids Res 29:275–276

Pfeifer K (2000) Mechanisms of genomic imprinting. Am J Hum Genet 67:777–787

Reik W, Walter J (2001) Genomic imprinting: parental influence on the genome. Nat Rev Genet 2:21–32

SAGE (1997) Statistical analysis for genetic epidemiology, release 3.1. Computer program package available from the Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland

Self SG, Liang K-Y (1987) Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. J Am Stat Assoc 82(398): 605–610

Sham PC, Cherny SS, Purcell S, Hewitt JK (2000) Power of linkage versus association analysis of quantitative traits, by use of variance-component models, for sibship data. Am J Hum Genet 66:1616–1630

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

Williams JT, Blangero J (1999) Power of variance component linkage analysis to detect quantitative trait loci. Ann Hum Genet 63:545–563