Linkage mapping of quantitative trait loci in humans: an overview

S. GHOSH[1], T. REICH[1] AND P. P. MAJUMDER[2]
[1] *Department of Psychiatry, Washington University School of Medicine, St Louis, MO, USA*
[2] *Anthropology and Human Genetics Unit, Indian Statistical Institute, Kolkata, India*

(Received 7.3.02. Accepted 4.7.02)

SUMMARY

In this article, we provide an overview of the different statistical procedures that have been developed for linkage mapping of quantitative trait loci. We outline the model assumptions, the data requirements and the underlying tests for linkage for the different methods.

INTRODUCTION

Many quantitative traits such as blood pressure and body mass index (BMI) are known to be determined primarily, though not exclusively, by inherited genetic factors. It is thus of considerable importance to identify chromosomal locations of the genes that control a quantitative character. Linkage analysis (Ott, 1999), which deals with the detection of linkage and estimation of recombination fractions among the loci controlling a qualitative/quantitative character and marker loci whose positions are known *a priori*, is widely used for localization of genes. Although statistical methodologies for mapping genes determining dichotomous qualitative characters in humans are well-developed, the development of such methodologies, especially those that are statistically and computationally efficient, for human quantitative traits is an active area of current research in human genetics. It has been emphasized that many traits that have traditionally been treated as qualitative are inherently quantitative in nature.

Although the idea of mapping quantitative trait loci (QTL mapping) can be traced back to Sax (1923), who studied the nature of association of seed size with seed-coat pattern and pigmentation in beans, the recent development of dense maps of highly polymorphic DNA markers in plants and animals has resulted in a resurgence of interest in QTL mapping. Statistical linkage relies on the nature and extent of co-inheritance of alleles at the trait and marker loci. For many plants and animals experimental crosses can be set up such that the trait locus genotype of an offspring can be unambiguously inferred. This simplifies the statistical investigation of co-inheritance of alleles at the trait and marker loci. However, it is not possible to set up experimental crosses for humans. Moreover, for experimental organisms, traits are often Mendelian in nature which facilitates the knowledge of trait genotypes. On the other hand, most human quantitative traits follow a complex mode of inheritance. Hence, QTL mapping in humans is statistically more difficult than in experimental plants and animals. In this article, we provide an overview, albeit non-exhaustive, of the different statistical procedures that have been developed for linkage mapping of QTLs.

MODELLING A QUANTITATIVE TRAIT

A quantitative trait (Y) can be modelled in a general way as $Y = G + E$, where G and E are the genetic and environmental contributions to the phenotype, respectively. While this general form of the model can be used in an exploratory way to provide some broad statistical inferences about

Correspondence: Saurabh Ghosh, Department of Psychiatry, Washington University School of Medicine, 660 S. Euclid, Campus Box 8134, St Louis, MO 63110-1093. Tel: (314) 362-8079; Fax: (314) 362-4247.
E-mail: saurabh@silver.wustl.edu

the quantitative trait, such as heretability of the trait, for making specific inferences or for QTL mapping, it is necessary to formulate a more detailed model. Often models are formulated on the basis of exploratory data analyses.

A quantitative trait may be determined, in addition to an environmental component whose expectation is usually assumed to be zero, by one or more loci, each biallelic or multiallelic, linked or unlinked. There may be dominance effects at various loci, and unlinked loci may also interact epistatically in the determination of the trait values.

For a quantitative trait that is determined by a single biallelic locus, a general model is: $Y|A_1A_1 \sim f_1(\mu_1, \sigma_1^2)$, $Y|A_1a_1 \sim f_2(\mu_2, \sigma_2^2)$ and $Y|a_1a_1 \sim f_3(\mu_3, \sigma_3^2)$, where A_1 and a_1 are the two alleles at the locus, and f_1, f_2 and f_3 are general probability distribution functions with means μ_is and variances σ_i^2s. If allelic effects are additive, that is, there is no dominance, then $\mu_1 = 2\alpha$, $\mu_2 = \alpha+\beta$ and $\mu_3 = 2\beta$, where α and β are the allelic effects of A_1 and a_1, respectively. In the presence of a dominance effect, δ, $\mu_1 = 2\alpha$, $\mu_2 = \alpha+\beta+\delta$ and $\mu_3 = 2\beta$. Even these specified forms of the model are statistically too complicated for QTL mapping. Therefore, the popular statistical model is: $Y|A_1A_1 \sim f_1(a, \sigma^2)$, $Y|A_1a_1 \sim f_2(d, \sigma^2)$ and $Y|a_1a_1 \sim f_3(-a, \sigma^2)$. A detailed discussion on modelling of quantitative traits is available in Falconer & Mackay (1996). Often f_1, f_2 and f_3 are assumed to be $N(.,.)$.

EARLY METHODS

One of the most popular approaches of analyzing human linkage data is based on sib-pairs. Some of the earliest contributions in these studies were made by Penrose. He assessed the efficiency of using concordant and discordant sib-pairs (in terms of quantitative trait values) in studying multifactorial disorders (Penrose, 1935). It was shown by Penrose (1947) based on a linkage study between the loci for phenylketonuria and the presence or absence of the B allele at the ABO locus, that the efficiency and complexity of detection and estimation of linkage can be increased by distinguishing the two types of identical sib-pairs. Penrose (1953) extended his earlier methods to multiple alleles using data on red-hair and the ABO locus restricted to a single generation. An extensive review of Penrose's contributions and the subsequent extensions to QTL mapping procedures using sib-pairs is presented in Edwards (1998).

HASEMAN–ELSTON AND ITS EXTENSIONS

A popular model-free linkage method is to utilise the inverse relationship between the difference between trait values of sib-pairs and their marker identity-by-descent (i.b.d.) scores. A pair of related individuals shares an allele i.b.d. if that allele has a common ancestral source. For sib-pairs, the common ancestors are their parents. Haseman & Elston (1972) developed a regression approach for detecting linkage based on the squared difference in quantitative trait values of sib-pairs (Y) and their estimated marker i.b.d. scores ($\hat{\pi}_m$). The basis of the regression is the equation:

$$E(Y|\hat{\pi}_m) = \alpha + \beta\hat{\pi}_m, \qquad (1)$$

where there is no dominance in the trait and $\beta = -2p(1-p)a^2(1-2\theta)^2$; p being the allele frequency of A_1, a the conditional expectation of the trait given genotype A_1A_1 and θ the recombination fraction between the QTL and the marker locus. Function of θ, a test for no linkage (i.e. $\theta = 0.5$) is equivalent to testing $\beta = 0$ in Equation (1). The test can be performed via the usual t statistic based on the least squares regression estimate of β.

Amos & Elston (1989) extended the above regression procedure to other relative pairs. For each type of relative pair, the regression parameter β is a different function of θ. However, the test for no linkage in each case is equivalent to testing $\beta = 0$. Amos *et al.* (1989) showed that in the presence of dominance in the trait, the least squares estimator of β is biased. They derived the conditional variance of Y given $\hat{\pi}_m$ as $\alpha_0+\beta_0\hat{\pi}_m+\gamma_0\hat{\pi}_m^2$. The test for linkage is based on the weighted least squares estimators of β_0 and γ_0, and is more

powerful than the original Haseman–Elston test (1972). Olson & Wijsman (1993) used generalised estimating equations to combine information from different types of relative pairs in a set of pedigree data. The test for no linkage between the QTL and the marker locus is equivalent to testing $\beta = 0$ where β is the vector of regression coefficients of Ys on $\hat{\pi}_m$s corresponding to the different types of relative pairs. The test statistic is of the form $\sqrt{N}c'\hat{\beta}/\{c'\mathrm{Var}(\hat{\beta})c\}^{1/2}$, where c is a vector of weights chosen proportional to $\{\mathrm{Var}(\hat{\beta})\}^{-1}\hat{\beta}$. Elston *et al.* (2000) suggested that the mean-corrected cross-product of the sib-pair trait values carry more linkage information than the squared sib-pair trait difference used in the traditional Haseman–Elston set-up (1972), and have implemented these regression procedures in the computer package SAGE. However, recent studies have shown both analytically and empirically that a combined least squares regression analysis with appropriate weighting of squared sib-pair sum and squared sib-pair difference (Drigalenko, 1998; Xu *et al.* 2000; Forrest, 2001; Visscher & Hopper, 2001) may be more powerful in detecting linkage than the traditional Haseman–Elston method (1972) or that proposed in Elston *et al.* (2000). Although the Haseman–Elston class of regression models does not assume any specific probability distribution for the trait values, it has been found that a t distribution approximation for the test statistic (based on the slope parameter) is often anti-conservative and leads to an inflated rate of false positives, especially when the sibship size is large (Elston *et al.* 2000).

Fulker & Cardon (1994) extended the Haseman–Elston (1972) regression equation to interval mapping. They proposed a method where the i.b.d. scores at the flanking markers (π_{m1} and π_{m2}) are estimated separately using marginal marker information and the trait i.b.d. score (π_t) is estimated using the equation:

$$\hat{\pi}_t = \rho_0 + \rho_1\hat{\pi}_{m1} + \rho_2\hat{\pi}_{m2}.$$

Y is regressed on $\hat{\pi}_t$ and the approximate position of the QTL is inferred based on the plot of $\hat{\beta}/\hat{s}.\hat{e}.(\hat{\beta})$, where $\hat{\beta}$ is the regression estimator of Y on $\hat{\pi}_t$. Olson (1995) suggested that in order to obtain maximum information, the marker i.b.d. scores be jointly estimated using all available marker data. The resultant regression equation was:

$$E(Y \mid \hat{\pi}_{m1}, \hat{\pi}_{m2}) = \beta_0 + \beta_1\hat{\pi}_{m1} + \beta_2\hat{\pi}_{m2},$$

where there is no dominance in the trait loci. Fulker, Cherny & Cardon (1995) extended the interval mapping procedure of Fulker & Cardon (1994) to take account of information from all marker loci simultaneously. They showed that the power of the traditional Haseman–Elston method (1972) can be substantially improved by this strategy when the markers differ in their information content. Their method has provided a framework for multipoint i.b.d. estimation not restricted to the class of Haseman–Elston regression methods.

Tiwari & Elston (1997) extended the traditional Haseman–Elston (1972) procedure to the case of two unlinked QTLs which might interact epistatically. They showed that under a fairly general model of epistasis, where they assumed that the marginal genotypic effects of the QTLs as well as those of the epistatic interactions are additive, the expectation of Y is a linear function of $\hat{\pi}_{m1}, \hat{\pi}_{m2}, f_1, f_2$ and their pairwise cross-product terms, where f_1 and f_2 are the probabilities that a sib-pair shares 1 and 2 alleles i.b.d., respectively. Under a restricted set-up, Ghosh & Majumder (2001) derived a regression equation for multiple unlinked QTLs using a generalized digenic interaction model (Kearsey & Pooni, 1996) and examined the marginal effects of the different trait and linkage parameters in mapping the underlying QTLs.

VARIANCE COMPONENTS

Another popular statistical approach for QTL mapping is to dissect the genetic variation within the quantitative trait. Although parametric in nature (i.e. the methods assume specific probability distributions for trait values), the advantage of using these methods is that larger sibships or entire pedigrees can be simultaneously

analyzed. Although Goldgar (1990) developed a variance components model which assumed that several genetic factors from a chromosomal region influence the quantitative trait, and Schork (1993) studied its power extensively, the basic framework for variance components linkage analysis was provided by Amos (1994).

The general variance components model is given by:

$$Y = \mu + g + G + e,$$

where μ is the overall mean of the quantitative trait, g is a random effect due to a major gene with additive variance σ_a^2 and dominance variance σ_d^2, G is a random polygenic effect with variance σ_G^2 and e is the non-shared environmental effect (or random error) with variance σ_e^2. The trait values of individuals in a pedigree are usually assumed to be distributed as multivariate normal with dispersion matrix V, where the variance of the trait value of each individual is $\sigma_a^2 + \sigma_d^2 + \sigma_G^2 + \sigma_e^2$ and the covariance between the trait values of two individuals is given by $\phi\sigma_a^2 + \Delta\sigma_d^2 + \phi\sigma_G^2$, where ϕ is the coefficient of relationship between the two individuals and Δ is the probability that the two individuals share both their alleles i.b.d. at the major locus (Amos, 1994). Conditioned on i.b.d. score (π) at a marker locus, the above covariance is given by $f(\theta, \pi)\sigma_a^2 + g(\theta, \Delta)\sigma_d^2 + \phi\sigma_G^2$, where θ is the recombination fraction between the QTL and the marker locus. The log-likelihood of the data is given by:

$$c - \frac{1}{2}\sum_P \log|V| - \frac{1}{2}\sum_P (Y - \mu 1)' V^{-1} (Y - \mu 1),$$

where c is a constant, Y and $\mu 1$ are respectively the vector of trait values and that of the means within a pedigree and the summation is over independent pedigrees. The variance components methods use the maximum likelihood method to estimate the parameters. The test for linkage is equivalent to testing $\sigma_a^2 = 0$ versus $\sigma_a^2 > 0$. The usual likelihood ratio test statistic is distributed as a 50:50 mixture of a χ^2 distribution with 1 d.f. and a χ^2 distribution with 0 d.f. (defined as a degenerate variable at 0). The model can also incorporate other environmental covariates.

Almasy & Blangero (1998) developed a general framework of multipoint i.b.d. probability calculations using pedigrees of arbitrary sizes. The correlations in i.b.d. scores were shown to be a function of the chromosomal distances for different relative pairs in a general pedigree. They extended the model of Amos (1994) to incorporate multiple QTLs. Their variance components method considers increase in log-likelihood of the data with sequential addition of QTLs and has been implemented in a computer package, SOLAR. The computer package GENEHUNTER 2 also includes a maximum likelihood-based variance components model with a provision of fixing the dominance variance of the underlying QTL and/or other unlinked QTLs at zero.

We emphasize here that the variance components methods are dependent on the assumption of a specific probability distribution (multivariate normal in most scenarios) for the trait values. If the underlying quantitative trait distribution is indeed normal, one would expect these methods to be much more powerful than distribution-free methods (discussed in the next section). However, it is often not feasible to verify distributional and other model assumptions. When underlying assumptions are violated, the behaviour of parametric methods is unclear as it could yield either a high rate of false positives or a high rate of false negatives. For example, leptokurtosis of trait distribution and the presence of gene-environment interaction can lead to inflated false positive error rates (Allison *et al.* 2000).

NON-PARAMETRIC ALTERNATIVES

Statistical methods for mapping QTLs, which involve assumptions of specific probability distributions for trait values, are often susceptible to deviations from underlying distributional assumptions. Some of the non-parametric (distribution-free) methods proposed a test statistic based on the rank correlation between the absolute differences in trait values of sib-pairs and their estimated marker i.b.d. scores. Kruglyak & Lander (1995*a*) proposed a Wilcoxon rank sum

test based on ranks of squared differences in sib-pair trait values and an indicator variable depending on the marker genotype. A detailed discussion on some of the distribution-based and distribution-free multipoint sib-pair linkage approaches, which have been implemented in the computer package MAPMAKER/SIBS, is presented in Kruglyak & Lander (1995*b*). The computer package GENEHUNTER 2 includes the Haseman–Elston class of regressions as well as the different analytical methods of MAPMAKER/SIBS. Ghosh & Majumder (2000*a*) have developed a two-stage linkage procedure, in which rank correlation between the squared sib-pair trait difference and their estimated marker i.b.d. score is used at the coarse-mapping stage and a non-parametric regression procedure based on kernel smoothing is implemented for fine-mapping.

EXTREME SIB-PAIRS

Risch & Zhang (1995) observed that analysis of extremely discordant sib-pairs (i.e. one sib has the quantitative trait value in the upper decile of the trait distribution, while the other has a trait value in the lower decile) yields more power than random sib-pairs, thereby reducing the sample size requirements for genotyping over conventional designs. However, it is often not feasible to obtain extremely discordant sib-pairs. Moreover, under oligogenic QTL models, where heterozygosities of different loci vary widely, using extremely discordant sib-pairs may not be an optimal strategy for mapping the more heterozygous loci (Allison *et al.* 1998). An alternative is to include extremely concordant sibs in the analysis (Eaves & Meyer, 1994; Zhang & Risch, 1996; Gu *et al.* 1996; Gu & Rao, 1997) which provides a compromise between the power to detect linkage and the availability of extreme sib-pairs.

OTHER METHODS

An interesting method for linkage analysis with pedigree data was proposed by Heath (1997), in which reversible jump Markov Chain Monte Carlo (MCMC) methods were used to implement a sampling scheme in which the Markov chain can jump between parameter subspaces corresponding to models with different numbers of QTLs. Though the method involves assumption of specific probability distributions for the trait, it avoids the problem of misspecification of the number of QTLs. The method has been implemented in a computer package, LOKI. Lee & Thomas (2000) have developed a refined MCMC procedure by improving on the marker-haplotype updating algorithm.

Another approach has been motivated by the classical LOD score statistic (Morton, 1955) using inclusion and exclusion mapping. Page *et al.* (1998) have proposed a QLOD score statistic for detecting linkage in QTLs, where the traditional critical values of 3 and -2 for the underlying sequential tests were used.

Alcais & Abel (1999) have developed a maximum-likelihood-binomial method of mapping QTLs using sibship data. The idea is to introduce a latent binary variable Z which captures linkage information between the QTL and the marker locus. The likelihood is formulated in terms of:

$$P(M_1, M_2 | Y) = \sum_Z P(Z | Y) P(M_1, M_2 | Z),$$

where Y is the observed phenotype and M_1, M_2 are the alleles at the marker locus. $P(Z|Y)$ is modelled by a probit distribution and $P(M_1, M_2|Z)$ by a Bernoulli distribution. The test for linkage is based on a likelihood ratio test of the Bernoulli parameter $= 0.5$.

COMPARATIVE STUDIES

There have been a few comparative studies between the different statistical techniques for QTL mapping in humans. Alcais & Abel (2000) showed that larger sibships contain more linkage information than independent sib-pairs. They also showed that their maximum-likelihood-binomial approach, which does not require decomposition of sibships into sib-pairs, is more powerful and cost-effective compared to extremely discordant sib-pair analyses. Visscher &

Hopper (2001) compared three sib-pair methods in the Haseman–Elston class of regressions and four maximum likelihood methods under the assumption of normality for the trait values. They showed that the Elston *et al.* (2000) method may be less powerful than both the traditional Haseman–Elston method and a complete maximum-likelihood analysis, especially if the sib-pair correlation is high. Efficiencies of variance components versus sib-pair based linkage methods was examined by Williams & Blangero (1999), where they observed that these have similar performances with respect to unbiasedness of the estimate of QTL location and Type I error rate; but within the single sib-pair and sibship sampling units, the variance components approach gave consistently superior power and efficiency of parameter estimation. However, Sham & Purcell (2001) have highlighted the asymptotic equivalence in power between a combined Haseman–Elston regression based on the squared sum and the squared difference of sib-pair trait values and variance components analyses.

MULTIVARIATE PHENOTYPES

One of the major current challenges in genetic epidemiology is to unravel genetic architectures of complex traits. Quantitative variables, possibly correlated, generally underlie complex traits. Many models and approaches have been developed, including variance components (Lange & Boehnke, 1983; Schork, 1993), regressive model (Bonney *et al.* 1998; Moldin & van Eerdewegh, 1995), multivariate extension of the Haseman–Elston model (Amos *et al.* 1990; Amos & Liang, 1996) and structural equations model (Eaves *et al.* 1996; Todorov *et al.* 1998) to jointly analyze data on several correlated quantitative phenotypes as a single multivariate phenotype. However, the power of a multivariate analysis to detect linkage can be substantially low (Ott & Rabinowitz, 1999). Data reduction techniques, such as principal components analysis or factor analysis, (Zlotnik *et al.* 1983; Hasstedt *et al.* 1994; Boomsma, 1996; Allison & Beasley, 1998; Ott & Rabinowitz, 1999) help in circumventing this problem of reduced power. However, it is important to realize that unless the variables included in a principal component are significantly correlated, inferences on linkage could be highly misleading (Majumder *et al.* 1998; Ghosh & Majumder, 2000*b*).

DISCUSSION

The aim of this article was to provide an overview of the different linkage methodologies developed for mapping quantitative trait loci. As mentioned in the Introduction, this is a non-exhaustive set of existing methods and we have simply tried to highlight the various statistical techniques along with the underlying data requirements and model assumptions.

While there is clearly no uniformly most powerful method for detecting linkage, certain methods are more optimal than others under relevant assumptions. As mentioned in a previous section, likelihood-based variance components methods are expected to perform better than distribution-free methods if assumptions (like normality) for the underlying quantitative trait distribution are valid. Non-parametric methods, which are more robust to deviations from underlying assumptions, can be viewed as complementary to the distribution-based approaches. Thus, a possible way to enhance confidence in a linkage finding is to verify whether multiple methods, under varying assumptions, replicate the finding not only with the same data but also with independent sets of data.

This work was supported by the NIAAA through UPHS grants U10AA08403 and U10AA08401.

REFERENCES

Alcais, A. & Abel, L. (1999). Maximum-likelihood-binomial method for genetic model-free linkage analysis of quantitative traits in sibships. *Genet. Epidemiol.* **17**, 102–117.

Alcais, A. & Abel, L. (2000). Linkage analysis of quantitative trait loci: sib-pairs or sibships? *Hum. Hered.* **50**, 251–256.

Allison, D. B. & Beasley, M. (1998). A method and computer program for controlling the family-wise alpha rate in gene association studies involving multiple phenotypes. *Genet. Epidemiol.* **15**, 87–101.

Allison, D. B., Heo, M. *et al.* (1998). Extreme selection strategies in gene mapping studies of oligogenic quantitative traits do not always increase power. *Hum. Hered.* **48**, 97–107.

Allison, D. B., Neale, M. C. *et al.* (2000). Testing the robustness of the likelihood-ratio test in a variance-component quantitative-trait loci-mapping procedure. *Am. J. Hum. Genet.* **65**, 531–544.

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

Amos, C. I. & Elston, R. C. (1989). Robust methods for the detection of genetic linkage for quantitative data from pedigrees. *Genet. Epidemiol.* **6**, 349–360.

Amos, C. I., Elston, R. C. *et al.* (1989). A more powerful robust sib-pair test linkage for quantitative trait. *Genet. Epidemiol.* **6**, 435–449.

Amos, C. I., Elston, R. C. *et al.* (1990). A multivariate approach for detecting linkage, with application to a pedigree with an adverse lipoprotein phenotype. *Am. J. Hum. Genet.* **47**, 247–254.

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

Amos, C. I. & Liang, A. E. (1996). A comparison of univariate and multivariate tests for genetic linkage. *Genet. Epidemiol.* **10**, 671–676.

Bonney, G. E., Lathrop, G. M. *et al.* (1988). Combined linkage and segregation analysis using regressive models. *Am. J. Hum. Genet.* **43**, 29–37.

Boomsma, D. I. (1996). Using multivariate genetic modeling to detect pleiotropic quantitative loci. *Behav. Genet.* **26**, 161–166.

Blackwelder, W. C. & Elston, R. C. (1985). A comparison of sib-pair linkage tests for disease susceptibility loci. *Genet. Epidemiol.* **2**, 85–97.

Drigalenko, E. (1998). Hoe sib-pairs reveal linkage. *Am. J. Hum. Genet.* **63**, 1242–1245.

Eaves, L. & Meyer, J. (1994). Locating human quantitative trait loci: guidelines for the selection of sibling pairs for genotyping. *Behav. Genet.* **24**, 443–455.

Eaves, L. J., Neale, M. C. *et al.* (1996). Multivariate multipoint linkage analysis of quantitative trait loci. *Behav. Genet.* **26**, 519–525.

Edwards, J. H. (1998). Penrose and sib-pairs. *Ann. Hum. Genet.* **62**, 365–377.

Elston, R. C., Buxbaum, S. *et al.* (2000). Haseman and Elston revisited. *Genet. Epidemiol.* **19**, 1–17.

Falconer, D. S. & Mackay, T. F. C. (1996). Introduction to quantitative genetics. 4th edition. Prentice Hall.

Forrest, W. (2001). Weighting improves the 'new Haseman–Elston' method. *Hum. Hered.* **52**, 47–54.

Fulker, D. W. & Cardon, L. R. (1994). A sib-pair approach to interval mapping of quantitative trait data. *Am. J. Hum. Genet.* **54**, 1092–1103.

Fulker, D. W., Cherny, S. S. *et al.* (1995). Multipoint interval mapping of quantitative trait loci, using sib pairs. *Am. J. Hum. Genet.* **56**, 1224–1233.

Ghosh, S. & Majumder, P. P. (2000*a*). A two-stage variable stringency semi-parametric method for mapping quantitative trait loci with the use of genome-wide scan data on sib pairs. *Am. J. Hum. Genet.* **66**, 1046–1061.

Ghosh, S. & Majumder, P. P. (2000*b*). Deciphering the genetic architecture of a multivariate phenotype. *Adv. in Genet.* **42**, 323–347.

Ghosh, S. & Majumder, P. P. (2001). Mapping a human quantitative trait. *Current Science* **80**, 1145–1160.

Goldgar, D. E. (1990). Multipoint analysis of human quantitative genetic variation. *Am. J. Hum. Genet.* **47**, 957–967.

Gu, C., Todorov, A. A. *et al.* (1996). Combining extremely concordant sibpairs with extremely discordant sibpairs provides a cost effective way to linkage analysis of quantitative trait loci. *Genet. Epidemiol.* **13**, 513–533.

Gu, C. & Rao, D. C. (1997). A linkage strategy for detection of human quantitative-trait loci. II. Optimization of study designs based on extreme sib pairs and generalized relative risk ratios. *Am. J. Hum. Genet.* **61**, 211–222.

Haseman, J. K. & Elston, R. C. (1972). The investigation of linkage between a quantitative trait and a marker locus. *Behav. Genet.* **2**, 3–19.

Hasstedt, S. J., Hunt, S. C. *et al.* (1994). Evidence for multiple genes determining sodium transport. *Genet. Epidemiol.* **11**, 553–568.

Heath, S. C. (1997). Markov chain Monte Carlo segregation and linkage analysis for oligogenic models. *Am. J. Hum. Genet.* **61**, 748–760.

Kearsey, M. J. & Pooni, H. S. (1996). *The Genetical Analysis of Quantitative Traits.* London: Chapman and Hall.

Kruglyak, L. & Lander, E. S. (1995*a*). A nonparametric approach for mapping quantitative trait loci. *Genet.* **139**, 1421–1428.

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

Lange, K. & Boehnke, M. (1983). Extensions to pedigree analysis. IV. Covariance components models for multivariate traits. *Am. J. Med. Genet.* **14**, 513–524.

Lee, J. K. & Thomas, D. C. (2000). Performance of Markov chain-Monte Carlo approaches for mapping genes in oligogenic models with an unknown number of loci. *Am. J. Hum. Genet.* **67**, 1232–1250.

Majumder, P. P., Moss, H. B. *et al.* (1998). Familial and nonfamilial factors in the prediction of disruptive behaviors in boys at risk for substance abuse. *J. Child Psychol. Psychiat.* **39**, 203–213.

Moldin, S. O. & van Eerdewegh, P. (1995). Multivariate genetic analysis of oligogenic disease. *Genet. Epidemiol.* **12**, 801–806.

Morton, N. E. (1955). Sequential tests for the detection of linkage. *Am. J. Hum. Genet.* **7**, 277–318.

Olson, J. M. (1995). Robust multipoint linkage analysis, an extension of the Haseman-Elston method. *Genet. Epidemiol.* **12**, 177–193.

Olson, J. M. & Wijsman, E. M. (1993). Linkage between quantitative trait and marker loci, methods using all relative pairs. *Genet. Epidemiol.* **10**, 87–102.

Ott, J. (1999). *Analysis of Human Genetic Linkage.* 3rd edition. Baltimore: Johns Hopkins University Press.

Ott, J. & Rabinowitz, D. (1999). A principal-components approach based on heritability for combining phenotype information. *Hum. Hered.* **49**, 106–111.

Page, G. P., Amos, C. I. *et al.* (1998). A quantitative lod score test statistic and sample size for exclusion and linkage of quantitative traits in human sibships. *Am. J. Hum. Genet.* **62**, 962–968.

Penrose, L. S. (1935). The detection of autosomal linkage in data which consist of pairs of brothers and sisters of unspecified parentage. *Ann. Eugen.* **8**, 133–138.

Penrose, L. S. (1947). A further note on the sib-pair linkage method. *Ann. Eugen.* **13**, 25–29.

Penrose, L. S. (1953). The general purpose of sib-pair linkage test. *Ann. Eugen.* **18**, 120–124.

Risch, N. & Zhang, H. (1995). Extreme discordant sibpairs for mapping quantitative trait loci in humans. *Science* **268**, 1584–1589.

Sax, K. (1923). The association of size differences with seed-coat pattern and pigmentation in *Phaseolus vulgaris*. *Genet.* **8**, 552–560.

Schork, N. J. (1993). Extended multipoint identity-by-descent analysis of human quantitative traits: efficiency, power, and modeling considerations. *Am. J. Hum. Genet.* **53**, 1306–1319.

Sham, P. C. & Purcell, S. (2001). Equivalence between Haseman–Elston and variance-components linkage analyses for sib pairs. *Am. J. Hum. Genet.* **68**, 1527–1532.

Tiwari, H. K. & Elston, R. C. (1997). Linkage of multilocus components of variance of polymorphic markers. *Am. Hum. Genet.* **61**, 253–261.

Todorov, A. A., Volger, G. P. *et al.* (1998). Testing causal hypotheses in multivariate linkage analysis of quantitative traits: general formulation and application to sib-pair data. *Genet. Epidemiol.* **15**, 263–278.

Visscher, P. & Hopper, J. (2001). Power of regression and maximum likelihood methods to map QTL from sib-pair and DZ twin data. *Ann. Hum. Genet.* **65**, 583–601.

Williams, J. T. & Blangero, J. (1999). Comparison of variance components and sibpair-based approaches to quantitative trait linkage analysis in unselected samples. *Genet. Epidemiol.* **16**, 113–114.

Xu, X., Weiss, S. *et al.* (2000). A unified Haseman–Elston method for testing linkage with quantitative traits. *Am. J. Hum. Genet.* **67**, 1025–1028.

Zhang, H. & Risch, N. (1996). Mapping quantitative-trait loci in humans by use of extreme concordant sib pairs: selected sampling by parental phenotypes. *Am. J. Hum. Genet.* **59**, 951–957.

Zlotnik, L. H., Elston, R. C. *et al.* (1983). Pedigree discriminant analysis: a method to identify monogenic segregation. *Am. J. Med. Genet.* **15**, 307–313.