Genet. Res., Camb. (2009), **91**, *pp.* 147–159.
doi:10.1017/S001667230900010X Printed in the United Kingdom

Mapping multiple quantitative trait loci under Bayes error control†

DANIEL SHRINER*
Center for Research on Genomics and Global Health, National Institutes of Health, Bethesda, MD 20892, USA

(*Received 14 November 2008 and in revised form 4 March 2009*)

Summary

In mapping of quantitative trait loci (QTLs), performing hypothesis tests of linkage to a phenotype of interest across an entire genome involves multiple comparisons. Furthermore, linkage among loci induces correlation among tests. Under many multiple comparison frameworks, these problems are exacerbated when mapping multiple QTLs. Traditionally, significance thresholds have been subjectively set to control the probability of detecting at least one false positive outcome, although such thresholds are known to result in excessively low power to detect true positive outcomes. Recently, false discovery rate (FDR)-controlling procedures have been developed that yield more power both by relaxing the stringency of the significance threshold and by retaining more power for a given significance threshold. However, these procedures have been shown to perform poorly for mapping QTLs, principally because they ignore recombination fractions between markers. Here, I describe a procedure that accounts for recombination fractions and extends FDR control to include simultaneous control of the false non-discovery rate, i.e. the overall error rate is controlled. This procedure is developed in the Bayesian framework using a direct posterior probability approach. Data-driven significance thresholds are determined by minimizing the expected loss. The procedure is equivalent to jointly maximizing positive and negative predictive values. In the context of mapping QTLs for experimental crosses, the procedure is applicable to mapping main effects, gene–gene interactions and gene–environment interactions.

1. Introduction

In genome-wide mapping of quantitative trait loci (QTLs), the goal is to characterize the genetic architecture of a trait by simultaneously identifying the entire subset of chromosomal intervals that affect the trait (Lynch & Walsh, 1998). Observed data consist of phenotypic values and marker genotypes for a mapped population. For sparse marker maps, pseudo-markers are routinely inserted at evenly spaced intervals, e.g. 1 cM (Sen & Churchill, 2001). Genotypes are unobservable except at completely informative markers, but genotype probabilities can be inferred.

* Center for Research on Genomics and Global Health, Building 12A, Room 4047, 12 South Dr., MSC 5635, Bethesda, MD 20892-5635, USA. Tel: +1 (301) 435 0068. Fax: +1 (301) 451 5426. e-mail: shrinerda@mail.nih.gov
† Presented in part at the 6th Annual Meeting of the Complex Trait Consortium, 26–29 May 2007 in Braunschweig, Germany.

In a QTL mapping analysis, performing hypothesis tests of linkage to a phenotype of interest across an entire genome induces a multiple testing problem. Furthermore, linkage among loci induces correlation among tests. With respect to the multiplicity problem, Bonferroni corrections are conservative and sacrifice power in order to maintain a low genome-wide false positive error rate. With respect to the correlation problem, Bonferroni corrections can become more conservative as correlation among tests increases.

In order to gain power, a less stringent threshold for claiming significance can be used, provided that some small number of false positive claims is acceptable. The false discovery rate (FDR) controls the proportion of hypothesis tests claimed to be positive that are falsely positive. Several groups have implemented various definitions of FDR control in the context of linkage analysis (Weller *et al.*, 1998; Heyen *et al.*, 1999; Weller, 2000; Zaykin *et al.*, 2000;

Beh *et al*., 2001; Mosig *et al*., 2001; Yekutieli, 2001; Lee *et al*., 2002; Sabatti *et al*., 2003; Varga *et al*., 2003; Bennewitz *et al*., 2004; Bernardo, 2004; Cnaani *et al*., 2004; Fernando *et al*., 2004; Simonsen & McIntyre, 2004; Zhang *et al*., 2004; Benjamini & Yekutieli, 2005).

Chen and Storey (Chen & Storey, 2006) discussed two issues with linkage statistics. First, under the global null hypothesis, linkage statistics randomly fluctuate across the genome. Both theoretical and resampling methods have been developed to address this noise (e.g. Lander & Botstein, 1989; Lander & Kruglyak, 1995). Second, there is dependence in the signal due to linkage between markers. Under certain types of dependence structures, the FDR can be controlled, though control may be conservative compared with the desired nominal FDR level (Sabatti *et al*., 2003; Benjamini & Yekutieli, 2005; Benjamini *et al*., 2006).

In the case of sparse marker maps, it may be appropriate to consider the number of markers as the effective number of tests because the distance between markers may be large enough for the markers to be effectively uncorrelated. However, the FDR can be manipulated by marker placement (Fernando *et al*., 2004; Chen & Storey, 2006). A routine practice in QTL mapping, as already mentioned, is to insert pseudomarkers and impute missing genotype data (Sen & Churchill, 2001). Inserting pseudomarkers provides the advantage of ensuring that the marker map is nearly balanced with respect to proportionally marking the regions of the genome that are and are not QTL, even if the observed markers are unbalanced. The availability of dense maps consisting of only observed markers resulting from high-throughput sequencing projects will likely eliminate the need for pseudomarkers while preserving the balance of marker maps.

In the case of dense marker maps, whether the map consists of markers only or markers supplemented with pseudomarkers, correlation between markers (and pseudomarkers) must be taken into account. Chen and Storey (Chen & Storey, 2006) proceeded under the simplification that, under the null hypothesis of no linkage to the phenotype, the chromosome is the unit of testing, implicitly assuming that linkage extends across an entire chromosome. Similarly, Benjamini and Yekutieli (Benjamini & Yekutieli, 2005) based their work on the simplification that 'for each trait all the hypotheses on a chromosome are either true or false depending on the presence of a QTL on the chromosome'.This choice ignores information provided by recombination fractions, which precludes localization of QTL on the subchromosomal scale, and is potentially biased if there are multiple QTLs on one chromosome (Kao *et al*., 1999). If meioses occur, particularly for long chromosomes or when recombination rates are high, linkage does not necessarily extend across an entire chromosome. For example, human chromosome 1 is $\sim$225 cM, corresponding to a recombination fraction of 0·50 using the Kosambi mapping function, which indicates that the ends of this chromosome are unlinked. A more general definition of the proper unit of testing, appropriate for any recombination rate, is a linkage region, i.e. a region of a chromosome in which the set of all markers (and pseudomarkers) are in tight linkage such that test statistics are positively correlated and reflect only one underlying outcome.

In this study, an expected loss function is developed for multiple QTL mapping, taking into account dependence among markers as measured by recombination fractions. Using this method, the numbers of truly significant, falsely significant, truly non-significant and falsely non-significant hypothesis tests can all be estimated. In the Bayesian framework, these quantities can be directly estimated from posterior probabilities. Then, the FDR, which is the proportion of significant tests that are falsely significant, and the false non-discovery rate (FNR), which is the proportion of non-significant tests that are falsely non-significant, can both be estimated. Finally, these two error rates can be combined into a single value, the Bayes error, and minimizing this expected loss quantity can be used as an optimality criterion for determining significance thresholds. Bayes error-based mapping procedures are presented for main effects, gene–gene interactions and gene–environment interactions.

2. Materials and methods

(i) *The multiple interacting QTL model*

For continuous traits, consider the linear regression framework of Yi *et al*. (Yi *et al*., 2007*b*) for multiple QTL models. The phenotypes of n individuals, $y=(y_1, \ldots, y_n)^T$, are expressed as

$$\mathbf{y}=\boldsymbol{\mu}+\mathbf{X}_{\mathrm{G}}\boldsymbol{\beta}_{\mathrm{G}}+\mathbf{X}_{\mathrm{GG}}\boldsymbol{\beta}_{\mathrm{GG}}+\mathbf{X}_{\mathrm{E}}\boldsymbol{\beta}_{\mathrm{E}}+\mathbf{X}_{\mathrm{GE}}\boldsymbol{\beta}_{\mathrm{GE}}+\mathbf{e}, \qquad (1)$$

in which $\boldsymbol{\mu}=(\mu, \ldots, \mu)^T$ is the overall mean; $\mathbf{X}_{\mathrm{G}}$, $\mathbf{X}_{\mathrm{GG}}$, $\mathbf{X}_{\mathrm{E}}$ and $\mathbf{X}_{\mathrm{GE}}$ are the design matrices for main effects, gene–gene interactions, environmental effects and gene–environment interactions, respectively; $\boldsymbol{\beta}_{\mathrm{G}}$, $\boldsymbol{\beta}_{\mathrm{GG}}$, $\boldsymbol{\beta}_{\mathrm{E}}$ and $\boldsymbol{\beta}_{\mathrm{GE}}$ are vectors of main effects, gene–gene interactions, environmental effects and gene–environment interactions, respectively; and $\mathbf{e}$ is a vector of independent normal errors with mean zero and variance $\sigma^2\mathbf{I}_n$. Under prior independence, any one effect or interaction is independent of all other effects and interactions. Constraints can be imposed to make interactions dependent on main effects (Yi *et al*.,

https://doi.org/10.1017/S001667230900010X Published online by Cambridge University Press

Table 1. *Discrete outcomes of testing* m-*independent hypotheses*

Hypothesis	Accept null	Reject null	Total
True null	U	V	m_0
True alternative	T	S	m_1
Total	W	R	m

2007*b*). With appropriate modifications to eqn (1), binary, categorical and ordinal traits can also be analysed (Yandell *et al*., 2007; Yi *et al*., 2007*a*).

Suppose the genome is partitioned into H loci, $\boldsymbol{\zeta}=\{\zeta_1, \ldots,\zeta_H\}$, corresponding to the set of all markers and pseudomarkers. From the perspective of model selection, a subset of $\boldsymbol{\zeta}$ explains phenotypic variation. Let $\boldsymbol{\lambda}=\{\lambda_1, \ldots, \lambda_L\}(\in \{\zeta_1, \ldots, \zeta_H\})$ be the positions of L QTL among the H loci. A vector of latent indicator variables $\boldsymbol{\gamma}$ is used to indicate which effects (main effects, gene–gene interactions and gene–environment interactions) are included in ($\gamma=1$) or excluded from ($\gamma=0$) the model (Yi, 2004; Yi *et al*., 2005, 2007*b*). Effect sizes, conditional on $\boldsymbol{\gamma}$, are distributed according to a mixture of a point mass at zero and a normal distribution (Yi & Shriner, 2008). In the Bayesian framework, the goal is to infer the posterior distribution of $(\boldsymbol{\gamma}, \boldsymbol{\lambda}_{\boldsymbol{\gamma}})$. A full description of the prior distributions and the Markov chain Monte Carlo (MCMC) algorithms can be found elsewhere (Yi, 2004; Yi *et al*., 2005, 2007*b*).

(ii) *Discrete definitions of operating characteristics*

Consider the possible outcomes when performing m hypothesis tests (Table 1). The number of true negatives is given by U, the number of false negatives is given by T, the number of true positives is given by S and the number of false positives is given by V. All four of these random numbers are unobservable. The marginal sums $m_0=U+V$, which is the number of true null hypotheses and $m_1=T+S$, which is the number of true alternative hypotheses, are likewise unobservable random numbers. The marginal sums $W=U+T$, which is the number of non-rejected hypotheses and $R=V+S$, which is the number of rejected hypotheses, are observable random numbers. The positive FDR (pFDR) is defined as

$$\text{pFDR}=E\left(\frac{V}{R}|R>0\right), \tag{2}$$

which is the expected proportion of discoveries (i.e. rejected null hypotheses) that are false, given at least one discovery (Storey, 2003). The marginal FDR (mFDR) is defined as

$$\text{mFDR}=\frac{E(V)}{E(R)}, \tag{3}$$

(Benjamini & Hochberg, 1995; Tsai *et al*., 2003). The pFDR and mFDR are asymptotically equivalent as the number of tests increases (Storey, 2003). Similarly, the positive FNR (pFNR) is defined as

$$\text{pFNR}=E\left(\frac{T}{W}|W>0\right), \tag{4}$$

which is the expected proportion on non-discoveries (i.e. accepted null hypotheses) that are false, given at least one non-discovery (Genovese & Wasserman, 2002; Storey, 2003). The marginal FNR (mFNR) is defined as

$$\text{mFNR}=\frac{E(T)}{E(W)}. \tag{5}$$

(iii) *The number of tests for main effects*

Under the global null hypothesis of no segregating QTL anywhere in the genome, equations have been derived to estimate the effective number of tests, m, under sparse and dense map cases (Lander & Botstein, 1989; Lander & Kruglyak, 1995). For sparse maps, the effective number of tests can be estimated simply by the number of markers, M. For dense maps, solve the equation (based on an Orenstein–Uhlenbeck diffusion process)

$$\alpha=(C+2Gt_\alpha)\chi^2(t_\alpha) \tag{6}$$

for t_α, in which α is the family-wide significance level (typically $\alpha=0{\cdot}05$), C is the number of chromosomes and G is the length of the genome in Morgans. For a recombinant inbred line design with selfing or with full-sib mating, G should be replaced with $2G$ or $4G$, respectively. $\chi^2(t_\alpha)$ is the one-tailed probability (i.e. the area under the curve to the right of t_α) from the distribution function for the central χ^2 distribution. The χ^2 distribution has one degree of freedom for backcross or recombinant inbred line designs (for one effect size parameter) and two degrees of freedom for an F_2 design (for two effect size parameters). The effective number of tests is given by $m=C+2Gt_\alpha$ (Lander & Botstein, 1989; Lander & Kruglyak, 1995). If one has a sufficiently dense marker map to efficiently test linkage, additional markers will be tightly linked with existing markers. Additional markers yield diminishing returns, increasing the amount of correlation more than increasing the information content (Darvasi *et al*., 1993). The average width of a test interval in cM is $100G/(C+2Gt_\alpha)$. Note that this interval defines the average linkage region, which is the proper unit of testing for dense marker maps.

https://doi.org/10.1017/S001667230900010X Published online by Cambridge University Press

Table 2. *Probabilistic outcomes for hypothesis testing*

Hypothesis	Accept null	Reject null	Total
True null	$\Pr(H=0,t\notin\Gamma)$[a]	$\Pr(H=0,t\in\Gamma)$	$\Pr(H=0)$
True alternative	$\Pr(H=1,t\notin\Gamma)$	$\Pr(H=1,t\in\Gamma)$	$\Pr(H=1)$
Total	$\Pr(t\notin\Gamma)$	$\Pr(t\in\Gamma)$	1

[a] $H=0$ indicates that the null hypothesis is true. $H=1$ indicates that the alternative hypothesis is true. $t\in\Gamma$ indicates that the test statistic is an element of the rejection region, Γ. $t\notin\Gamma$ indicates that the test statistic is not an element of the rejection region, Γ. $\Pr(\cdot)$ denotes the probability of an event. $\Pr(\cdot,\cdot)$ denotes the joint probability of two events.

(iv) *The direct posterior probability approach to FDR control*

Suppose there is a test statistic t_i for each of $i=1,\ldots,m$ tests. Define a binary indicator variable $H_i=0$ if the null hypothesis is true for the ith test and $H_i=1$ if the alternative hypothesis is true for the ith test. In the context of linkage analysis, the null hypothesis is that the tested locus is not linked to the phenotype and the alternative hypothesis is that the tested locus is linked to the phenotype. Assume that (t_i, H_i) are identically and independently distributed random variables. Suppose the rejection region Γ is held constant for testing all hypotheses. A test statistic t_i is either an element of the rejection, $t_i\in\Gamma$, or it is not, $t_i\notin\Gamma$. Let $\Pr(\cdot)$ denote a probability measure. Then, the pFDR for the ith test is

$$\text{pFDR}_i(\Gamma)=\frac{\Pr(H_i=0,t_i\in\Gamma)}{\Pr(t_i\in\Gamma)}=\Pr(H_i=0|t_i\in\Gamma) \quad (7)$$

(Table 2) (Storey, 2003).

Each locus in $\boldsymbol{\zeta}$ may affect a trait through its main effects and/or interactions with other loci (epistasis) or environmental effects (Yi *et al.*, 2005). The posterior inclusion probability of the hth locus ζ_h, $\Pr(\gamma_h=1|\mathbf{y})$, is estimated by the frequency that the locus ζ_h appears in the posterior sample given the data $\mathbf{y}$. Furthermore, $\Pr(\gamma=1|\mathbf{y})$ is a direct estimate of $\Pr(H=1)$. For a given rejection region Γ, $t_i\in\Gamma$ if t_i is greater than or equal to a defined threshold of $\Pr(\gamma=1|\mathbf{y})$. The test statistic t_i is the cumulative posterior inclusion probability for an interval. The cumulative posterior inclusion probability over a chromosomal interval bracketed by loci h_L and h_R is given by $\sum_{h=h_L}^{h_R}\Pr(\gamma_h=1|\mathbf{y})$, for an interval in which all loci have posterior probabilities in the rejection region. (Care must be taken because the cumulative posterior inclusion probability may exceed one for a sufficiently large interval, indicating the presence of more than one QTL in that interval.) The length of the interval is given by $\zeta_{hR}-\zeta_{hL}$, which corresponds to $\frac{\zeta_{h_R}-\zeta_{h_L}}{100G/(C+2Gt_\alpha)}$ tests. The probability that an interval declared positive is truly positive, $\Pr(H_i=1|t_i\in\Gamma)$, is $\sum_{h=h_L}^{h_R}\Pr(\gamma_h=1|\mathbf{y})$. The probability that an interval declared positive is falsely positive, $\Pr(H_i=0|t_i\in\Gamma)$, is given by $1-\Pr(H_i=1|t_i\in\Gamma)$, or $1-\sum_{h=h_L}^{h_R}\Pr(\gamma_h=1|\mathbf{y})$. The pFDR can then be estimated by the sum (over all intervals declared significant) of $1-\sum_{h=h_L}^{h_R}\Pr(\gamma_h=1|\mathbf{y})$ divided by the sum (over all intervals declared significant) of $\frac{\zeta_{h_R}-\zeta_{h_L}}{100G/(C+2Gt_\alpha)}$ (Newton *et al.*, 2004; Storey *et al.*, 2005).

(v) *The direct posterior probability approach to FNR control*

For a given significance threshold Γ, $t_i\notin\Gamma$ if t_i is less than a defined threshold of $\Pr(\gamma=1|\mathbf{y})$. The probability that an interval declared negative is falsely negative,

$$\text{pFNR}_i(\Gamma)=\frac{\Pr(H_i=1,t_i\notin\Gamma)}{\Pr(t_i\notin\Gamma)}=\Pr(H_i=1|t_i\notin\Gamma) \quad (8)$$

(Table 2), is given directly by $\sum_{h=h_L}^{h_R}\Pr(\gamma_h=1|\mathbf{y})$, for an interval in which all loci have posterior probabilities not in the rejection region. Thus, the pFNR is estimated by the sum of $\sum_{h=h_L}^{h_R}\Pr(\gamma_h=1|\mathbf{y})$ for all tests not declared significant divided by $\frac{\zeta_{h_R}-\zeta_{h_L}}{100G/(C+2Gt_\alpha)}$ for all tests not declared significant.

(vi) *Probabilistic definitions of operating characteristics*

The eight operating characteristics for testing a hypothesis can be written in probabilistic form as follows (Table 2):

$$\text{Type I error rate}=\frac{\Pr(H=0,t\in\Gamma)}{\Pr(H=0)}=\Pr(t\in\Gamma|H=0),$$

$$\text{Type II error rate}=\frac{\Pr(H=1,t\notin\Gamma)}{\Pr(H=1)}=\Pr(t\notin\Gamma|H=1),$$

$$\text{Sensitivity}=\frac{\Pr(H=1,t\in\Gamma)}{\Pr(H=1)}=\Pr(t\in\Gamma|H=1),$$

$$\text{Specificity}=\frac{\Pr(H=0,t\notin\Gamma)}{\Pr(H=0)}=\Pr(t\notin\Gamma|H=0),$$

$$\text{Positive predictive value}=\frac{\Pr(H=1,t\in\Gamma)}{\Pr(t\in\Gamma)}=\Pr(H=1|t\in\Gamma),$$

$$\text{Negative predictive value}=\frac{\Pr(H=0,t\notin\Gamma)}{\Pr(t\notin\Gamma)}=\Pr(H=0|t\notin\Gamma),$$

https://doi.org/10.1017/S001667230900010X Published online by Cambridge University Press

$$\mathrm{FDR} = \frac{\Pr(H=0, t \in \Gamma)}{\Pr(t \in \Gamma)} = \Pr(H=0 | t \in \Gamma),$$

$$\mathrm{FNR} = \frac{\Pr(H=1, t \notin \Gamma)}{\Pr(t \notin \Gamma)} = \Pr(H=1 | t \notin \Gamma).$$

Sensitivity is equivalent to average power. Note that the positive predictive value and the FDR are complementary, as are the negative predictive value and the FNR. Coverage is the proportion of replicates for which an interval contains the true QTL location. Accuracy is a function of the difference between the estimated peak location and the true QTL location. Precision is an inverse measure of the width of an interval. Estimates of each of these operating characteristics exist for each of the m tests. Averaging over all tests can be done either marginally (the ratio of averages) or jointly (the average of ratios); in this study, averaging was done marginally.

(vii) *Bayes error*

One possible optimality criterion is to minimize the expected loss, BE, by minimizing the weighted average of the pFDR and the pFNR, $(w)\mathrm{pFDR} + (1-w)\mathrm{pFNR}$, for a user-defined cost w (Genovese & Wasserman, 2002; Storey, 2003; Chen & Sarkar, 2006). Under this scheme, the significance threshold, which in turn defines the rejection region, is objectively determined by the data rather than being subjectively set by the user. Due to complementarities, minimizing a weighted average of the FDR and FNR is equivalent to maximizing a weighted average of the positive and negative predictive values.

(viii) *Extensions to gene–gene and gene–environment interactions*

Whereas mapping main effects involves optimization over a two-dimensional curve of posterior probability as a function of genomic location, mapping gene–gene interactions involves optimization over a three-dimensional surface of posterior probability as a function of two genomic locations. Under backcross or recombinant inbred line designs, there is one type of gene–gene interaction effect. Under an F_2 design, there are four types of gene–gene interaction effects, referred to as additive–additive, additive–dominance, dominance–additive and dominance–dominance effects. The effective number of tests increases to $\binom{m}{2}$. Storey *et al.* (Storey *et al.*, 2005) developed a sequential approach for detecting gene–gene interactions that detects interactions only if the main effects for both loci are significant. The implementation described herein allows for interaction effects independent of main effect sizes, which increases the type I error rate but also increases power for pairs of loci with primarily interaction effects. Mapping gene–environment interactions is analogous to mapping main effects, with one two-dimensional posterior probability curve for each interacting environmental effect.

(ix) *Implementation in R/qtlbim*

The freely available QTL mapping package R/qtlbim is an extensible, interactive environment for Bayesian analysis of multiple interacting QTL in experimental crosses (Yandell *et al.*, 2007). It provides several efficient MCMC algorithms for evaluating posterior probabilities of genetic architectures, i.e. the number and locations of QTLs, main effects and gene–gene and gene–environment interactions. R/qtlbim provides tools to monitor mixing behaviour and convergence of the simulated Markov chain, and provides extensive informative graphical and numerical summaries of the MCMC output to infer and interpret the genetic architecture of complex traits. Code implementing Bayes error calculations will be included in R/qtlbim, which facilitates the general usage of Bayesian methodologies for genome-wide interacting QTL analysis.

(x) *Simulation study*

Data simulation and Bayesian QTL mapping were performed using the R package qtlbim (Yandell *et al.*, 2007). The phenotypic trait was assumed to be normally distributed. An F_2 design was simulated. The simulated genome consisted of four autosomal chromosomes with a total length of 400 cM. Using the approximation for dense maps (Lander & Botstein, 1989; Lander & Kruglyak, 1995) presented in eqn (6) for $\alpha = 0{\cdot}05$, these choices yielded 130 effective tests, for which the comparable LOD (likelihood of odds) threshold, equal to $\chi^2(t_\alpha)2\ln 10$, was 3·41. Under the global null hypothesis, the average test spanned $\sim 3{\cdot}1$ cM. Pseudomarkers were inserted at 1 cM intervals using the Kosambi mapping function and genotype probabilities were estimated using the multipoint method. The experimental conditions are presented in Table 3. For every set of conditions, 1000 independent replicate data sets were generated and analysed.

The prior number of QTLs in the MCMC analysis was the true value, except for Experiment 10 in which the prior number was misspecified as five (greater than the number of chromosomes). For a correctly constructed MCMC algorithm, the Markov chain will converge to a unique stationary distribution that is the target posterior distribution, regardless of the initialization values, provided that a sufficiently long chain is run (Geyer, 1992). The main effect of misspecified priors is to increase the burn-in period before the Markov chain converges to the stationary

https://doi.org/10.1017/S001667230900010X Published online by Cambridge University Press

Table 3. *Simulation conditions*

Experiment	Chromosome length	Marker map[a]	Estimated number of tests	Sample size	Main effect QTL[b]	$G \times G$ QTL[c]	$G \times E$ QTL[d]
1	100-100-100-100	10-10-10-10	130	250	(1, 45, 0·25, 0·25)		
2	150-75-125-50	10-10-10-10	130	250	(1, 45, 0·25, 0·25)		
3	150-75-125-50	14-8-12-6	130	250	(1, 45, 0·25, 0·25)		
4	100-100-100-100	50-10-10-10	130	250	(1, 45, 0·25, 0·25)		
5	100-100-100-100	20-20-20-20	130	250	(1, 45, 0·25, 0·25)		
6	100-100-100-100	50-10-10-10	130	250	(1, 45, 0·25, 0·25)		
7	100-100-100-100	10-10-10-10	130	250	(1, 0·5, 0·25, 0·25)		
8	100-100-100-100	10-10-10-10	130	250	(1, 45, 0·5, 0·25)		
9	100-100-100-100	10-10-10-10	130	500	(1, 45, 0·25, 0·25)		
10[e]	100-100-100-100	10-10-10-10	130	250	(1, 45, 0·25, 0·25)		
11	100-100-100-100	10-10-10-10	130	250	(1, 45, 0·25, 0·25) (2, 60, 0·25, 0·25)		
12	100-100-100-100	10-10-10-10	130	250	(1, 45, 0·25, 0·25) (2, 45, 0·25, 0·25) (3, 45, 0·25, 0·25) (4, 45, 0·25, 0·25)		
13	100-100-100-100	10-10-10-10	14 993	250		(1, 45, 2, 60, 0·5, 0, 0, 0)	
14	100-100-100-100	10-10-10-10	14 993	250		(1, 45, 2, 60, 0·63, 0, 0, 0)	
15	100-100-100-100	10-10-10-10	14 993	250		(1, 45, 2, 60, 0·5, 0, 0, 0) (3, 45, 4, 60, 0·5, 0, 0, 0)	
16	100-100-100-100	10-10-10-10	14 993	500		(1, 45, 2, 60, 0·5, 0, 0, 0)	
17	100-100-100-100	10-10-10-10	130	250			(1, 45, 0·5, 0·5)
18	100-100-100-100	10-10-10-10	130	250			(1, 45, 0·7, 0·7)
19	100-100-100-100	10-10-10-10	130	250			(1, 45, 0·5, 0·5) (2, 60, 0·5, 0·5)
20	100-100-100-100	10-10-10-10	130	500			(1, 45, 0·5, 0·5)
21	100-100-100-100	10-10-10-10	130-14 993-130[f]	250	(1, 45, 0·25, 0·25)	(1, 45, 2, 60, 0·5, 0, 0, 0)	(1, 45, 0·5, 0·5)
22	100-100-100-100	10-10-10-10	159	250	(1, 45, 0·25, 0·25)		
23	100-100-100-100	10-10-10-10	199	250	(1, 45, 0·25, 0·25)		

[a] The notation for marker maps indicates the number of markers on each chromosome. For Experiment 4, the 50 markers are evenly distributed across chromosome 1. For Experiment 6, 40 markers are evenly distributed between 32·5 and 57·5 cM and the other 10 markers are evenly distributed across the rest of the chromosome.
[b] The notation for main effects is (chromosome, location, additive effect and dominance effect).
[c] The notation for gene–gene interactions is (chromosome 1, location 1, chromosome 2, location 2, additive–additive effect, additive–dominance effect, dominance–additive effect and dominance–dominance effect).
[d] The notation for gene–environment interactions is (chromosome, location, additive effect and dominance effect).
[e] In Experiment 10, the prior number of QTLs was misspecified as five.
[f] In Experiment 21, the estimated number of tests is given separately for main effects, gene–gene interactions and gene–environment interactions.

Table 4. *Mean operating characteristics for mapping main effects as a function of chromosome lengths and the marker map*

	Experiment					
	1	2	3	4	5	6
Significance threshold	0·011	0·011	0·012	0·010	0·011	0·008
pFDR	0·11	0·11	0·10	0·16	0·15	0·17
pFNR	0·0090	0·0105	0·0106	0·0081	0·0087	0·0074
Bayes error	0·060	0·062	0·058	0·085	0·080	0·087
Sensitivity	0·86	0·87	0·85	0·88	0·87	0·89
Specificity	0·98	0·94	0·96	0·96	0·96	0·95
Positive predictive value	0·89	0·89	0·90	0·84	0·85	0·83
Negative predictive value	0·99	0·99	0·99	0·99	0·99	0·99
Proportion of true null hypotheses	0·90	0·84	0·86	0·87	0·88	0·86
Proportion of true alternative hypotheses	0·10	0·16	0·14	0·13	0·12	0·14
Number of false positive intervals	2·1	3·9	3·2	4·3	3·7	4·8
False positive interval width	11·9	11·5	11·3	8·9	9·0	9·1
False positive interval posterior probability	0·19	0·14	0·15	0·12	0·13	0·11
Coverage	0·80	0·76	0·82	0·81	0·80	0·81
True positive peak location	44·2	46·5	45·4	44·4	44·7	44·5
True positive interval width	24·6	34·4	28·0	28·4	26·5	28·5
True positive interval posterior probability	0·65	0·62	0·66	0·67	0·64	0·73

distribution. The MCMC algorithm was run for 400 000 iterations, with a thinning value of 20 and no burn-in. The cost in the Bayes error function was 0·5.

3. Results

Seven sets of simulations were performed. The first set of simulations was designed to investigate stability to chromosomal lengths and the marker map (Table 4). In Experiment 2 compared with Experiment 1, the lengths of the chromosomes differed but the number of markers on each chromosome was the same. This had the effect of increasing the intermarker distance on the chromosome with the true QTL, so that the estimation of the QTL location was less precise. In Experiment 3 compared with Experiment 1, the lengths of the chromosomes differed and the number of markers differed such that the average intermarker distance was the same. In Experiment 4 compared with Experiment 1, chromosome 1 (on which was one QTL) was saturated with markers. In Experiment 5 compared with Experiment 1, marker density was increased evenly across all chromosomes. In Experiment 6 compared with Experiment 1, only the interval of chromosome 1 surrounding the true QTL was saturated. The total number of markers in Experiments 4–6 was the same. The significance threshold, the pFDR, and the pFNR were not sensitive to chromosome lengths or marker placement.

The second set of simulations was designed to investigate mapping main effects as a function of QTL location, effect size and sample size (Table 5). When a QTL was located close to the edge of a chromosome (Experiment 7), the interval for the QTL location was truncated, leading to smaller widths and posterior probabilities. The estimation of the true QTL location was less accurate, with the bias being away from the edge of the chromosome. Also, it was more difficult to differentiate false versus true positive QTL, as can be seen by larger pFDR estimates and smaller positive predictive values. There were more false positive QTL, although the average posterior probability for a false positive QTL was smaller. With a larger effect size (Experiment 8), the significance threshold was markedly increased, implying a more stringent test. The pFDR was much smaller and the positive predictive value was much larger for a QTL with a larger effect. Sensitivity was reduced but specificity improved, consistent with increased stringency. Coverage and the true positive posterior probability both increased and the true positive width decreased with larger effects. The number of false positive QTLs decreased but the false positive posterior probability and the false positive width both increased. With a larger sample size (Experiment 9), similar to increasing the effect size, the significance threshold was increased, implying a more stringent test. The pFDR was much smaller and the positive predictive value was much larger for a larger sample size. Sensitivity was reduced but specificity increased, consistent with a more stringent test. Coverage and the true positive posterior probability both increased and the true positive width decreased with larger effects. The number of false positive QTLs decreased but the false positive posterior probability and the false positive width both increased. A misspecified prior (Experiment 10) had no effect.

https://doi.org/10.1017/S001667230900010X Published online by Cambridge University Press

Table 5. *Mean operating characteristics for mapping main effects as a function of QTL location, effect size, sample size and prior (mis)specification*

	Experiment				
	1	7	8	9	10
Significance threshold	0·011	0·009	0·018	0·013	0·010
pFDR	0·11	0·18	0·031	0·047	0·11
pFNR	0·0090	0·0082	0·0097	0·0076	0·0089
Bayes error	0·060	0·092	0·021	0·027	0·061
Sensitivity	0·86	0·86	0·77	0·85	0·87
Specificity	0·98	0·96	1·00	0·99	0·98
Positive predictive value	0·89	0·82	0·97	0·95	0·89
Negative predictive value	0·99	0·99	0·99	0·99	0·99
Proportion of true null hypotheses	0·90	0·89	0·94	0·93	0·89
Proportion of true alternative hypotheses	0·10	0·11	0·06	0·07	0·11
Number of false positive intervals	2·1	3·5	0·4	0·7	2·4
False positive interval width	11·9	11·1	14·8	13·4	11·5
False positive interval probability	0·19	0·13	0·45	0·27	0·17
Coverage	0·80	0·83	0·94	0·94	0·80
True positive peak location	44·2	2·6	43·9	43·9	44·4
True positive interval width	24·6	15·5	14·6	20·2	26·8
True positive interval posterior probability	0·65	0·46	0·93	0·85	0·63

Table 6. *Mean operating characteristics for mapping main effects as a function of QTL number*

	Experiment		
	1	11	12
Significance threshold	0·011	0·012	0·011
pFDR	0·11	0·084	0·073
pFNR	0·0090	0·0095	0·0099
Bayes error	0·060	0·047	0·042
Sensitivity	0·86	0·88	0·91
Specificity	0·98	0·98	0·97
Positive predictive value	0·89	0·92	0·93
Negative predictive value	0·99	0·99	0·99
Proportion of true null hypotheses	0·90	0·86	0·77
Proportion of true alternative hypotheses	0·10	0·14	0·23
Number of false positive intervals	2·1	1·8	1·9
False positive interval width	11·9	12·3	10·9
False positive interval probability	0·19	0·22	0·21
Coverage	0·80	0·69	0·67
True positive peak location	44·2	43·8/58·6	43·9
True positive interval width	24·6	27·0	27·4
True positive interval posterior probability	0·65	0·68	0·67

The third set of simulations was designed to investigate mapping multiple QTLs (Table 6). Increases in the amount of true positive signal due to multiple true QTLs had effects as expected on all of the operating characteristics (Experiments 11 and 12). The pFDR and the number of false positive intervals were smaller with multiple QTLs. Power increased as the number of QTLs increased, or equivalently, as the proportion of true alternative hypotheses increased.

The fourth set of simulations was designed to investigate mapping gene–gene interactions (Table 7). Not surprisingly, the pFDR was much larger for gene–gene interactions than for main effects (Experiment 13). The vast majority of claims of significance were in fact false positives, but with vanishingly small posterior probabilities. On average, it was $\sim$100 times easier to distinguish false positives from true positives for gene–gene interactions than for main effects. Increasing the effect size (Experiment 14), the number of gene–gene interactions (Experiment 15) or the sample size (Experiment 16) had the same consequences as for main effects. As expected, large sample

https://doi.org/10.1017/S001667230900010X Published online by Cambridge University Press

Table 7. *Mean operating characteristics for mapping gene–gene interactions as a function of effect size, QTL number and sample size*

	Experiment			
	13	14	15	16
Significance threshold	0·0010	0·0014	0·0014	0·0015
pFDR	0·38	0·20	0·26	0·10
pFNR	0·0001	0·0001	0·0001	0·0001
Bayes error	0·19	0·10	0·13	0·051
Sensitivity	0·68	0·82	0·78	0·90
Specificity	0·99	1·00	1·00	1·00
Positive predictive value	0·62	0·80	0·74	0·90
Negative predictive value	1·00	1·00	1·00	1·00
Proportion of true null hypotheses	0·999	0·998	0·998	0·998
Proportion of true alternative hypotheses	0·001	0·002	0·002	0·002
Number of false positive intervals	581	259	300	82
False positive interval width 1	0·21	0·22	0·24	0·26
False positive interval width 2	0·21	0·22	0·24	0·27
False positive interval probability	0·0005	0·0009	0·0010	0·0023
Coverage 1	0·45	0·66	0·33	0·78
True positive peak location 1	43·7	43·6	43·6	44·1
True positive interval width 1	18·1	15·4	17·7	14·6
True positive peak location 2	58·4	59·0	59·0	59·0
True positive interval width 2	19·9	17·8	19·4	16·8
True positive interval posterior probability 1	0·34	0·52	0·40	0·63
Coverage 2			0·32	
True positive peak location 3			44·4	
True positive interval width 3			16·4	
True positive peak location 4			58·9	
True positive interval width 4			18·1	
True positive interval posterior probability 2			0·34	

sizes were critical for obtaining high levels of predictive values and posterior probabilities for true hypotheses (Experiment 16).

The fifth set of simulations was designed to investigate mapping gene–environment interactions (Table 8). Under the specific conditions tested, gene–environment interactions were easier to map than main effects (Experiments 17–20). Otherwise, all of the trends observed for main effects and gene–gene interactions also held for gene–environment interactions. The sixth set of simulations was designed to investigate simultaneous mapping of main effects, gene–gene interactions and gene–environment interactions (Table 9). Error estimates were slightly larger in the presence of all three types of genetics effects, but there was no impediment to this type of joint analysis (Experiment 21).

The seventh set of simulations was designed to investigate the effect of the estimated effective number of tests (Table 10). This was accomplished by solving eqn (6) for $\alpha = 0{\cdot}05$ (Experiment 1), $\alpha = 0{\cdot}01$ (Experiment 22) and $\alpha = 0{\cdot}001$ (Experiment 23). As the estimated effective number of tests increased, the estimated values of the various operating characteristics converged to their respective asymptotic limits.

4. Discussion

For Mendelian traits, statistical hypothesis testing during QTL mapping is complicated by the fact that genome-wide testing involves multiple comparisons. For complex traits with multiple QTLs, a second problem is induced by the fact that some unknown proportion of tests, but usually many more than just one test, are expected to be rejected. Under the latter case, traditional corrections for multiple comparisons, such as Bonferroni corrections, are generally conservative with respect to both multiplicity and correlation. Recently developed methods, such as FDR-controlling methods, retain more power, ideally increase in power with an increasing number of QTLs, and account for dependence among tests. In this study, optimizing predictive value led to more conservative control ($1 - \text{specificity} < 0{\cdot}05$) and suboptimal coverage (coverage $< 0{\cdot}95$) than expected by a traditional experiment-wide significance level of 0·05.

Insertion of pseudomarkers at even intervals yields a nearly balanced marker map, such that the false nulls and true nulls are proportionally represented. If the map of markers and pseudomarkers is sufficiently dense, then tests of markers or intervals between

https://doi.org/10.1017/S001667230900010X Published online by Cambridge University Press

Table 8. *Mean operating characteristics for mapping gene–environment interactions as a function of effect size, QTL number and sample size*

	Experiment			
	17	18	19	20
Significance threshold	0·012	0·015	0·012	0·012
pFDR	0·056	0·021	0·047	0·018
pFNR	0·0056	0·0055	0·0061	0·0037
Bayes error	0·031	0·013	0·026	0·011
Sensitivity	0·88	0·85	0·92	0·89
Specificity	0·99	1·00	0·99	1·00
Positive predictive value	0·94	0·98	0·95	0·98
Negative predictive value	0·99	0·99	0·99	1·00
Proportion of true null hypotheses	0·93	0·95	0·88	0·96
Proportion of true alternative hypotheses	0·07	0·05	0·12	0·04
Number of false positive intervals	0·7	0·2	0·8	0·2
False positive interval width	13·6	14·5	13·9	15·0
False positive interval probability	0·26	0·46	0·33	0·47
Coverage 1	0·93	0·97	0·87	0·99
True positive peak location 1	43·8	43·9	43·9	43·8
True positive interval width 1	19·8	14·1	21·1	13·4
True positive interval posterior probability 1	0·82	0·95	0·84	0·96
Coverage 2			0·80	
True positive peak location 2			58·9	
True positive interval width 2			23·8	
True positive interval posterior probability 2			0·82	

Table 9. *Simultaneous mapping of main effects, gene–gene interactions and gene–environment interactions* (*Experiment 21*)

	Main effect	Gene–gene interaction	Gene–environment interaction
Significance threshold	0·0095	0·00096	0·014
pFDR	0·16	0·44	0·091
pFNR	0·0084	0·0001	0·0087
Bayes error	0·083	0·22	0·050
Sensitivity	0·88	0·64	0·86
Specificity	0·96	0·99	0·98
Positive predictive value	0·84	0·56	0·91
Negative predictive value	0·99	1·00	0·99
Proportion of true null hypotheses	0·87	0·999	0·88
Proportion of true alternative hypotheses	0·13	0·001	0·12
Number of false positive intervals	3·8	636	2·3
False positive interval width 1	10·4	0·22	13·0
False positive interval width 2		0·22	
False positive interval probability	0·13	0·0005	0·21
Coverage	0·75	0·37	0·85
True positive peak location 1	44·0	59·4	59·3
True positive interval width 1	26·7	19·8	23·6
True positive peak location 2		44·1	
True positive interval width 2		18·1	
True positive interval posterior probability	0·52	0·30	0·74

markers are correlated through linkage. Thus, it is crucial to account for linkage by estimating the effective number of tests. The 'width' of a test defines a proper unit of testing that accounts for all markers and pseudomarkers in tight linkage. This test definition prevents the manipulation of the FDR by marker placement (Fernando *et al*., 2004; Chen & Storey, 2006). Herein, I describe a new implementation of a Bayesian FDR-controlling method. The proposed method extends previous ones by controlling the overall error rate by simultaneously controlling the FDR and the FNR. The proposed method

https://doi.org/10.1017/S001667230900010X Published online by Cambridge University Press

Table 10. *Mean operating characteristics for mapping main effects as a function of the estimated effective number of tests*

	Experiment		
	1	22	23
Estimated number of tests	130	159	199
Significance threshold	0·011	0·011	0·011
pFDR	0·11	0·095	0·077
pFNR	0·009	0·0074	0·0059
Bayes error	0·060	0·051	0·041
Sensitivity	0·86	0·88	0·90
Specificity	0·98	0·98	0·98
Positive predictive value	0·89	0·91	0·92
Negative predictive value	0·99	0·99	0·99
Proportion of true null hypotheses	0·90	0·89	0·89
Proportion of true alternative hypotheses	0·10	0·11	0·11
Number of false positive intervals	2·1	2·3	2·4
False positive interval width	11·9	11·2	10·4
False positive interval posterior probability	0·19	0·17	0·15
Coverage	0·80	0·82	0·82
True positive peak location	44·2	43·6	44·0
True positive interval width	24·6	24·4	25·2
True positive interval posterior probability	0·65	0·64	0·64

accounts for the effective number of tests in a genome-wide linkage scan, although the relevant parameter α is merely a nuisance parameter and could be eliminated by integration. Critically, the proposed method accounts for the correlation among markers within a linkage region. The proposed method is applicable for continuous, binary and ordinal traits and for multiple interacting QTL mapping, i.e. with an arbitrary number of QTLs with arbitrary genetic effects.

For multiple QTLs mapping, it is unlikely that any single model sampled by the MCMC algorithm would correctly contain all QTLs. This problem is addressable through the use of Bayesian model averaging to estimate parameters for each QTL marginally. Also, the overall Bayes error rate decreased with multiple QTLs. With multiple QTLs, a greater proportion of alternative hypotheses are true, and pFDR-controlling methods tend to be more powerful in these situations (Storey, 2002). Since different types of genetic effects can be estimated with differing levels of accuracy and precision, it seems reasonable to optimize mapping for main effects, gene–gene interactions and gene–environment interactions separately.

The most commonly used summary statistic in Bayesian hypothesis testing is the Bayes factor, defined as the ratio of the posterior odds to the prior odds (Jeffreys, 1961; Kass & Raftery, 1995). Equivalently, the Bayes factor is the ratio of the marginal likelihoods of the data under the two hypotheses. The Bayes factor can be seen as the Bayesian counterpart to the likelihood ratio statistic, which is based on maximization (rather than integration) in frequentist analysis. Under certain regularity conditions, the Bayesian information criterion (Schwarz, 1978) is approximately equal to $-2\ln$(Bayes factor) (Kass & Wasserman, 1995). Thus, Bayes factors can be useful not only in a manner similar to LOD scores but also in the model selection framework.

On the other hand, there are disadvantages with using Bayes factors. First, the interpretation of Bayes factors, just as with the interpretation of likelihood ratio statistics in the frequentist framework, depends on a subjective significance threshold (Jeffreys, 1961; Kass & Raftery, 1995). Increasing the stringency of the threshold will increase specificity at the cost of reducing sensitivity. Minimizing the expected loss function provides an objective, data-driven solution to this problem. Second, Bayes factors depend critically upon the prior distributions. Whereas this dependency has not been particularly problematic for single trait analysis, it appears to be problematic for testing pleiotropy in multiple trait analysis (Banerjee *et al*., 2008). The prior probability that a locus affects none of the traits, i.e., the intersection of all of the null hypotheses, tends to become vanishingly small as the number of traits increases, thereby potentially yielding enormously large Bayes factors that are unstable and difficult to calibrate. In this context, analysing the posterior probability profile appears to be more tractable than analysing Bayes factors (Banerjee *et al*., 2008).

False positive error rates in Bayesian analyses can be affected by the specification of prior distributions. The posterior distribution of the number of QTLs is influenced by the prior distribution of the number of QTLs, whereas the Bayes factor tends to be relatively

https://doi.org/10.1017/S001667230900010X Published online by Cambridge University Press

insensitive to the prior distribution of the number of QTLs (Yi & Shriner, 2008). In contrast, the prior distribution of effect sizes tends to more strongly affect false positive error rates. The point-normal prior induces shrinkage of effect sizes towards zero, which has the advantages of reducing model space by removing most loci from the model and reducing the false positive error rate, independent of whether posterior inference is made through Bayes factors or the Bayes error.

The merit of optimizing prediction rather than classification remains an open question. In this study, optimization of prediction tended to be conservative, as evident by sub-optimal coverage and large specificities. Maximizing the positive predictive value sacrifices detection of smaller effects in favour of higher positive predictive values for larger effects. Thus, optimizing prediction may have particular benefit in genome-wide analysis by improving replicability of findings.

This research was supported by National Institutes of Health grant DK062710 and the Intramural Research Program of the National Human Genome Research Institute, National Institutes of Health. I thank the Enabling Technology Laboratory in the Department of Mechanical Engineering, University of Alabama at Birmingham for assistance with high performance computing. I thank Howard Wiener and Grier Page for their suggestions, Śaunak Sen for critical discussion and the anonymous reviewers for their helpful comments.

References

Banerjee, S., Yandell, B. S. & Yi, N. (2008). Bayesian quantitative trait loci mapping for multiple traits. *Genetics* **179**, 2275–2289.

Beh, K. J., Callaghan, M. J., Leish, Z., Hulme, D. J., Lenane, I. & Maddox, J. F. (2001). A genome scan for QTL affecting fleece and wool traits in Merino sheep. *Wool Technology and Sheep Breeding* **49**, 88–97.

Benjamini, Y. & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B* **57**, 289–300.

Benjamini, Y., Krieger, A. M. & Yekutieli, D. (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika* **93**, 491–507.

Benjamini, Y. & Yekutieli, D. (2005). Quantitative trait loci analysis using the false discovery rate. *Genetics* **171**, 783–790.

Bennewitz, J., Reinsch, N., Guiard, V., Fritz, S., Thomsen, H., Looft, C., Kühn, C., Schwerin, M., Weimann, C., Erhardt, G., Reinhardt, F., Reents, R., Boichard, D. & Kalm, E. (2004). Multiple quantitative trait loci mapping with cofactors and application of alternative variants of the false discovery rate in an enlarged granddaughter design. *Genetics* **168**, 1019–1027.

Bernardo, R. (2004). What proportion of declared QTL in plants are false? *Theoretical and Applied Genetics* **109**, 419–424.

Chen, J. & Sarkar, S. K. (2006). A Bayesian determination of threshold for identifying differentially expressed genes in microarray experiments. *Statistics in Medicine* **25**, 3174–3189.

Chen, L. & Storey, J. D. (2006). Relaxed significance criteria for linkage analysis. *Genetics* **173**, 2371–2381.

Cnaani, A., Zilberman, N., Tinman, S., Hulata, G. & Ron, M. (2004). Genome-scan analysis for quantitative trait loci in an F_2 tilapia hybrid. *Molecular Genetics and Genomics* **272**, 162–172.

Darvasi, A., Weinreb, A., Minke, V., Weller, J. I. & Soller, M. (1993). Detecting marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. *Genetics* **134**, 943–951.

Fernando, R. L., Nettleton, D., Southey, B. R., Dekkers, J. C., Rothschild, M. F. & Soller, M. (2004). Controlling the proportion of false positives in multiple dependent tests. *Genetics* **166**, 611–619.

Genovese, C. & Wasserman, L. (2002). Operating characteristics and extensions of the false discovery rate procedure. *Journal of the Royal Statistical Society Series B* **64**, 499–517.

Geyer, C. J. (1992). Practical Markov chain Monte Carlo. *Statistical Science* **7**, 473–483.

Heyen, D. W., Weller, J. I., Ron, M., Band, M., Beever, J. E., Feldmesser, E., Da, Y., Wiggans, G. R., Van Raden, P. M. & Lewin, H. A. (1999). A genome scan for QTL influencing milk production and health traits in dairy cattle. *Physiological Genomics* **1**, 165–175.

Jeffreys, H. (1961). *Theory of Probability*, 3rd edn. Oxford, UK: Oxford University Press.

Kao, C. H., Zeng, Z. B. & Teasdale, R. D. (1999). Multiple interval mapping for quantitative trait loci. *Genetics* **152**, 1203–1216.

Kass, R. E. & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association* **90**, 773–795.

Kass, R. E. & Wasserman, L. (1995). A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion. *Journal of the American Statistical Association* **90**, 928–934.

Lander, E. & Kruglyak, L. (1995). Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nature Genetics* **11**, 241–247.

Lander, E. S. & Botstein, D. (1989). Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **121**, 185–199.

Lee, H., Dekkers, J. C., Soller, M., Malek, M., Fernando, R. L. & Rothschild, M. F. (2002). Application of the false discovery rate to quantitative trait loci interval mapping with multiple traits. *Genetics* **161**, 905–914.

Lynch, M. & Walsh, B. (1998). *Genetics and Analysis of Quantitative Traits*. Sunderland, MA: Sinauer Associates, Inc.

Mosig, M. O., Lipkin, E., Khutoreskaya, G., Tchourzyna, E., Soller, M. & Friedmann, A. (2001). A whole genome scan for quantitative trait loci affecting milk protein percentage in Israeli-Holstein cattle, by means of selective milk DNA pooling in a daughter design, using an adjusted false discovery rate criterion. *Genetics* **157**, 1683–1698.

Newton, M. A., Noueiry, A., Sarkar, D. & Ahlquist, P. (2004). Detecting differential gene expression with a semiparametric hierarchical mixture method. *Biostatistics* **5**, 155–176.

Sabatti, C., Service, S. & Freimer, N. (2003). False discovery rate in linkage and association genome screens for complex disorders. *Genetics* **164**, 829–833.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics* **6**, 461–464.

Sen, Ś. & Churchill, G. A. (2001). A statistical framework for quantitative trait mapping. *Genetics* **159**, 371–387.

Simonsen, K. L. & McIntyre, L. M. (2004). Using alpha wisely: improving power to detect multiple QTL. *Statistical Applications in Genetics and Molecular Biology* **3**, Article 1.

Storey, J. D. (2002). A direct approach to false discovery rates. *Journal of the Royal Statistical Society Series B* **64**, 479–498.

Storey, J. D. (2003). The positive false discovery rate: a Bayesian interpretation and the *q*-value. *The Annals of Statistics* **31**, 2013–2035.

Storey, J. D., Akey, J. M. & Kruglyak, L. (2005). Multiple locus linkage analysis of genomewide expression in yeast. *PLoS Biology* **3**, e267.

Tsai, C.-A., Hsueh, H.-M. & Chen, J. J. (2003). Estimation of false discovery rates in multiple testing: application to gene microarray data. *Biometrics* **59**, 1071–1081.

Varga, L., Müller, G., Szabó, G., Pinke, O., Korom, E., Kovacs, B., Patthyc, L. & Sollerd, M. (2003). Mapping modifiers affecting muscularity of the myostatin mutant ($Mstn^{Cmpt\text{-}dl1Abc}$) compact mouse. *Genetics* **165**, 257–267.

Weller, J. I. (2000). Using the false discovery rate approach in the genetic dissection of complex traits: a response to Zaykin *et al*. *Genetics* **154**, 1918.

Weller, J. I., Song, J. Z., Heyen, D. W., Lewin, H. A. & Ron, M. (1998). A new approach to the problem of multiple comparisons in the genetic dissection of complex traits. *Genetics* **150**, 1699–1706.

Yandell, B. S., Mehta, T., Banerjee, S., Shriner, D., Venkataraman, R., Moon, J. Y., Neely, W. W., Wu, H., von Smith, R, & Yi, N. (2007). R/qtlbim: QTL with Bayesian interval mapping in experimental crosses. *Bioinformatics* **23**, 641–643.

Yekutieli, D. (2001). Theoretical results needed for applying the false discovery rate in statistical problems. PhD thesis, Department of Statistics and Operations Research, Tel Aviv University, Tel Aviv, Israel.

Yi, N. (2004). A unified Markov chain Monte Carlo framework for mapping multiple quantitative trait loci. *Genetics* **167**, 967–975.

Yi, N., Banerjee, S., Pomp, D. & Yandell, B. S. (2007 *a*). Bayesian mapping of genomewide interacting quantitative trait loci for ordinal traits. *Genetics* **176**, 1855–1864.

Yi, N. & Shriner, D. (2008). Advances in Bayesian multiple quantitative trait loci mapping in experimental crosses. *Heredity* **100**, 240–252.

Yi, N., Shriner, D., Banerjee, S., Mehta, T., Pomp, D. & Yandell, B. S. (2007 *b*). An efficient Bayesian model selection approach for interacting quantitative trait loci models with many effects. *Genetics* **176**, 1865–1877.

Yi, N., Yandell, B. S., Churchill, G. A., Allison, D. B., Eisen, E. J. & Pomp, D. (2005). Bayesian model selection for genome-wide epistatic quantitative trait loci analysis. *Genetics* **170**, 1333–1344.

Zaykin, D. V., Young, S. S. & Westfall, P. H. (2000). Using the false discovery rate approach in the genetic dissection of complex traits: a response to Weller *et al*. *Genetics* **154**, 1917–1918.

Zhang, D., Zhang, M. & Wells, M. T. (2004). Variable selection for large p small n regression model with incomplete data: application to QTL mapping. Technical Report. Department of Biostatistics and Computational Biology, University of Rochester Medical Center.