

Report

Genomic Imprinting and Linkage Test for Quantitative-Trait Loci in Extended Pedigrees

Sanjay Shete, Xiaojun Zhou, and Christopher I. Amos

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

Genomic imprinting is a mechanism in which only one of the two copies of a gene is expressed. Some genes that affect development and behavior in mammals are known to be imprinted. Deregulation of imprinted genes has been found in a number of human diseases. Incorporating imprinting information into linkage analysis results in a more powerful test for linkage. Here, we propose an efficient method to test for linkage and imprinting of quantitative traits in extended pedigrees. We compared the results obtained by using the extended-pedigree-analysis approach proposed in this study with other existing approaches. We found that the proposed method is more powerful and uses extended-pedigree information most efficiently.

Genomic imprinting is a mechanism in which only one copy of the gene pair is expressed and expression is determined by the parental origin of the copy. Several imprinted genes have been identified in humans. Many imprinted genes contribute to growth. In mammals, imprinted genes have evolved over time to fine-tune the growth of the fetus. It has been hypothesized that paternally expressed genes generally enhance growth, whereas maternally expressed genes appear to suppress growth (Reik and Walter 2001).

Deregulation of imprinted genes has been found in a number of human diseases, including insulin-like growth factors in Beckwith-Wiedemann syndrome and growth inhibitors, such as the *GRB10* gene in Russell-Silver, Prader-Willi, and Angelman syndromes and Albright hereditary osteodystrophy. In addition, normal and abnormal genomic imprinting contributes to a wide range of malignancies. Morison et al. (2001) created the Imprinted Gene Catalogue, a database of >150 imprinted genes. Imprinted diseases are characterized by complex patterns of mutations and associated phenotypes that affect prenatal and postnatal growth and neurological functions.

Received April 18, 2003; accepted for publication July 7, 2003; electronically published September 16, 2003.

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

© 2003 by The American Society of Human Genetics. All rights reserved. 0002-9297/2003/7304-0019\$15.00

Expression of imprinted genes is regulated by allele-specific epigenetic modifications of DNA and chromatin. These modifications affect central regulatory elements that control allele-specific expression of several neighboring genes. Imprinted genes have had a strong impact on biomedical research and have provided interesting models for studying the mechanisms and effects of epigenetic gene control (Walter and Paulsen 2003). (For reviews of genomic imprinting mechanisms, see Pfeifer [2000], Reik and Walter [2001], and Li [2002]).

Incorporating imprinting information into linkage analysis can result in a more powerful test for linkage (Hanson et al. 2001; Shete and Amos 2002). Methods for detection of linkage and imprinting in sibship data have been developed recently (Strauch et al. 2000; Hanson et al. 2001; Shete and Amos 2002). However, dividing larger pedigrees into sibships generally results in a loss of power to detect linkage (Wijsman and Amos 1997). A method to assess imprinting for affected relative pair has been described in Karason et al (2003). Here, we report a test for linkage and imprinting in extended pedigrees for quantitative traits. We follow the general framework of Shete and Amos (2002). Let X_i be the phenotypic value for the i th individual in a pedigree. We can write

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

where μ is the overall mean, g_i is the major gene effect,

G_i is the polygenic effect, values of β_k are covariate effects that are assumed not to be correlated 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, the first allele is derived from the father, and the second allele is derived from the mother. If d is the dominance effect and I the imprinting effect, then $d = (d_1 + d_2)/2$ and $I = (d_1 - d_2)/2$. When $d_1 = d_2$, then there is no imprinting.

The genetic variance, σ_g^2 , at this locus can be divided into three parts: an additive component from the paternally derived allele, $\sigma_{a_f}^2$; an additive component from the maternally derived allele, $\sigma_{a_m}^2$; and the usual dominance component, σ_d^2 . These components are equal to

$$\sigma_{a_f}^2 = pq[(a + I) - d(p - q)]^2$$

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 (Shete and Amos 2002). Also, $\sigma_{a_f}^2 + \sigma_{a_m}^2 = \sigma_a^2$.

When the imprinting coefficient I is zero, then $\sigma_{a_f}^2$ 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, then I is zero. Hence, a test for equality of these two parent-specific additive variances can be used to test for imprinting.

We define “parent-specific identical by descent (IBD) sharing between a pair of relatives i and j ” as follows:

$$\pi_{ff,ij} = \begin{cases} 1 & \text{if } i \text{ and } j \text{ share} \\ & \text{an allele ibd that is} \\ & \text{derived from their fathers} \\ 0 & \text{otherwise} , \end{cases}$$

and

$$\pi_{fm,ij} = \begin{cases} 1 & \text{if } i \text{ and } j \text{ share an} \\ & \text{allele ibd that is derived} \\ & \text{from the father of } i \\ & \text{and the mother of } j \\ 0 & \text{otherwise} . \end{cases}$$

We define “ $\pi_{mf,ij}$ ” and “ $\pi_{mm,ij}$ ” similarly.

Then, from formula (1), the phenotypic covariance is $Cov[X_i, X_j | (\pi_{ff,ij}, \pi_{mm,ij}, \pi_{fm,ij}, \pi_{mf,ij})]$

$$= \begin{cases} \sigma_{a_f}^2 + \sigma_{a_m}^2 + \sigma_d^2 + \sigma_G^2 + \sigma_e^2 & \text{if } i = j \\ \pi_{ff,ij}\sigma_{a_f}^2 + \pi_{mm,ij}\sigma_{a_m}^2 \\ \quad + (\pi_{fm,ij} + \pi_{mf,ij})(\sigma_a^2/2 - 2pqI^2) & \\ \quad + \Delta_{ij}\sigma_d^2 + \phi_{ij}\sigma_G^2 & \text{if } i \neq j \end{cases} , \quad (2)$$

where $\sigma_{a_f}^2$, $\sigma_{a_m}^2$, and σ_d^2 are as defined above; σ_G^2 and σ_e^2 are variances owing to the polygenic component and environmental component, respectively; Δ_{ij} is the probability that a pair of relatives share both alleles IBD, known as “the coefficient of fraternity” (Lynch and Walsh 1997); and ϕ_{ij} is the coefficient of relationship. This model will be useful for testing linkage when polymorphic markers are available within or very near the candidate gene. Although the model is developed for a two-allele system, this model can still be fitted to the multiallelic case. In this instance, the parameters will reflect average effects of the dominance and imprinting from the alleles.

From equation (2), it can be seen that the coefficients of $\pi_{ff,ij}$, $\pi_{mm,ij}$, and $(\pi_{fm,ij} + \pi_{mf,ij})$ are equal if and only if $\sigma_{a_f}^2$ and $\sigma_{a_m}^2$ are equal, and $\sigma_{a_f}^2$ and $\sigma_{a_m}^2$ are equal if and only if the imprinting parameter I is zero (i.e., there is no parental imprinting). Hence, the null hypothesis of no imprinting can be tested by using the likelihood-ratio test for equality of these coefficients. Ordinarily, in a genome scan, one will test the joint null hypotheses of no linkage and no imprinting by testing $\sigma_{a_f}^2 = \sigma_{a_m}^2 = 0$. If the data consist solely of sibships, then the coefficient of $(\pi_{fm,ij} + \pi_{mf,ij})$ is zero. To understand when the gain will be maximal for testing linkage in the presence of imprinting with the proposed model in formula (2) compared with the model that is based on sibship, we plotted values of $(\sigma_a^2/2 - 2pqI^2)$ for the range of values of imprinting (I) and dominance (d) in figure 1. From figure 1, we observe that, for testing for linkage, the gain in power is higher when there is either moderate-to-high dominance or moderate-to-low imprinting. The power for linkage detection decreases with increasing values of imprinting and decreasing values of dominance with no gain when imprinting is complete and there is no dominance (*lower left corner of the figure*). The part of the gain in the test for imprinting depends on the value of $2pqI$. Increasing values of I give larger deviations, as shown in figure 1, by decreasing values of $(\sigma_a^2/2 - 2pqI^2)$. There will also be contributions from bilinear pairs.

Multipoint parent-specific identity by descent is computed as follows: There are four alleles at a single locus for the relative pair i and j . Let us denote the two alleles for individual i by a vector (i_m, i_f) , where i_m and i_f are maternal and paternal alleles, respectively. Similarly, we define the vector (j_m, j_f) for individual j . There are 15 possible ordered states of identity by descent between these two individuals (Lynch and Walsh 1997). Of these 15

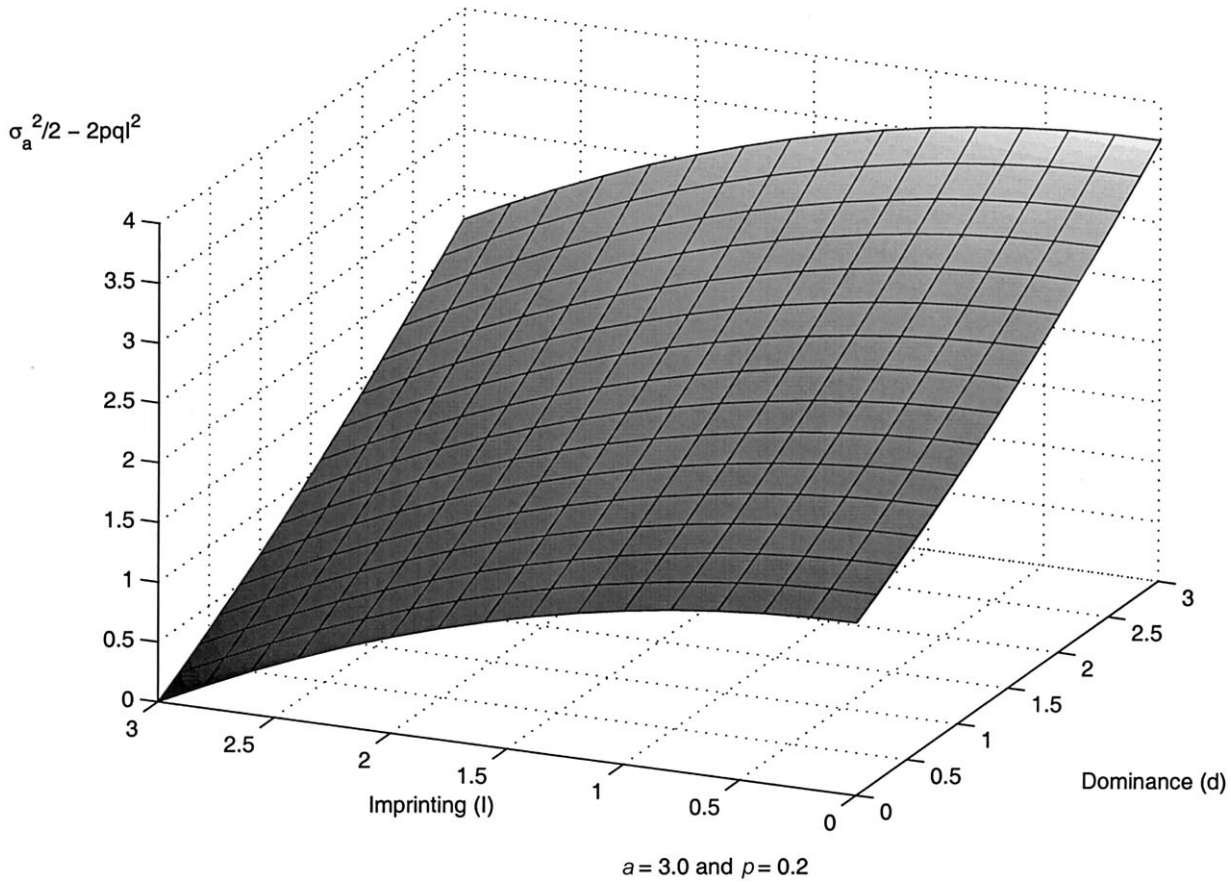


Figure 1 The region in which the new model will be more powerful for testing for linkage in the presence of imprinting, in comparison with the sibship-based model. The coefficient ($\sigma_a^2/2 - 2pql^2$) of $(\pi_{fm,ij} + \pi_{mf,ij})$ is plotted against different values of imprinting and dominance.

states, only 7 are essential for computation of identity-by-descent sharing in outbred populations. Using the notations of SimWalk2 (Weeks et al. 1995; Sobel and Lange 1996), we define probabilities of these seven states as

$$\begin{aligned}
 S_9 &= (i_m, j_m)(i_f, j_f) , \\
 S_{10} &= (i_m, j_m)(i_f)(j_f) , \\
 S_{11} &= (i_m)(i_f, j_f)(j_m) , \\
 S_{12} &= (i_m, j_f)(i_f, j_m) , \\
 S_{13} &= (i_m, j_f)(i_f)(j_m) , \\
 S_{14} &= (i_m)(i_f, j_m)(j_f) ,
 \end{aligned}$$

and

$$S_{15} = (i_m)(i_f)(j_m)(j_f) .$$

In these states, the pairs of alleles inside each pair of parentheses are IBD. For example, in S_9 , the maternal alleles and paternal alleles of i and j are IBD, and in S_{10} , the maternal allele of i is IBD with the maternal allele of j , and the paternal alleles are not IBD. The state S_{12} occurs in outbred populations for some double cousins, such as double first cousins related through opposite-sex parents. On the basis of these seven states, we can calculate the parent-specific identity by descent as

$$\begin{aligned}
 \pi_{ff,ij} &= S_{11} + S_9 , \\
 \pi_{mm,ij} &= S_{10} + S_9 , \\
 \pi_{fm,ij} + \pi_{mf,ij} &= S_{13} + S_{14} + 2 * S_{12} ,
 \end{aligned}$$

and

$$\Delta_{ij} = S_9 + S_{12} .$$

Table 1
Power of Different Tests of Linkage and Imprinting

	Usual VC Method ^a	Sibship-Imprinting Method ^b	Pedigree-Imprinting Method ^c
Average log likelihood	-278.13	-271.56	-237.85
$2(\ln L_1 - \ln L_0)$	3.66	6.28	8.36
<i>P</i>	.0279	.0169	.0057
Power	.65	.83	.93

^a The usual variance components model, which does not allow for imprinting.

^b The method of testing for linkage and imprinting that is based on sibship data—setting the coefficient of $(\pi_{fm,ij} + \pi_{mf,ij})$ to zero in equation (2).

^c The proposed method of testing for linkage and imprinting that uses extended pedigree most effectively.

The identity states S_9 – S_{15} can be calculated by using the software program SimWalk2 (Weeks et al. 1995; Sobel and Lange 1996).

All the unknown parameters can be estimated by using the maximum-likelihood method, under the assumption that the data are approximately multivariate normally distributed. We do not estimate parameters p and I from equation (2). We estimate three parameters from equation (2)—namely, σ_{af}^2 , σ_{am}^2 , and $(\sigma_a^2/2 - 2pqI^2)$ —treating the coefficient of $(\pi_{fm,ij} + \pi_{mf,ij})$ in equation (2) as a single unknown quantity. When the data are not normally distributed, trait-value winsorization or transformation-type techniques can be used (Etzel et al. 2003; Shete et al., in press). We implemented the maximum-likelihood-estimation method by using MAXFUN (Sorant and Elston 1994) to estimate the unknown parameters. To test $H_0: \sigma_{af}^2 = \sigma_{am}^2 = 0$, we used the likelihood-ratio test. The likelihood-ratio test is twice the difference in log likelihood between a model in which σ_{af}^2 and σ_{am}^2 are free to vary and at least one of them is positive and a model in which σ_{af}^2 and σ_{am}^2 are set to zero. The test statistic $LR_{imp} = 2(\ln L_1 - \ln L_0)$ is asymptotically a mixture of 0, χ_1^2 , and χ_2^2 in the proportions 1/4, 1/2, and 1/4, respectively, under the null hypothesis (Self and Liang 1987). Although we use the additional parameter $(\sigma_a^2/2 - 2pqI^2)$ for the coefficient of $(\pi_{fm,ij} + \pi_{mf,ij})$ in equation (2), this parameter is highly correlated with other parameters— σ_{af}^2 and σ_{am}^2 —in the model. A regression of σ_{af}^2 and σ_{am}^2 onto $(\sigma_a^2/2 - 2pqI^2)$ over the parameter space cuboid defined by all values of a and p between 0 and 1 and of d and I from -1 to 2 gave an R^2 value >0.95 . Hence, we do not need to include a degree of freedom for this parameter. From the simulation study below, we found that a mixture of 0, χ_1^2 , and χ_2^2 in the proportions 1/4, 1/2, and 1/4, respectively, under the null hypothesis is a good approximation for controlling the type 1 error at a nominal significance level. Under the alternative hypothesis, the distribution of the likelihood-ratio test is a noncentral χ^2 . The usual test, $H_0: \sigma_a^2 = 0$, is also tested by using likelihood-ratio test between a model in

which σ_a^2 is free to vary and a model in which it is fixed to zero. The test statistics $LR_U = 2(\ln L_1 - \ln L_0)$ is asymptotically a mixture of 0 and χ_1^2 in the proportions 1/2 and 1/2.

We performed simulations to test the validity of the proposed test. We generated a large 3-generation pedigree with 40 individuals (fig. A [online only]). The first-generation parents had three sons and three daughters. Each of these children subsequently had four or five children. We generated 500 replicate samples, using the software SLINK (Ott 1989; Weeks et al. 1990). A trait locus with two alleles was simulated with the frequency of the deleterious allele equal to 0.20. The trait value was simulated using a normal distribution with means depending on genotypes at the trait locus (see definition of g_i above) and residual variance of 1. We used $a = 3$, $I = 2$, and “no dominance” in the definition of g_i . We simulated an unlinked marker with two equiproportional alleles as a “polygene.” A value at the polygene was simulated by using a normal distribution, with means depending on the genotypes of the polygene (-1 and $+1$ for the two homozygous genotypes and 0 for the heterozygous genotype) and a variance of 1. Residual environmental values were simulated from the standard normal distribution. The trait locus, polygene, and environmental values were added to obtain an individual’s phenotype. A completely linked marker with 10 equiproportional alleles was simulated to test for linkage and imprinting. An unlinked marker with 10 equiproportional alleles was simulated for type 1 error-assessment.

When we fix the coefficient of $(\pi_{fm,ij} + \pi_{mf,ij})$ to zero, then the model in equation (2) reduces to the one proposed for sibship data (Hanson et al. 2001; Shete and Amos 2002). The same model can be used for testing of imprinting and linkage in the extended pedigree (termed “sibship-imprinting method” in table 1) but is less efficient because it does not use all the information. We compared the proposed imprinting method (termed “pedigree-imprinting method” in table 1), which uses the extended-pedigree information most effectively, with the usual vari-

ance-components model (termed “usual VC method” in table 1), which does not allow for imprinting, and with the sibship-imprinting method. The results of the simulation are presented in table 1. The proposed method was significantly more powerful than the other methods. The P value corresponding to average test statistics that were based on the proposed method was .0057, compared with P values of .0169 and .0279 for the sibship-imprinting model and the usual VC method, respectively. We calculated the power reported in table 1 as the proportion of replicates in which the null hypothesis is rejected at $\alpha = 0.0001$. The type 1 errors of all three methods were well controlled at the nominal level of 5% and 1% significance. Of 500 replicate samples, we found 25 and 4 false positives in the usual VC method and 30 and 5 false positives in both the sibship-imprinting method and the pedigree-imprinting method, at 5% and 1% levels of significance, respectively.

The power of the proposed extended-pedigree-based imprinting method was better than that of the usual variance-components test and the sibship-imprinting method, because, in our test, we incorporated the imprinting parameter in the analysis of the extended pedigree most efficiently. However, when there is no imprinting, the usual variance-components test will be more powerful (Hanson et al. 2001; Shete and Amos 2002). While studying imprinted genes, it is important to use sex-specific recombination fractions because of differences in male and female recombination fractions in some regions of the genome. However, it has been shown that these tests are not sensitive to modest differences between male and female recombination fractions (Hanson et al. 2001; Shete and Amos 2002). A version of the MULTIC program of the software ACT, which incorporates imprinting and the program used to compute multipoint parent-specific identity-by-descent sharing from SimWalk2 output, can be obtained, at no cost, from the authors or at the Epigenetic Web site.

Acknowledgments

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

Electronic-Database Information

URLs for data presented herein are as follows:

Epigenetic, <http://www.epigenetic.org/Linkage/Imprint/> (for imprinting programs for performing extended-pedigree im-

printing and linkage test and pedigree structure used in this simulation)

Imprinted Gene Catalogue, <http://cancer.otago.ac.nz/IGC/Web/home.html>

References

- Etzel CJ, Shete S, Beasley TM, Fernandez JR, Allison DB, Amos CI (2003) Effect of box-cox transformation on power of Haseman-Elston and maximum-likelihood variance components tests to detect quantitative trait loci. *Hum Hered* 55: 108–116
- 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
- Karason A, Gudjonsson JE, Upmanyu R, Antonsdottir AA, Hauksson VB, Runasdottir EH, Jonsson HH, Gudbjartsson DF, Frigge ML, Kong A, Stefansson K, Valdimarsson H, Gulcher JR (2003) A susceptibility gene for psoriatic arthritis maps to chromosome 16q: evidence for imprinting. *Am J Hum Genet* 72:125–131
- Li E (2002) Chromatin modification and epigenetic reprogramming in mammalian development. *Nat Rev Genet* 3:662–673
- 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
- Ott J (1989) Computer-simulation methods in human linkage analysis. *Proc Natl Acad Sci USA* 86:4175–4178
- 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
- Self SG, Liang K-Y (1987) Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under non-standard conditions. *J Am Stat Assoc* 82:605–610
- Shete S, Amos CI (2002) Testing for genetic linkage in families by a variance-components approach in the presence of genomic imprinting. *Am J Hum Genet* 70:751–757
- Shete S, Beasley M, Etzel CJ, Fernandez JR, Chen J, Allison DB, Amos CI. Effect of winsorization on power and type 1 error of variance components and related methods of QTL detection. *Behav Genet* (in press)
- Sobel E, Lange K (1996) Descent graphs in pedigree analysis: applications to haplotyping, location scores, and marker-sharing statistics *Am J Hum Genet* 58:1323–1337
- Sorant AJM, Elston RC (1994) A subroutine package for function maximization (a user's guide to MAXFUN version 6.0). Department of Biometry and Genetics, Louisiana State University Medical Center, New Orleans
- 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
- Walter J, Paulsen M (2003) Imprinting and disease. *Semin Cell Dev Biol* 14:101–110
- Weeks DE, Ott J, Lathrop GM (1990) SLINK: a general simu-

lation program for linkage analysis. *Am J Hum Genet* 47:
A204
Weeks DE, Sobel E, O'Connell JR, Lange K (1995) Computer
programs for multilocus haplotyping of general pedigrees. *Am*

J Hum Genet 56:1506–1507
Wijsman EM, Amos CI (1997) Genetic analysis of simulated
oligogenic traits in nuclear and extended pedigrees: summary
of GAW10 contributions. *Genet Epidemiol* 14:719–735