Power of Competing Strategies of Linkage Analysis for Complex Traits

Jianzhong Ma^a E. Warwick Daw^b Christopher I. Amos^a

^aDepartment of Epidemiology, The University of Texas M. D. Anderson Cancer Center, Houston, Tex., and

^bDivision of Statistical Genomics, Washington University School of Medicine, Saint Louis, Miss., USA

Key Words

Variance components • Linkage analysis • Location score • Multipoint analysis • Model-free methods • Markov chain Monte Carlo • Statistical power

Abstract

Variance components (VC) and the Bayesian Markov chain Monte Carlo (MCMC) analysis are two of the widely used linkage analysis approaches to mapping genes for complex quantitative traits. Both approaches can handle extended pedigrees and multiple markers and do not require a pre-specified genetic model. In this study, we used simulated data to compare the performance of these two approaches with the traditional parametric linkage analysis. Using simulated data sets without linkage between a quantitative trait and the markers, we estimated a critical value for various test scores used in VC or MCMC and the location (LOC) score at a fixed level of significance (5%). These critical values were then used to determine the power for the three methods for simulated data sets with linkage. We found that both the VC and MCMC approaches worked well, compared with the LOC score, when there was only one gene underlying the quantitative trait; however, VC had higher power than the other methods in a simulation study of a complex phenotype influenced by more than one gene. We also compared two im-

plementations of MCMC analysis, finding interpretation of results using the log of placement score was more accurate for linkage inference than the Bayes factor but required much more intensive simulation studies.

Copyright © 2010 S. Karger AG, Basel

Introduction

Gene mapping using linkage analysis is characterized by several features. The traits of interest are often complex in that they are determined by more than one gene as well as environmental or host-specific factors and there may be both allelic and locus heterogeneity. Information on the exact trait model, including the number of causal loci, is usually not available. Moreover, in addition to multiple markers, the approaches for linkage analysis are more powerful if they can handle extended pedigrees [1].

Two commonly used approaches that can handle these complications are variance components (VC) [2] and the Bayesian Markov chain Monte Carlo (MCMC) analysis [3]. Neither approach requires knowledge of a particular

A preprint was submitted to Elsevier Science on January 12, 2010.

trait mode of inheritance. Using the VC approach, the trait-determining mechanism is modeled only in terms of contributions of factors, including genes, to the total variance; hence, no detailed trait model is needed. Information provided by multiple markers is utilized through the multipoint identity by descent (IBD) scores. Only information about pairwise IBD sharing is used in the analysis. Calculation of the multipoint IBD scores for large pedigrees usually requires sampling-based approaches, such as MCMC. The Bayesian MCMC approach, on the other hand, is actually a joint segregation and linkage analysis in the sense that it models the explicit genetic model of an unknown number of genes. All the marker data as well as trait phenotypes are directly incorporated in the explicit model. In addition, the Bayesian MCMC jointly models data from entire pedigrees, unlike the VC model, which only considers pairs of individuals jointly. Therefore, Bayesian MCMC modeling would be expected to yield the most accurate procedure for linkage identification when the genetic model is unknown.

Statistical evaluations of linkage evidence are performed in different ways for the VC and Bayesian MCMC approaches. For the VC approach, the evidence of linkage is assessed using the likelihood ratio test (LRT) statistic. This statistic is usually assumed to be distributed as a 50:50 mixture of a $\chi^2(1)$ and a point mass at zero. This assumption is made because negative variance components are inadmissible [4]. However, this assumed distribution is violated in some circumstances, such as when pedigrees are sampled through individuals with extreme trait values or when the trait does not follow a multivariate normal distribution. Because of its Bayesian nature, the MCMC approach does not result in a p value or any test statistics that can be interpreted in the usual manner in the frequentist manner. Instead, the strength of evidence of linkage is measured using various scores defined in the Bayesian context through comparing the posterior to the prior distributions of the linkage parameter [5]. These scores, such as the Bayes factor, can be treated as empirical statistics. An empirical p value can thus be obtained through simulation, although the simulation is usually extremely time-consuming for an MCMC approach.

In this paper, we report on our evaluation of the performance of the VC approach as implemented in the program Multic in the package ACT [2] and the Bayesian MCMC approach as implemented in the package Loki [3]. A joint analysis of results from ACT versus other software programs, including SOLAR [6] and Mx [7], showed that all the VC procedures yielded similar tests [8]. We also

performed the traditional parametric location (LOC) score analysis, as a gold standard, as implemented in the program LINKMAP in the package FASTLINK [9], to our simulated data. For this approach the genetic model influencing trait values at the linked locus was known and specified, so that this method serves as a gold standard for the best performance of a linkage test statistic, which would not ordinarily be achieved due to usual uncertainty about the genetic model. Even this optimal approach has its limitation since the LINKMAP procedure assumes no residual correlation in the trait, given the single linked locus it models. We used extensive simulation to determine the critical values at a fixed significance level and then used them to estimate the power of these two programs for various test statistics. Performance of two MCMC-based test scores was also compared. Only random ascertainment was considered in this study. We evaluated the performance of these methods using either a very simple genetic model, which we conjectured should indicate highest power for the parametric test. The highly complex model that was contrasted was a plausible model for a cardiovascular disease risk factor and represented the complexity we anticipate for most complex genetically influenced traits. We anticipated power would be highest for MCMC procedures and least for parametric linkage for this complex phenotype, in which the trait phenotypes are correlated among pedigree members, due to effects from multiple unlinked loci and shared environmental effects.

Methods

Variance Components for Linkage Analysis

The mixed model for a quantitative trait [2] is described as

$$\vec{y}_i = \mu + X_i\vec{\beta} + \vec{a}_i + \vec{g}_{1i} + \vec{g}_{2i} + \vec{e}_i \quad (1)$$

where $\vec{y}_i$ is a vector of trait values for the i -th pedigree; μ is the overall mean; X_i is the matrix for the observed values of the covariates and $\vec{\beta}$ is the vector of the corresponding regression coefficients; $\vec{a}_i$ is an unobserved vector of additive polygenic effects with $\vec{a}_i \sim N(0, \sigma_a^2 G_i)$ with G_i the kinship coefficient matrix; $\vec{g}_{1i}$ and $\vec{g}_{2i}$ are unobserved vectors of two major genes' effects; and $\vec{e}_i$ is the random residual effect with $\vec{e}_i \sim N(0, \sigma_e^2 I_i)$ with I_i the identity matrix.

The variance-covariance matrix, defined as $V_i = \text{Cov}(\vec{y}_i | X_i, Z_i, G_i)$, can be expressed in terms of the variance components as follows

$$V_i = \sigma_a^2 G_i + \sigma_{g1}^2 Z_i + \sigma_{g2}^2 Z_i + \sigma_e^2 I_i \quad (2)$$

where σ_{gk}^2 is the variance attributed to major gene k and Z_i is the matrix of estimated proportion of IBD for all pairs of individuals in the pedigrees.

Assuming a multivariate normal distribution for the trait, we have the log likelihood as follows

$$\ln L_i = -\frac{k_i}{2} \ln 2\pi - \frac{1}{2} \ln |V_i| - \frac{1}{2} (\vec{y}_i - \mu - X_i \vec{\beta})^T V_i^{-1} (\vec{y}_i - \mu - X_i \vec{\beta}) \quad (3)$$

with k_i the number of individuals in pedigree i . If there is only one major gene in the model, test of linkage can be performed using the following score

$$\text{LRT} = 2 \ln \frac{L(g)}{L(g=0)} \quad (4)$$

where $L(g)$ and $L(g=0)$ represent likelihood with and without the major gene, respectively. When detecting two trait loci simultaneously, we use the following two LRT scores

$$\text{LRT}_1 = 2 \ln \frac{L(g_1, g_2)}{L(g_1=0, g_2)} \quad (5)$$

$$\text{LRT}_2 = 2 \ln \frac{L(g_1, g_2)}{L(g_1, g_2=0)} \quad (6)$$

where $L(g_1, g_2)$ is the likelihood when both genes are present in the model, while $L(g_1=0, g_2)$ and $L(g_1, g_2=0)$ are the likelihoods when one of them is absent. If g_2 is estimated to be greater than zero, then

$$2\text{LRT}_1 \sim \frac{1}{2} \chi_1^2 + \frac{1}{2} \chi_0^2$$

[10], but if g_2 is estimated to be zero, then the distribution has a complex degenerated form. The required multipoint IBD sharing scores for VC analysis were calculated at each marker location using the Loki package.

MCMC Segregation and Linkage Analysis

The linear model for a quantitative trait [3] is

$$y_i = \mu_b + X\beta + \sum_{j=1}^k \mu_{G_{ji}} + e_j \quad (7)$$

where y_j is the trait value of individual j ; μ_b is the baseline; k is the number of loci related to the trait; $\mu_{G_{ji}}$ is the genotypic mean with respect to μ_b of genotype G_{ji} for the i -th locus; $X\beta$ represents the effects of covariates; and e_j is the residual assumed to be normally distributed as $N(0, \sigma_e^2)$. Polygenic effects are not explicitly modeled here but are rather modeled as effects from multiple loci each of which can have small effects.

The trait model, together with the usual inheritance model and a set of prior distributions for the model parameters, gives rise to a Bayesian framework for the joint estimation of the segregation and linkage parameters. The joint posterior density of all parameters and the unobserved genotypes (including those of the markers), given observed trait data Y_T and marker data Y_M , can be expressed as

$$P(\vec{\mu}, \sigma_e^2, \vec{q}_T, \vec{q}_M, \vec{\lambda}, G_T, G_M | Y_T, Y_M) \propto P(Y_T, Y_M | G_T, G_M, \vec{\mu}, \sigma_e^2, \vec{q}_T, \vec{q}_M, \vec{\lambda}) P(G_T, G_M | \vec{q}_T, \vec{q}_M, \vec{\lambda}) P(\vec{\mu}, \sigma_e^2, \vec{q}_T, \vec{q}_M, \vec{\lambda}) \quad (8)$$

where $\vec{q}_T$ and $\vec{q}_M$ are vectors of the allele frequencies for the trait and marker loci, respectively; G_T and G_M denote the complete genotypes of all individuals for the trait loci and the marker loci, respectively; and $\vec{\lambda}$ represents the linkage parameters, including the locations of the trait loci and their statuses of linkage to the chromosomes. A series of Gibbs samplers are then used to generate samples of parameters from this distribution. A reversible-

jump MCMC sampler is used in Loki because the number of loci responsible for the trait k is an unknown parameter.

After a certain number (usually 500,000) of MCMC iterations are performed, one explores if there is evidence of linkage from the posterior distribution of the linkage parameter λ . Several scores have been proposed based on the posterior distributions. The L -score (LS) that is implemented in Loki is a variation of the Bayes Factor and is defined as [11]

$$LS = \frac{1}{N} \sum_{j: \#QTLs \geq 1} \frac{\Pr(\text{linkage})}{\Pr_{\text{prior}}(\text{linkage} | k \text{ QTLs in iteration } j)} \quad (9)$$

where N is the total number of iterations; $\Pr(\text{linkage})$ is the posterior probability of linkage in iteration j ; and $\Pr_{\text{prior}}(\text{linkage} | k \text{ QTLs in iteration } j) = 1 - (1 - m/L)^k$ with L the total genome length and m the length of interval over which LS is calculated, assuming equally spaced intervals are analyzed. The Bayes Factor score and the intensity ratio score, defined by Chapman et al. [12], have been shown to be highly correlated to the LS [5] and were not included in our performance comparison in this paper.

A second score we considered was the log of placement (LOP) posterior probability ratio, proposed by Daw et al. [13]. To obtain the LOP score, one needs to pair each of the chromosomes of interest with a pseudochromosome that has the same structure as the corresponding true chromosome, such as the marker locations and the marker allele frequencies. Only the marker genotypes on the pseudochromosome are different from those on the true one and are simulated using gene dropping following the Mendelian law. The LOP is defined as

$$\text{LOP}(l, c) = \log_{10} \left[\frac{P(l, c | T)}{P(l, c | P)} \right] \quad (10)$$

with l the location on the chromosome and c the size of the gene. Here, T stands for the true chromosome and P for the pseudo-chromosome.

In this study, the LS and LOP scores were calculated over 2-cM intervals. Their chromosome-wide maxima, also denoted hereafter as LS and LOP, respectively, were used as empirical test statistics of linkage for a performance comparison with the LRT score of the VC approach.

Parametric Linkage Analyses

When the model parameters are available, it is generally expected that the traditional parametric linkage analysis using the location (LOC) score as the test statistic has the greatest power in detecting linkage, provided that the model describing the phenotype-genotype relation is correctly specified. The linear model and inheritance model used in the parametric linkage analysis as implemented in the LINKMAP program are the same as those used in the Bayesian MCMC approach, except that only one trait locus is allowed. These approaches differ in two aspects. First, the usual parametric approach performs simple linkage analysis; it does not estimate the model parameters. Instead, the segregation parameters describing trait inheritance in the pedigree need to be specified. Second, the parametric approach does not need priors and instead uses the likelihood method to define a test statistic, that is, the LOC score for the test of linkage between a trait locus and a marker or markers, defined as [14]

$$S(x) = 2 \ln \frac{L(x)}{L(\infty)} \quad (11)$$

where $L(x)$ is the likelihood at position x of the disease locus, and $L(\bullet\bullet)$ the likelihood that the disease locus is off the map. Since we were using simulated data in this study, we were able to use the generating values for the model parameters to perform a parametric linkage analysis. We expected the LOC score would act as a gold standard. We used a multipoint linkage analysis, combining data from sets of four markers at a time since LINKMAP can only use a limited number of markers.

The linear model used in the parametric analysis assumes a single gene and a normally distributed residual. For our simulated data, there were either multiple trait loci or a polygenic factor (see below). For a parametric run using LINKMAP, we set the segregation parameters by putting the variances contributed by a polygenic factor, trait loci other than the one of interest, or a covariate into the residual.

Simulation

We used the following strategy to compare the performance of the VC and MCMC approaches. We first created 100 replicates of null data sets for which there was no linkage between the trait and markers, and then we performed linkage analysis using both approaches on each of the replicates. A critical value for each of the test statistics for a fixed significance level (5%) was determined by ranking the resultant 100 values of the corresponding statistics. Because we were studying evidence for linkage over an entire chromosome, we did not expect the test statistics to follow a simple χ^2 distribution for the VC or LOC score methods and therefore derived empirical critical values. Then, we created 100 replicates of linked data sets for which the trait locus was located somewhere on the simulated chromosome. The power for a test statistic was then obtained by performing linkage analysis for all 100 replicates of linked data sets and counting the number of replicates for which the test statistic exceeded the critical value.

We used two trait models for our comparison. The first model involved only a single major gene, in addition to a polygenic component and residual noise. We first created 130 extended pedigrees using a method described previously [15]. There were 52 individuals in each of the pedigrees. We then used the GENE-DROP [16] program to simulate the genotypes for a chromosome with 30 equally distanced (5 cM) markers, each of which had eight alleles with equal frequencies (0.125). We created 100 replicates, and each of the replicates was used to generate a null and a linked data set. For the linked data set, the trait locus was located at 1 cM to the right of marker 10 (i.e. with a location at 66 cM). For the null data set, the trait locus was unlinked to the chromosome.

There were two alleles, A and B, with frequencies of 0.74 and 0.26, respectively, for the trait locus. This trait locus was used to generate a quantitative trait using a model described as follows: baseline 100.0; genotypic means 0.0 (AA), 10.0 (AB), and 25.0 (BB); polygenic variance 50.0; and residual variance 440.0. The variance attributed to the major gene was 50.1.

The second model we used was from the Genetic Analysis Workshop 13 (GAW13) data [17]. For each of the 100 replicates, there were 4,692 individuals in 330 families containing 7 to 84 people. There were 399 microsatellite markers on 22 autosomal chromosomes. Among the eight phenotypes simulated, we considered only height, where 10 simulated trait loci contributed to the sex-specific variance (9.12 for males, and 7.29 for females) of 40, 20, 10, 5, 2, 2, 2, 1, 1, and 1%, respectively. We concentrated on the two largest genes, located at 80.41 and 184.89 cM (sex-aver-

aged) on chromosomes 5 and 7, respectively. We performed linkage analysis using the VC, MCMC, and parametric approaches using marker data for chromosomes 5 and 7 to test their power in detecting these two genes. For details of the trait model and marker locations and allele frequencies, readers are referred to Daw et al. [17]. Originally, no null data sets were simulated for GAW13. We therefore paired the phenotype (height) of one replicate with the marker data of another in a sequential order (i.e. phenotypes of replicate 1 were replaced by those of replicate 2, phenotypes of replicate 2 were replaced by those of replicate 3, and so on) to form 100 replicates of null data sets for the purpose of estimating critical values.

It should be noted that the effects of the trait loci do not exactly fit the assumptions of any of the models we used for linkage analyses (VC, MCMC, or parametric approaches). First, all of the effects were influenced by small trait-specific effects in addition to the direct effects from the genetic factors of interest. Sex affected the trait through a two-way interaction with the ten genes in addition to having a small additive effect. None of the three approaches models this interaction effect. Second, some of the ten genes (such as the largest one located on chromosome 5) had more than two alleles, while Loki uses a two-allelic model. For the parametric approach, a three-allelic model was used for the analysis with chromosome 5. Finally, since only chromosomes 5 and 7 were used in our analyses, genes located on other chromosomes were thus treated as residual, resulting in a deviation of the residual from a normal distribution. We anticipate that in linkage analysis of complex traits none of the existing models for analysis used by programs will be accurate, and we chose to study this example as reflecting the performance of commonly used linkage methods when usually applied.

Results

Figures 1–3 show the four test statistics computed using the VC, MCMC and parametric approaches for the null and linked data sets and their pairwise Spearman correlations. The two MCMC-based linkage measures, LS and LOP, showed fairly strong correlations ($r > 0.70$) for both the linked and null data sets. Correlations between the measures of VC and the parametric approach, LRT and LOC, were also strong. Weaker correlations appeared for other combinations of scores from different approaches. For those pairs of scores, correlations for the null data sets were even weaker than those for the linked data sets. The weakest correlations occurred between the LOC score and others for the null data sets. This may indicate that different approaches detect linkage using different information, given that they showed similar power (see below).

The critical values obtained from the replicates of null data sets for all four statistics and the corresponding power calculated using the replicates of linked data sets are given in table 1. For the three cases we considered (i.e. the

Fig. 1. Pairwise Spearman correlation of LOP, LS, LRT, and LOC from simulation using the one-gene model. Statistics plotted on each panel are indicated on the diagonal. Results from linked data sets appear above the diagonal, and results from null data sets appear below.

one-gene model and the two chromosomes in the multi-gene model), the critical values of LRT, LOP and LOC varied only slightly while the critical value of LS varied significantly between the two models we used for simulating data. For simulated data using the one-gene model, power obtained using LRT (88%) was equivalent to that using the LOP score (89%), but was lower than the power obtained using LS (97%) and LOC (96%). A two-sided Fisher's exact test showed that LS of Loki had significantly higher power than LRT of Multic ($p = 0.029$) and than LOP of Loki ($p = 0.049$). The slightly lower power of VC than MCMC likely reflects the more accurate modeling afforded by the analytical approach used by MCMC. In our simulated one-gene model, the major gene was designed to have a fairly large dominant effect ($\sigma_d^2 = 0.26$) compared to the additive one ($\sigma_a^2 = 1.57$). Since the MCMC approach models the explicit genetic model, this dominant effect was appropriately modeled by Loki. However, in performing linkage analysis using the program Multic, we used the option that models only the additive component, which is

Table 1. Critical values and power at 5% significance level for the test statistics

Model	LRT (Multic)	LOP (Loki)	LS (Loki)	LOC (LINKMAP)
One-gene ^a	6.8/88%	2.3/89%	3.9/97%	7.49/96%
Multigene (chr 5) ^b	8.8/92%	2.2/84%	7.1/88%	8.19/84%
Multigene (chr 7) ^c	8.9/34%	2.4/28%	9.2/32%	9.39/18%

^a Chromosome length: 150 cM. Number of markers: 30.

^b Chromosome length: 220 cM. Number of markers: 26. ^c Chromosome length: 200 cM. Number of markers: 22.

usual practice in initially testing for linkage using a VC approach. In addition, MCMC approaches model the entire inheritance of trait alleles in the pedigree while VC procedures only model joint IBD for pairs of individuals. Hence, MCMC procedures should have higher power to detect linkage when the genetic model is approximately correctly specified compared to VC procedures.

Fig. 2. Pairwise Spearman correlation of LOP, LS, LRT, and LOC from simulation for chromosome 5 using the multigene model. Statistics plotted on each panel are indicated on the diagonal. Results from linked data sets appear above the diagonal, and results from null data sets appear below.

For the multigene model, the VC-based LRT had the greatest power (92%) in detecting the largest gene on chromosome 5, while the two MCMC-based statistics and the parametric approach were relatively less powerful, although these differences were not statistically significant ($p > 0.1$). For the MCMC method, this was probably because the largest gene had no dominant effect as simulated; and more importantly, there were three alleles in the model, but Loki uses a two-allelic model for the trait loci. In this case, the VC was robust because it does not use an explicit trait model. For the parametric approach, first, the multigene effects, especially the sex-specific effect, were poorly modeled by the parametric approach; second, since there were six genotypes for the largest gene and only three could be modeled, the model was partly misspecified. For the second largest gene on chromosome 7, all statistics had poor performance with low power because the effect of this gene was designed to have a much weaker effect on height and to have a relatively strong dominant effect (85% additive and 15%

dominant). However, here the LRT had slightly higher power than the MCMC procedures, and statistically significantly higher power than the LOC of LINKMAP ($p = 0.015$). Also, the LS of Loki was significantly more powerful than the LOC of LINKMAP ($p = 0.033$).

Discussion

We have compared the performance of four statistics in linkage analysis of quantitative traits: the LRT used in the VC approach, the LS and the LOP score used in the MCMC approach, and the LOC score of the traditional parametric approach. We found that it was hard to declare a clear winner between the VC and MCMC approaches. The MCMC approach, especially when the LS was used, was more powerful when the underlying trait model was a simple one-gene model and had a dominant effect, while the VC worked better if the trait-determining gene had more than two additively acting alleles. Both

Fig. 3. Pairwise Spearman correlation of LOP, LS, LRT, and LOC from simulation for chromosome 7 using the multigene model. Statistics plotted on each panel are indicated on the diagonal. Results from linked data sets appear above the diagonal, and results from null data sets appear below.

approaches had similar or even higher power than the traditional parametric linkage analysis for the complex genetic models that we studied. For genes with very small effects, all three approaches worked poorly.

For the two MCMC-based statistics, the LS seemed to have greater power, but the corresponding critical value varied largely across different conditions in our simulations. In contrast, the LOP score was less powerful, but its critical value was similar for different conditions. This indicates that, when using LS score for a measure of linkage, one usually has to resort to simulation in order to determine a reliable critical value for the specific settings. This difference is a consequence of their different definitions: while the LS compares the posterior with the prior, the LOP compares the posterior on the true chromosome with that on the pseudochromosome. The stability of the LOP is probably a result of information canceling out on variations in marker information along the chromosome by using both the true chromosome and pseudochromosome. For the LRT score of the VC approach, the critical

value had a stable value around 7 to 9, while its power was comparable to that of the LS.

The surprisingly lower power of the LOC score, except in the case of the diallelic gene with a smaller effect, was probably due to the fact that the linear model used in the parametric approach requires a normal distribution with no correlations for the residual. Although the Bayesian MCMC approach has the same requirement, the oligogenic model makes the Bayesian MCMC approach more robust when more than one gene is contributing to the trait.

In conclusion, we recommend that both the VC and Bayesian MCMC approaches be applied when using linkage to locate a gene underlying a quantitative trait. If the gene has only two alleles, the MCMC approach may have slightly greater power. Otherwise, the VC approach should be more robust. For complex traits, either of these methods appears preferable to the LOC score approach. In particular, in this simulation study we knew what the correct parameters were to describe the trait locus for the

parametric LOC score analysis and hence our evaluation of power is more optimistic than might be applied in practice, when the parameters would have to be selected or estimated in some fashion. When using the MCMC approach, measuring evidence of linkage with the LS should be done with simulations because the null distribution depends on the model. Finally, we note that our conclusion is based on a small number of null replicates (100), and thus should be taken with caution, since the

difference between the critical values and hence the difference between powers of different statistics might simply be due to chance.

Acknowledgment

This work was supported by National Institute of Health grants R01ES09912, R01CA133996, P01CA034936, and P50CA016672.

References

- 1 Wijsman EM, Amos CI: Genetic analysis of simulated oligogenic traits in nuclear and extended pedigrees: summary of GAW10 contributions. *Genet Epidemiol* 1997;14:719–735.
- 2 Amos CI: Robust variance-components approach for assessing genetic linkage in pedigrees. *Am J Hum Genet* 1994;54:535–543.
- 3 Heath SC: Markov chain Monte Carlo segregation and linkage analysis for oligogenic models. *Am J Hum Genet* 1997;61:748–760.
- 4 Searle SR, Casella G, McCulloch CE: *Variance Components*. Wiley Series in Probability and Mathematical Statistics Section. New York: John Wiley and Sons, 1992.
- 5 Igo RP Jr, Wijsman EM: Empirical significance values for linkage analysis: trait simulation using posterior model distributions from MCMC oligogenic segregation analysis. *Genet Epidemiol* 2008;32:119–131.
- 6 Almasy L, Blangero J: Multipoint quantitative-trait linkage analysis in general pedigrees. *Am J Hum Genet* 1998;62:1198–1211.
- 7 Neale MC: *Mx: Statistical Modeling*, ed 4. Medical College of Virginia, Richmond, 1997.
- 8 Allison DB, Neale MC, Zannolli R, Schork NJ, Amos CI, Blangero J: Testing the robustness of the likelihood-ratio test in a variance-component quantitative-trait loci-mapping procedure. *Am J Hum Genet* 1999;65:531–544.
- 9 Gottingham R, Idury RM, Shaffer AA: Faster sequential genetic linkage computations. *Am J Hum Genet* 1993;53:252–263.
- 10 Self SG, Liang KL: Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *J Am Stat Assoc* 1987;82:605–610.
- 11 Shmulewitz D, Heath SC: Genome scans for Q1 and Q2 on general population replicates using Loki. *Genet Epidemiol* 2001;21(suppl 1):S686–S691.
- 12 Chapman NH, Igo RP, Thomson JB, Matsushita M, Brkanac Z, Holzman T, Berninger VW, Wijsman EM, Raskind WH: Linkage analyses of four regions previously implicated in dyslexia: confirmation of a locus on chromosome 15q. *Am J Med Genet (Neuropsychiatr Genet)* 2004;131B:67–75.
- 13 Daw EW, Wijsman EM, Thompson EA: A score for Bayesian genome screening. *Genet Epidemiol* 2003;24:181–190.
- 14 Ott J: *Analysis of Human Genetic Linkage*, ed 3. The Johns Hopkins University Press, Baltimore, 1999.
- 15 Ma J, Amos CI, Daw EW: Ascertainment correction for Markov chain Monte Carlo segregation and linkage analysis of a quantitative trait. *Genet Epidemiol* 2007;31:594–604.
- 16 MORGAN, <http://www.stat.washington.edu/thompson/Genepi/MORGAN/Morgan.shtml>.
- 17 Daw EW, Morrison J, Zhou X, Thomas DC: Genetic Analysis Workshop 13: simulated longitudinal data on families for a system of oligogenic traits. *BMC Genet* 2003;4(suppl):S3.