

Bayesian Inference for Genomic Imprinting Underlying Developmental Characteristics

Runqing Yang¹, Xin Wang¹ and Yuehua Cui^{2*}

¹*School of Agriculture and Biology, Shanghai Jiaotong University, Shanghai 200240, People's Republic of China*

²*Department of Statistics and Probability, Michigan State University, East Lansing, MI 48864, USA*

SUMMARY. The identification of imprinted genes is becoming a standard procedure in searching for quantitative trait loci (QTL) underlying complex traits. When a developmental characteristic such as growth or drug response is observed at multiple time points, understanding the dynamics of gene function governing the underlying feature should provide more biological information regarding the genetic control of an organism. Recognizing that differential imprinting can be development-specific, mapping imprinted genes considering the dynamic imprinting effect can provide additional biological insights into the epigenetic control of complex traits. In this study, we propose a Bayesian iQTL mapping framework considering the dynamics of imprinting effects and model multiple iQTLs with an efficient Bayesian model selection procedure. The method overcomes the limitation of the likelihood-based mapping procedure, and can simultaneously identify multiple iQTLs with different gene action modes across the whole genome with high computational efficiency. An inference procedure using Bayes factor to distinguish different imprinting patterns of the detected iQTL is proposed. The utility of the approach is illustrated through an analysis of a body weight growth dataset in an F₂ family derived from LG/J and SM/J mouse strains. The proposed Bayesian mapping method provides an efficient and computationally feasible framework for genome-wide multiple iQTL inference.

KEY WORDS: Developmental traits; Quantitative trait loci; Genomic imprinting; Imprinting pattern; Bayesian model selection

Running Head: Bayesian iQTLs mapping for developmental traits

*To whom correspondence should be addressed. E-mail: cui@stt.msu.edu

1. Introduction

Genomic imprinting is a genetic phenomenon in which the same genes are expressed differently, depending on their parental origin (Reik and Walter, 2001). On the molecular level, genomic imprinting may result from DNA methylation, histone modification, noncoding RNAs (ncRNA), and even long distance interchromosomal interactions (Wood and Oakey, 2006). Genomic imprinting has been broadly identified in plants (e.g., Alleman and Doctor, 2000), animals (e.g., Nezer et al., 1999; Van Laere et al., 2003) and humans (e.g., Fall et al., 1999; McInnis et al., 2003).

The role of genomic imprinting in shaping an organism's development has been unanimously recognized (Isles and Holland, 2005; Tycko and Morison, 2002; Constanica et al., 2004). The imprinting effect on traits of interest can be characterized as different types. When the paternal allele at a gene is expressed and the maternal allele is inactivated, this feature of imprinting is referred to as paternal imprinting. Similarly, maternal imprinting can be defined. Genomic imprinting has been traditionally viewed as a mono-allelic expression with complete maternal or paternal silence. The definition has been revised by the inclusion of partial imprinting which signifies the different levels of expression for alleles inherited from different parents (see Naumova and Croteau 2004; Sandovici et al. 2005). Noted that these classifications are all based on the additive effect of an imprinting locus, often imprinting can cause the change of interactions between alleles. Cheverud et al. (2008) recently illustrated a scheme for characterizing the potential diversity of imprinting patterns, in which, imprinting patterns are classified as either parental expression (paternal or maternal) or dominance (bipolar and polar). This is so far the most complete classification list for genomic imprinting.

Recent studies have shown the powerfulness of genetic mapping in the identification of epigenetic modification of imprinted genes or imprinted quantitative trait loci (iQTLs) on

complex traits. Methods are developed based on different mapping populations for the purpose to identify iQTLs with different genetic designs, for example, the variance components methods for family-based pedigree data in human linkage analysis (e.g., Hanson et al., 2001; Haghghi and Hodge, 2002; Shete and Amos, 2002); the variance components methods for experimental crosses (Li and Cui, 2009, 2010); the regression-based approaches for controlled crosses between outbred parents (e.g., Knott et al. 1998; de Koning et al. 2002) and between inbred lines (e.g., Cui et al. 2006; Cui 2007). In practice, when two reciprocal heterozygotes along with two homozygotes at each marker loci are fully informative or distinguishable in a mapping population, the imprinting effect of an iQTL can be uniquely determined by means of conditional probabilities of QTL genotypes on flanking markers. However, if two reciprocal heterozygotes are not fully informative or distinguishable, then the information about sex-specific differences in recombination fraction can be used to infer the imprinting effect of an iQTL (Cui et al. 2006). This allows us to infer in which fashion an iQTL is inherited in a segregation population without knowing specific allelic parental origin, such as in an F2 population (Cui et al. 2006).

Most imprinted genes play important roles in controlling embryonic and postnatal growth and development in mammal (Isles and Holland 2005; Tycko and Morison 2002; Constancia et al. 2004). As a highly complex process, genomic imprinting is involved in a number of growth axes operating coordinately at different development stages (Bartolomei and Tilghman 1997), and shows time-dependent effect during development (Villar et al. 1995). The unbalanced expression of an imprinted gene that occurs during a development stage challenges the traditional paradigm of inheritance and mapping methods. We argue that traditional methods, by treating a trait measured at a certain developmental stage as mapping subject, without considering the correlation information at different developmental stages, are less powerful in dissecting the dynamic iQTL effects. Cui et al. (2008a) recently proposed a

functional iQTL mapping framework underlying developmental characteristics, which incorporates a mathematical function that best describes a developmental feature into an iQTL mapping framework. Such an approach can estimate and test time-specific imprinting effect at specific developmental stages, and displays several merits over traditional iQTL mapping methods.

However, the current mapping procedures for iQTL inference are all single iQTL models, estimating and testing one locus at a time without considering the effects of other iQTLs. When multiple iQTLs are presented in the genome, such approaches are less efficient under the likelihood-based framework (Kao et al. 1999). For a dynamic trait, the number of parameters being estimated is several folds larger than those for a univariate trait. In our previous QTL mapping model, we demonstrated that a Bayesian mapping method can handle this issue well with high computational efficiency (Yang et al. 2006). In this study, we unified the two endeavors, Bayesian mapping of developmental traits and iQTL inference, into a unified framework called Bayesian functional multiple iQTL mapping (Bafmim). We proposed an efficient Bayesian model selection strategy for multiple iQTL inference for a developmental trait. The inference for the number, position and effect of multiple iQTLs as well as for different imprinting patterns were provided. The behavior of the proposed method was illustrated by a simulation study. A real data set was re-analyzed with the new method. We identified several new imprinted genes which otherwise can not be detected by current methods. The proposed method has great implications in understanding the function of imprinted genes governing developmental characteristics.

2. Statistical Method

2.1 The imprinting model

In a mapping population, assume that there are four distinguishable genotypes, denoted by

$Q_M Q_P$, $Q_M q_P$, $q_M Q_P$ and $q_M q_P$, at each locus where the subscript letter M and P refers to an allele inherited from the maternal and paternal parents, respectively. A set of codominant molecular markers can be genotyped and phenotypes for a developmental trait which is measured at m time points on n individuals. In general, the additive effect a , is defined as half of the phenotypic difference between two homozygotes; the dominance effect d , is defined as the difference between the joint mean of both heterozygotes and the mean of both homozygotes; and the imprinting effect i , is defined as the difference between both heterozygotes (Falconer and Mackay 1996; Knott et al. 1998). Following the definitions about the genetic parameters, an imprinting model for a phenotype measured for individual k at time t , denoted as $y_k(t)$, can be formulated as

$$y_k(t) = \mu(t) + \sum_{j=1}^q [z_{kj}a_j(t) + w_{kj}d_j(t) + s_{kj}i_j(t)] + \xi_k(t) + e_k(t) \quad (1)$$

where q is the number of potential iQTLs in the genome; $\mu(t)$ is the population mean at time t ; $a_j(t)$, $d_j(t)$ and $i_j(t)$ for $j=1, 2, \dots, q$ are the additive, dominance and imprinting effects of the j th QTL at time point t ; $\xi_k(t)$ ($k=1, 2, \dots, n$) is an individual-specific time-dependent random environmental effect following normal distribution, i.e., $N(0, \sigma_\xi^2(t))$; and $e_k(t)$ is a random environmental error assumed to be normally distributed with mean zero and variance σ^2 . Note that z_{kj} , w_{kj} and s_{kj} are genotype-specific indicator variables related to genetic effects a_j , d_j and i_j , which are defined by Mantey et al. (2005) as

$$z_{kj} = \begin{cases} +1 \\ 0 \\ 0 \\ -1 \end{cases}, \quad w_{kj} = \begin{cases} 0 \\ +1 \\ +1 \\ 0 \end{cases}, \quad s_{kj} = \begin{cases} 0 & \text{for } Q_M Q_P \\ +1 & \text{for } Q_M q_P \\ -1 & \text{for } q_M Q_P \\ 0 & \text{for } q_M q_P \end{cases}$$

We use Legendre polynomial of order r to fit changing trajectories of the population mean and the effects of each iQTL (Yang and Xu 2007; Cui et al. 2008b). Let $\psi(t)$ be the basis of

the Legendre polynomial and have that $\mu(t) = \psi(t)\mu$, $a_j(t) = \psi(t)a_j$, $d_j(t) = \psi(t)d_j$, $i_j(t) = \psi(t)i_j$ and $\xi_k(t) = \psi(t)\xi_k$, where each one of μ , a_j , d_j , i_j and ξ_k is a vector of $r+1$ dimensions. Model (1) can be then rewritten as

$$y_k(t) = \psi(t)\mu + \sum_{j=1}^q [z_{kj}\psi(t)a_j + w_{kj}\psi(t)d_j + s_{kj}\psi(t)i_j] + \psi(t)\xi_k + e_k(t) \quad (2)$$

where ξ_k is a vector of random effects, which is assumed to be multivariate normal with mean zero and a $(r+1)(r+1)$ positive definite covariance matrix Σ .

For simplicity, we assume that each individual is measured at m time points and the time points are common for all individuals. Let $y_k = [y_k(t_0), y_k(t_1), \dots, y_k(t_m)]^T$ be a $(m+1) \times 1$ column vector for the repeated measurements of a developmental trait, and define $\psi = [\psi^T(t_0) \quad \psi^T(t_1) \quad \dots \quad \psi^T(t_m)]$ as a $(r+1)(m+1)$ matrix. In matrix notation, Model (2) becomes

$$y_k = \psi^T \mu + \sum_{j=1}^q [z_{kj}\psi^T a_j + w_{kj}\psi^T d_j + s_{kj}\psi^T i_j] + \psi^T \xi_k + e_k \quad (3)$$

where $e_k = [e_k(t_0), \dots, e_k(t_m)]^T$ is an $(m+1) \times 1$ vector for the environmental errors, distributed as $e_k \sim N(0, I\sigma^2)$ with I being an $(m+1) \times (m+1)$ identity matrix.

2.2 Bayesian model selection for genetic parameters

Model (3) is a mixed-effect model where population mean and genetic effects are fixed and the time-dependent environmental effect is random. Also noted that Model (3) is not a regular linear mixed-effect model, since the number of independent variables for the fixed effects and the associated indicator variables are unknown due to unknown number of iQTLs. In principle, iQTLs can be distributed anywhere in the genome, and hence any genomic positions can be potential iQTL locations. Thus, we approximate positions for all possible iQTLs by partitioning the entire genome into evenly spaced loci, covering all observed

markers and additional loci between flanking markers. The expected values for elements in a relative design matrix to each locus based on the conditional probabilities of the locus genotypes on two flanking markers can be calculated (Cui et al. 2006). For a supersaturated model where each genomic location could reside a potential iQTL, a huge number of genetic effects is almost impossible to be estimated. So we preset an upper bound on the number of iQTLs in the model (see Yi et al. 2005). The upper bound should be larger than the potential number of detectable iQTLs in a given data set.

Given an upper bound on the number of iQTLs, these iQTLs can be drawn from densely spaced loci over the genome. Even with a moderate number of upper bound, there are many genetic effects being estimated in Model (3). For inferring the existence of these effects, we introduce a random binary variable γ to indicate which genetic effects should be included in or excluded from the model, corresponding to $\gamma=1$ or $\gamma=0$ (George and McCulloch 1997; Kuo and Mallick 1998; Chipman et al. 2001). Model (3) then becomes

$$y_k = \psi^T \mu + \sum_{j=1}^q \left[\gamma_{aj} z_{kj} \psi^T a_j + \gamma_{dj} w_{kj} \psi^T d_j + \gamma_{ij} s_{kj} \psi^T i_j \right] + \psi^T \xi_k + e_k \quad (4)$$

where γ_{lj} ($l=a, d, \text{ or } i$) is the indicator variable for genetic effects $a, d, \text{ or } i$. Within the framework of Bayesian model selection, Bayesian sampling for unknown parameters in Model (4), including $\mu, \gamma, a, d, i, \text{ and } \Sigma$, is implemented with the MCMC algorithm.

Considering the same forms of posterior distributions for each γ as well as a, d and i , we

simplify $\sum_{j=1}^q [\bullet]$ item in Model (4) as $\sum_{j=1}^q \gamma_j x_{kj} \psi \beta_j$. Also noted that the released sampling

value for binary variable γ at a previous round determines which genetic effects and position of an iQTL should be drawn or estimated at the next round. We leave the details on specification of prior distributions for each parameter (see Yang et al. 2006; Yi et al. 2005, 2007), and focus our presentation on detailed steps in the proposed Bayesian model selection:

- (1) Calculate the expected values for the associated design matrix with all loci in the genome:

$$E(z) = \pi_{QQ} - \pi_{qq}, E(w) = \pi_{Qq} + \pi_{qQ}, \text{ and } E(s) = \pi_{Qq} - \pi_{qQ}$$

with $\pi_{QQ}, \pi_{Qq}, \pi_{qQ}$ and π_{qq} being the conditional probabilities of the tested iQTL genotype $Q_M Q_P, Q_M q_P, q_M Q_P$ and $q_M q_P$ on two flanking markers.

- (2) Set an upper bound on the number of imprinting loci, estimated by

$$L = l_0 + 3\sqrt{l_0}$$

with l_0 being prior expected number of imprinted loci that is determined according to initial investigations with traditional methods (e.g., Cui et al. 2008a). Thus, the prior

inclusion probability for an iQTL effect is $1 - \left[1 - \frac{l_0}{L}\right]^{\frac{1}{3}}$.

- (3) Initialize all variables with some initial values or values sampled from their prior distributions (see Yang et al. 2006; Yi et al. 2005, 2007).

- (4) Update population mean μ by sampling from a normal distribution with mean

$(n\psi V^{-1}\psi^T)^{-1}\psi V^{-1}\sum_{k=1}^n (y_k - M_k + \psi^T \mu)$ and covariance matrix $(n\psi V^{-1}\psi^T)^{-1}$, where

$M_k = \psi^T \mu + \sum_{j=1}^q \gamma_j x_{kj} \psi \beta_j$ ($k=1, 2, \dots, n$) is the conditional expectation and

$V = \psi^T \Sigma \psi + I\sigma^2$ is the covariance matrix in Model (4) for given fixed effects.

- (5) Update the genetic effects a , d and i corresponding to $\gamma_j=1$ by drawing from a normal distribution with mean

$\hat{\beta}_j = \left[\left(1 + \frac{1}{c}\right) \sum_{k=1}^n x_{kj}^2 \psi V^{-1} \psi^T \right]^{-1} \psi V^{-1} \sum_{k=1}^n x_{kj} (y_k - M_k + \gamma_j x_{kj} \psi \beta_j)$ and variance

$\hat{\Sigma}_j = \left[\left(1 + \frac{1}{c}\right) \sum_{k=1}^n x_{kj}^2 \psi V^{-1} \psi^T \right]^{-1}$. In general, c takes value of n . Note that if $\gamma_j=0$, then the

corresponding β_j is taken as a zero vector.

- (6) Update the binary indicators γ_l by adopting an efficient Metropolis - Hastings algorithm (Kohn et al. 2001; Yi et al. 2007) with the probability of acceptance as $\min(1, \rho)$, where

$$\rho = \left(\frac{wR}{1-w} \right)^{1-2\gamma_l} \quad \text{with} \quad R = \sqrt{\frac{c}{c+1}} \exp\left(-\frac{1}{2} \hat{\beta}_j^T \hat{\Sigma}_j^{-1} \hat{\beta}_j\right).$$

Note that γ_l represents any one of the binary variables defined in Model (4).

- (7) Update individual-specific time-dependent random environmental effect ξ_k by sampling from normal distribution with mean $\Sigma \psi V^{-1} (y_k - M_k)$ and covariance matrix $\Sigma - \Sigma \psi V^{-1} \psi^T \Sigma$.

- (8) Update covariance matrix Σ of ξ_k by drawing from an inverse Wishart distribution with the form $IW(\nu_h + n, \sum_{k=1}^n \xi_k \xi_k^T + S_h)$, where ν_h and S_h are prior hyperparameters.

- (9) Update the residual variance σ^2 by sampling from an inverse Chi-square distribution with parameters $\nu_e + n$ and $(\nu_e + n) S_e + \left(\sum_{k=1}^n e_k^T e_k \right)^{-1}$, where ν_e and S_e are prior hyperparameters, and $e_k = y_k - M_k - \psi^T \xi_k$.

- (10) Update the iQTL position by drawing from all spaced loci over the genome. Herein, the imprinting locus is sampled as long as the binary variable equals 1 for at least one of the genetic effects a , d and i at that locus. Each locus is drawn from a variable interval whose boundaries are the positions of adjoining QTLs. Metropolis - Hastings algorithm is used to decide whether each proposed (new) position should be accepted or not (see Wang et al. 2005; Zhang & Xu 2005).

- (11) Repeat steps (4) – (10) until the Markov chain reaches a desirable length.

Post MCMC analysis includes the monitor of the mixing behavior and convergence rates of the MCMC algorithm, and the assessment of characteristics of the imprinting genetic

architecture. The former can be checked by visually inspecting trace plots of the sample values of scalar quantities of interest or formal diagnostic methods provided in the package R/coda (Plummer et al. 2004). The latter can use model averaging which accounts for model uncertainty and average over possible models weighted by their posterior probabilities (see Raftery et al. 1997; Ball 2001; Sillanpää and Corander 2002). The posterior inclusion probability for each locus is estimated as its frequency in the posterior samples. Bayes factor (BF) is used as a measurement for inclusion against exclusion at each iQTL locus (Kass and Raftery 1995). Generally, a threshold of BF is empirically determined as 3, or $2 \ln BF = 2.1$, for declaring statistical significance for each iQTL Locus.

2.3 Bayesian inference for imprinting mode of action

Generally speaking, an iQTL detected with the above Bayesian algorithm can not be declared as an imprinted QTL, until we do further imprinting inference. After an iQTL is detected, we can adopt the idea of Bayes factor to infer statistical significance for its imprinting effect with the form

$$BF = \frac{p \cdot \frac{1-p}{p}}{1-p}$$

where p is a prior probability and p is a posterior probability for a certain genetic effect, which is calculated as the proportion of samples in which $\gamma_i=1$ in MCMC sampling rounds. If the BF is greater than 3 (or $2 \ln BF > 2.1$) for the imprinting effect i , then the detected iQTL can be claimed as a true iQTL, otherwise as a Mendelian QTL.

According to the relative estimated effects for a , d and i , Cheverud et al (2008) classified imprinting patterns as parental imprinting, i.e., $a = \pm i$ and $d = 0$, including paternal ($a = i$) and maternal ($a = -i$) imprinting subtypes; and dominance imprinting with $a = 0$ but $i \neq 0$, which can be further distinguished as bipolar imprinting in which $d = 0$ and $i \neq 0$ and polar imprinting in which $d = \pm i$. This classification provides a comprehensive dissection of

imprinting pattern an iQTL may possess.

With the above defined imprinting pattern, we substitute the related hypotheses into the genomic imprinting model (1) to differentiate imprinting patterns and obtain different reduced models. Table 1 shows all possible imprinting patterns. The corresponding hypotheses as well as the reduced models are also listed. Based on the reduced models, we also adopted Bayesian model selection to estimate genetic effects of each locus. In order to infer imprinting patterns of the detected iQTL, that is, to statistically assess the hypotheses that a or $d = \pm i$, a new Bayes factor is formulated by comparing posterior probability for detected locus between the full and the reduced model, denoted by

$$BF = \frac{P_{full}}{P_{reduced}}$$

where the prior probabilities for the full and reduced models at the detected iQTL are the same and therefore are integrated out.

In fact, the Gibbs sampling for associated genetic effects with different imprinting patterns can be carried out together with Bayesian sampling of the full model, or can be done separately after the sampling of the full model. Apparently, the later one is more efficient and is computationally faster.

3. Simulation Studies

We conducted simulation studies to evaluate the performance of the proposed Bayesian functional multiple iQTL mapping approach. A genome consisting of a single large chromosome of 600cM was simulated covering 61 evenly spaced markers. The growth pattern of a dynamic trait was assumed to be controlled by one QTL inherited in a Mendelian fashion and 4 iQTLs with their imprinting patterns, positions and effects listed in Table 2. The order of the Legendre polynomial that generates the growth trajectory was assumed to be $r=3$. We simulated a dynamic trait measured at eight time points assuming different sample

sizes ($n=250, 500$) with inbred F_2 individuals. The marker and QTL genotypes in the F_2 family were generated by mimicking sex-specific recombination fractions (see Cui et al. 2006 for more details). The population mean and the individual-specific environmental error covariance matrix were set the same as described in Yang and Xu (2006), and the residual variance was set as 4.

In all analyses for simulated data, we set the prior number of main-effect iQTLs as 4. The upper bound of the number of iQTLs was then equal to $L = 4 + 3\sqrt{4} = 10$. The actual values for the hyperparameters used here mimic the results obtained in real data analyses (see the real data analysis section). The initial values of all variables were sampled from their prior distributions. The MCMC is run for 10000 cycles as a burn-in period (deleted) and then for an additional 150,000 cycles after the burn-in. The chain is then thinned to reduce serial correlation by saving one observation in every 50 cycles. The posterior sample contained 3000 observations for the post-MCMC analysis. Note that the length of the burn-in is judged by visually inspecting the plots of some posterior samples across rounds and is set to enough cycles for ensuring the MCMC convergence. The simulation experiment is replicated 100 times for evaluating the statistical power of our method.

Table 3 shows the mean estimates as well as their standard deviation (in parenthesis) for the parameters given in Table 2. The relative statistical power to detect each QTL is also listed. Overall, the Bayesian mapping approach is able to estimate the regression effects of the iQTLs with reasonable precision. All the four QTL positions can be accurately estimated with high precision. As we expected, increasing sample size always leads to small bias, increased parameter estimation precision, and high mapping power. For example, the mapping power for iQTL 1 increases from 70% to 85% when sample size is increased from 250 to 500. Even with small sample size ($n=250$), the iQTL position can also be estimated with high precision. This indicates the power of Bayesian mapping for multiple iQTL

inference. It is also worthy to emphasize that we can accurately infer the imprinting pattern of the detected locus using Bayes factor (data not shown). The simulation indicates the robustness of the proposed method in multiple iQTL detection for dynamic traits with moderate sample size.

4. Real Data Analysis

We illustrate the application of our proposed approach by reanalyzing a mouse body weight growth dataset with an F_2 mating population derived from two inbred strains, the Large (LG/J) and the Small (SM/J). Total 502 F_2 mice were genotyped for 96 microsatellite markers located on 19 autosomal chromosomes. A linkage map of a total length of 1780cM has been constructed (for details, see Vaughn et al. 1999). The body mass is measured on each mouse at 10 weekly intervals starting at day 7. The raw weights were adjusted for the effects of each covariate due to dam, litter size at birth and parity, and sex (Vaughn et al. 1999). The dataset has been analyzed by Cui et al. (2008a) with likelihood-based functional mapping.

By fitting the mean change of weight growth over age, we choose the Legendre polynomial of order 4 as the base model to describe the changing trajectory for each component except for residuals, described in Model (1). The female-to-male recombination rate of 1.25:1 is used to estimate conditional probabilities for the four iQTL genotypes (Cui et al. 2006). The expected number of main-effect iQTLs was set as $l_0 = 4$ according to the results by Cui et al. (2008a) and the upper bounds of the number of iQTLs are then calculated as $L = 4 + 3\sqrt{4} = 10$. Thus, the prior inclusion probability for iQTL effects is calculated as

$1 - \left[1 - \frac{4}{10}\right]^{\frac{1}{3}} = 0.156$. The actual values for the hyperparameters are set as $S_h = S_e = 0.5I$,

$v_h = r + 1$ and $v_e = 0$. The initial values of all variables were sampled from their prior distributions. The MCMC is run 200,000 cycles after the burn-in period of 10,000 cycles.

Figure 1 plots the profiles of $2\log\text{BF}$ obtained with the Bayesian model selection. The top figure shows that there are eight peaks for $2\log\text{BF}$ s exceeding the horizontal reference line with an empirical critical value 2.1, indicating that eight QTLs are detected on chromosomes 2, 4, 6, 7, 9, 10, 11 and 15. Among these QTLs detected, six (on chromosomes 2, 4, 6, 7, 10 and 15) shows significant imprinting effect, as their relative $2\log\text{BF}$ s are greater than 2.1 for imprinting effects (the bottom figure in Figure 1). The ones on chromosomes 9 and 11 show Mendelian inheritance. Table 4 tabulates the position on each chromosome and the estimated effects (additive, dominance and imprinting) for the 6 detected iQTLs. The estimated regression coefficients for each iQTL have no biological meaning, but they can be used to predict the effects of an iQTL at any time points by substituting a time point into the Legendre polynomial with the regression coefficients.

We further evaluated the imprinting pattern of the 6 detected iQTLs. Two of them (on chromosomes 4 and 10) show polar imprinting, two (on chromosomes 2 and 6) show maternal imprinting and two (on chromosomes 7 and 15) show paternal imprinting, based on the results obtained from the significance analysis given in Table 5. Compared to the likelihood-based method (Cui et al. 2008a), the Bayesian method identified four more QTLs (on chromosomes 2, 4, 9 and 11).

5. Discussion

The epigenetic phenomenon in genomic imprinting has been constantly challenging and revising the traditional paradigm of inheritance. The inheritable property of imprinting provides clues for complicated genetic disorders (Falls et al. 1999). In the meantime, it also brings challenges for statistical modeling and mapping. We developed a Bayesian model selection method to identifying multiple iQTLs for developmental traits illustrated in an F_2 mating population. It can also be extended to mating designs such as a reciprocal backcross design (Cui et al. 2007). The Bayesian method has shown relative merits in multiple QTL

mapping partly due to its flexibility to handle a large parameter space (Yang et al. 2006; Yang and Xu 2007; Yi et al. 2005, 2007). Both simulation and real data analysis indicate the relative power of the proposed Bayesian multiple iQTL mapping for functional traits.

The current method is developed specifically for longitudinal or functional traits. Recently, Hayashi and Awata (2008) proposed a Bayesian mapping approach which can simultaneously map multiple QTLs, and further discriminating Mendelian and imprinting expressions of a QTL. Although the approach shows improvement in iQTL detection, it is limited by a number of facts. For example, drawing number of QTLs with a reversible-jump MCMC procedure may have low convergence efficiency. Moreover, the method is developed for univariate traits and ignores the dynamics of gene effects. In an earlier paper, Yang et al. (2010) developed a Bayesian multiple iQTL mapping method for univariate traits. The current work is an extension of our previous work, but taking the challenges of modeling the dynamics of genetic effects.

Our method assumes a maximum number of detectable iQTLs and introduces latent binary variables to indicate which main effects for a putative iQTL should be included or excluded from the model. Compared to the likelihood-based method (e.g., Cui et al., 2008a), it allows MCMC sampling for iQTL parameters to carry out in the reduced model space, thus enhancing the computational efficiency of Bayesian multiple iQTL mapping with many parameters. In addition, it facilitates statistical inference for imprinting patterns of the detected iQTLs with appropriately defined Bayes factors. We took the dynamic course of a developmental trait as the mapping subject, and therefore identified more iQTLs altering the developmental trajectory than separately performing iQTL mapping at each time point (Cui et al. 2006). In terms of computational speeds, the likelihood-based method is slower than the Bayesian method due to the slow convergence rate of parameter estimation.

The key issue for iQTL mapping of developmental traits is to choose appropriate

submodels for imprinting inference. In the likelihood-based mapping framework, Cui et al. (2008b) proposed to describe the changes of iQTL genotypic effects by a logistic growth curve. This parametric assumption is, however, difficult to implement in a multiple iQTL model for Bayesian mapping due to the nonlinearity of different submodels. We relaxed the parametric assumption, and adopted an orthogonal polynomial mapping approach for dynamic iQTL inference. Appropriately selecting optimal polynomial order is essential in determining the shape of phenotypic trajectories of a dynamic trait. Cui et al. (2008b) proposed two methods to choose an optimal order: (1) assuming the same order for different phenotypic trajectories which can be chosen under the null hypothesis; or (2) assuming different orders for different phenotypic trajectories. The second method, however, incurs huge computational burden. Thus, we adopted the first method and assume the same order to fit changes in QTL genetic effects and time-dependent environmental effects. The polynomial regression coefficients determine the shape of the population mean and each genetic effect. Certainly, an optimal mapping strategy is to choose the polynomial of different orders to best model each component in the imprinting model, which we leave it for future investigation.

In a functional mapping, how to model the correlation structure for repeated measurements is also a challenging problem. In most functional mapping studies, a parametric residual covariance structure such as the autoregressive model with order 1 [AR(1)] is often assumed (Ma et al. 2002). However, it is very difficult to implement a Bayesian method for this structure because of the complicated structure for the marginal posterior distribution. In contrast, the covariance structure described by $\psi^T \Sigma \psi + I\sigma^2$ is more flexible than the parametric structure because we can actually choose different degrees of polynomial order to fit a covariance structure with a large degree of complexity (Yang and Xu 2007). Moreover, we can easily sample the covariance matrix Σ from a closed form of marginal posterior distribution. Yap et al. (2009) recently proposed a nonparametric method

for covariance structure modeling in functional mapping. More work is needed to integrate these techniques into a Bayesian mapping framework to improve mapping power.

The multiple iQTL model for developmental traits proposed herein could be treated as a general form of the model for analyzing genomic imprinting of a quantitative trait. For instance, let $\psi = 1$ and $\xi_i = 0$ in scale, that is, only one fixed time point is measured on each individual, leading to a multiple interacting iQTL model for a univariate quantitative trait; take ψ to be an identity matrix of order m , and ξ_i to be a zero vector, resulting in a multiple interacting iQTL model for a multiple quantitative trait. If ξ_i is assigned to be non-zero in above two cases, then a multiple iQTL model for a univariate and a multivariate quantitative trait can also be able to make use of repeated records on the phenotype. Corresponding Bayesian model selection approaches can be likewise obtained by taking different values for ψ and ξ_i .

ACKNOWLEDGMENTS

We wish to thank James Cheverud for providing the mouse data set. The research was supported in part by the Chinese National Natural Science Foundation Grant 30471236 to RY, and by the US National Science Foundation grant DMS-0707031 to YC.

REFERENCES

- Alleman, M. and Doctor, J. (2000). Genomic imprinting in plants: observations and evolutionary implications. *Plant Molecular Biology* **43**, 147–161.
- Ball, R. D. (2001). Bayesian methods for quantitative trait loci mapping based on model selection: approximate analysis using the Bayesian information criterion. *Genetics* **159**, 1351–1364.
- Bartolomei, M.S. and Tilghman, S.M. (1997). Genomic imprinting in mammals. *Annual Review of Genetics* **31**, 493–525.
- Cheverud, J.M., Hager, R., Roseman, C., Fawcett, G., Wang, B., and Wolf, J.B. (2008).

- Genomic imprinting effects on adult body composition in mice. *Proceedings of the National Academy of Sciences* **11**: 4253–4258.
- Chipman, H., Edwards, E.I., and McCulloch, R.E. (2001) The practical implementation of Bayesian model selection, pp. 65–116 in *Model Selection* (edited by P. Lahiri). Institute of Mathematical Statistics, Beachwood, OH.
- Constancia, M, Kelsey, G., and Reik, W. (2004). Resourceful imprinting. *Nature* **432**, 53–57.
- Cui, Y.H. (2007). A statistical framework for genome-wide scanning and testing imprinted quantitative trait loci. *Journal of Theoretical Biology* **244**, 115–126.
- Cui, Y.H., Cheverud, J.M. and Wu, R.L. (2007). A statistical model for dissecting genomic imprinting through genetic mapping. *Genetica* **130**, 227-239.
- Cui, Y.H., Lu, Q., Cheverud, J.M., Littell, R.C., and Wu, R.L. (2006). Model for mapping imprinted quantitative trait loci in an inbred F2 design. *Genomics* **87**, 543–551.
- Cui, YH, Li, S., and Li, G. (2008a) Functional mapping imprinted quantitative trait loci underlying developmental characteristics. *Theoretical Biology and Medical Modeling* **5**, 1–15.
- Cui, Y.H., Wu, R., Casella, G., and Zhu, J. (2008b). Nonparametric functional mapping quantitative trait loci underlying programmed cell death. *Statistical Applications in Genetics and Molecular Biology* Vol. 7, Iss. 1, Article 4.
- de Koning, D.-J., Bovenhuis, H., and van Arendonk, J.A.M. (2002). On the detection of imprinted quantitative trait loci in experimental crosses of outbred species. *Genetics* **161**, 931–938.
- Falconer, D.S., and Mackay, T.F.C. (1996). *Introduction to quantitative genetics* 4th ed. London: Longman.
- Falls, J.G., Pulford, D.J., Wylie, A.A., and Jirtle, R.L. (1999). Genomic imprinting: implications for human disease. *American Journal of Pathology* **154**, 635–647.
- George, E.I., and McCulloch, R.E. (1997). Approaches for Bayesian variable selection. *Statistical Science* **7**, 339–373.
- Hanson, R.L., Kobes, S., Lindsay, R.S., and Knowler, W.C. (2001). Assessment of parent-of-origin effects in linkage analysis of quantitative traits. *American Journal of Human Genetics* **68**, 951–962.
- Haghighi, F., and Hodge, S.E. (2002). Likelihood formulation of parent-of-origin effects on segregation analysis, including ascertainment. *American Journal of Human Genetics* **70**, 142–156.
- Isles, A.R., and Holland, A.J. (2005). Imprinted genes and mother-offspring interactions.

- Early Human Development* **81**, 73-77.
- Kao, C.H., Zeng, Z.B., and Teasdale, R. D. (1999). Multiple interval mapping for quantitative trait loci. *Genetics* **152**, 1203–1216.
- Kass, R.E. and Raftery, A.E. (1995). Bayes factors. *Journal of American Statistical Association* **90**, 773–795.
- Knott, S.A., Marklund, L., Haley, C.S., Andersson, K., Davies, W., et al. (1998). Multiple marker mapping of quantitative trait loci in a cross between outbred wild boar and large white pigs, *Genetics* **149**, 1069-1080.
- Kohn, R., Smithand, M., and Chen, D. (2001). Nonparametric regression using linear combinations of basis functions. *Statistical Computing* **11**, 313–322.
- Kuo, L., and Mallick, B. (1998). Variable selection for regression models. *Sankhya Ser. B* **60**, 65–81.
- Li, G.X. and Cui, Y.H. (2009). A statistical variance components framework for mapping imprinted quantitative trait loci in experimental crosses. *Journal of Probability and Statistics* vol. 2009, Article ID 689489, doi:10.1155/2009/689489.
- Li, G.X. and Cui, Y.H. (2010). A general statistical framework for dissecting parent-of-origin effects underlying triploid endosperm traits in flowering plants. *Annals of Applied Statistics* (in press).
- Ma, C.X., Casella, G., and Wu, R. (2002) Functional mapping of quantitative trait loci underlying the character process: a theoretical framework. *Genetics* **161**, 1751-1762.
- Mantey, C., Brockmann, G. A., Kalm, E., and Reinsch, N. (2005). Mapping and Exclusion Mapping of Genomic Imprinting Effects in Mouse F2 Families. *Journal of Heredity* **4**, 329–338.
- McInnis, M.G., Lan, T.H., Willour, V.L., McMahon, F.J., Simpson, S.G. et al. (2003). Genome-wide scan of bipolar disorder in 65 pedigrees: supportive evidence for linkage at 8q24, 18q22, 4q32, 2p12, and 13q12. *Mol Psychiatry* **8**, 288–298.
- Naumova, A.K., and Croteau, S. (2004). Mechanisms of epigenetic variation: polymorphic imprinting. *Current Genomics* **5**, 417–429.
- Nezer, C., Moreau, L., Brouwers, B., Coppieters, W., Detilleux, J. et al. (1999). An imprinted QTL with major effect on muscle mass and fat deposition maps to the IGF2 locus in pigs. *Nature Genetics* **21**, 155–156.
- Raftery, A. E. Madigan, D., and Hoeting, J.A. (1997). Bayesian model averaging for linear regression models. *Journal of American Statistical Association* **92**, 179–191.
- Reik, W., and Walter, J. (2001). Genomic imprinting: parental influence on the genome.

- Nature Review Genetics* **2**, 21-32.
- Plummer, M. N. *et al.* (2004). Output Analysis and Diagnostics for MCMC, v. 0.9–5. (<http://www-fis.iarc.fr/coda/>).
- Sandovici, I., Kassovska-Bratinova, S., Loredano-Osti, J.C., Leppert, M., Suarez, A. *et al.* (2005). Interindividual variability and parent of origin DNA methylation differences at specific human Alu elements. *Human Molecular Genetics* **14**, 2135–2143.
- Sillanpää, M.J. and Corander, J. (2002). Model choice in gene mapping: what and why. *Trends Genetics* **18**, 301–307.
- Shete, S., and Amos, C.I. (2002). Testing for genetic linkage in families by a variance components approach in the presence of genomic imprinting, *American Journal of Human Genetics* **70**, 751–757.
- Tycko, B., and Morison, I.M. (2002). Physiological functions of imprinted genes. *Journal of Cell Physiology* **192**, 245-258.
- Van Laere, A.S., Nguyen, M., Braunschweig, M., Nezer, C., Collette, C. *et al.* (2003). A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig. *Nature* **425**, 832–836.
- Vaughn, T.T., Pletscher, L.S., Peripato, A., King-Ellison, K., Adams, E., Erikson, C., and Cheverud, J.M. (1999). Mapping quantitative trait loci for murine growth—A closer look at genetic architecture. *Genetics Research* **74**, 313-322.
- Villar, A.J., Eddy, E.M., and Pedersen, R.A. (1995). Developmental regulation of genomic imprinting during gametogenesis. *Developmental Biology* **172**, 264.
- Wang, H., Zhang, Y.M., Li, X., Masinde, G.L., Mohan, S., Baylink, D.J., and Xu, S. (2005). Bayesian Shrinkage Estimation of Quantitative Trait Loci Parameters. *Genetics* **170**, 465–480.
- Wood, A.J. and Oakey, R.J. (2006). Genomic imprinting in mammals: emerging themes and established theories. *PLoS Genetics* **2**, 1677–1685.
- Yang, R., Tian, Q. and Xu, S. (2006). Mapping QTL for longitudinal traits in line crosses. *Genetics* **173**, 2339-2356.
- Yang, R. and Xu, S. (2007). Bayesian shrinkage analysis of quantitative trait loci for dynamic traits. *Genetics* **176**, 1169-1185.
- Yang, R., Wang, X., Wu, Z., Prows, D.R., and Lin, M. (2010). Bayesian model selection for characterizing genomic imprinting effects and patterns. *Bioinformatics* **26**, 235-241.

- Yap, J.S., Fan, J. and Wu, R. (2009). Nonparametric modeling of longitudinal covariance structure in functional mapping of quantitative trait loci. *Biometrics* **65**, 1068- 1077.
- Yi, N., Yandell, B.S., Churchill, G.A., Allison, D.B., Eisen, E.J., and Pomp, D. (2005) Bayesian model selection for genome-wide epistatic quantitative trait loci analysis. *Genetics* **170**, 1333–1344.
- Yi, N., Banerjee, S., Pomp, D., and Yandell, B.S. (2007). Bayesian mapping of genome-wide interacting quantitative trait loci for ordinal traits. *Genetics* **176**, 1855–1864.
- Zhang, Y.M. and Xu, S. (2005). Advanced statistical methods for detecting multiple quantitative trait loci. *Recent Research Development in Genetics and Breeding* **2**, 1-23.

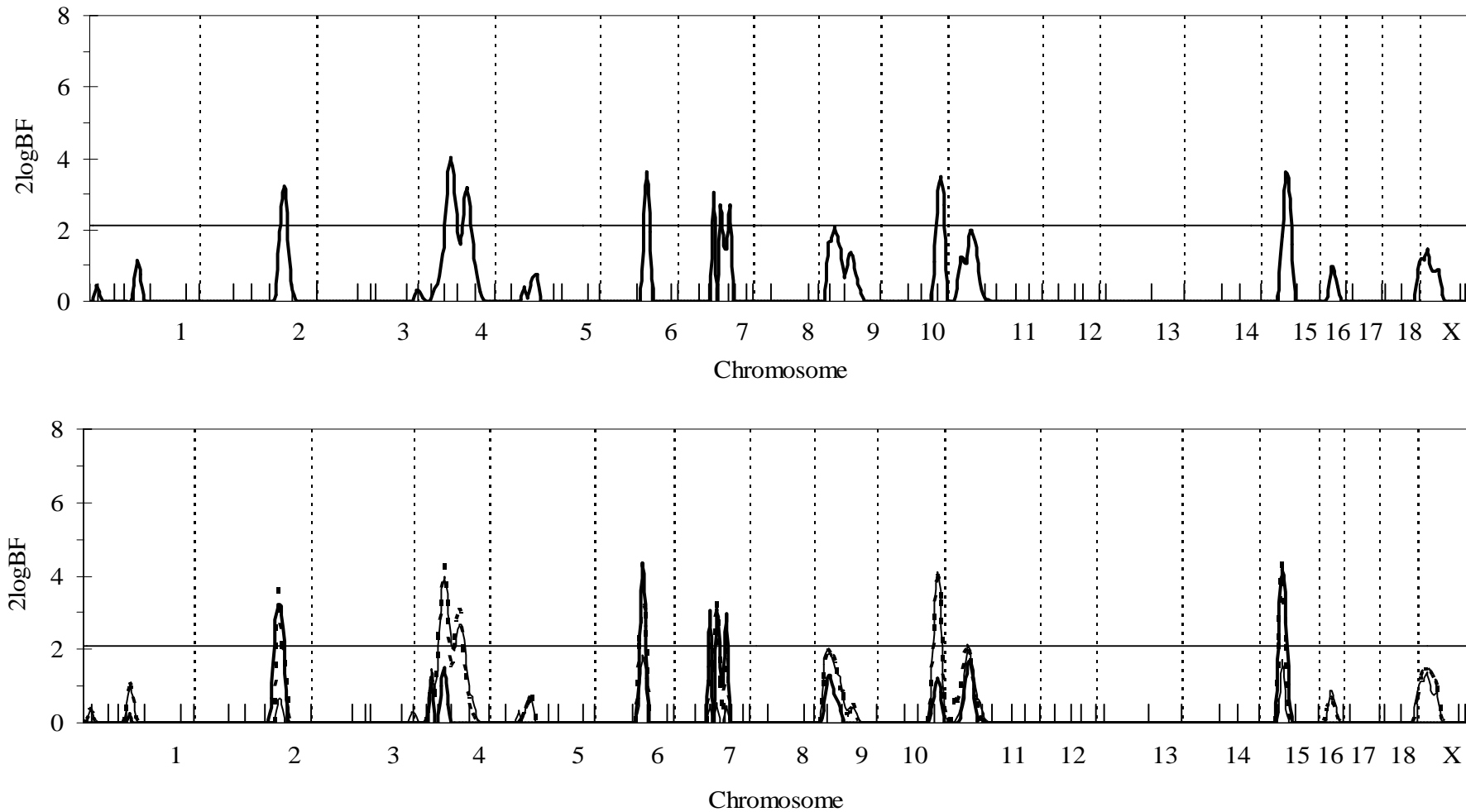


Figure 1— The profiles of $2\log\text{BF}$ for genome-wide iQTL scan (top) and for various genetic effects inference of iQTLs (bottom) obtained with the Bafmim in mouse weight growth. In both figures, linkage groups are separated by the vertical dotted lines and marker positions are indicated by the ticks on the horizontal axis. The horizontal reference line is the empirical critical value 2.1 for $2\log\text{BF}$. In the bottom figure, the thick solid, thin solid and dashed lines represent additive, dominance and imprinting effects, respectively.

Table 1: Imprinting patterns, hypothesis and corresponding reduced models

Imprinting Pattern	Imprinting direction	Hypothesis	Reduced model	
Parental Expression	Paternal	$a = i \neq 0$ and $d = 0$	$y_k = \psi^T \mu + \sum_{j=1}^q [(z_{kj} + s_{kj}) \psi^T a_j] + \psi^T \xi_k + e_k$	
	Maternal	$a = -i \neq 0$ and $d = 0$	$y_k = \psi^T \mu + \sum_{j=1}^q [(z_{kj} - s_{kj}) \psi^T a_j] + \psi^T \xi_k + e_k$	
	Bipolar	$a = d = 0$ and $i \neq 0$	$y_k = \psi^T \mu + \sum_{j=1}^m s_{kj} \psi^T i_j + \psi^T \xi_k + e_k$	
Dominance Imprinting	Polar	Over-dominance	$a = 0$ and $d = i$	$y_k = \psi^T \mu + \sum_{j=1}^q [(w_{kj} + s_{kj}) \psi^T d_j] + \psi^T \xi_k + e_k$
		Under-dominance	$a = 0$ and $d = -i$	$y_k = \psi^T \mu + \sum_{j=1}^q [(w_{kj} - s_{kj}) \psi^T d_j] + \psi^T \xi_k + e_k$

Table 2: The imprinting type and parameters (Regression effects) of iQTLs used in simulation

QTL No.	Imprinting type	Position	a_0	a_1	a_2	a_3	d_0	d_1	d_2	d_3	i_0	i_1	i_2	i_3
1	Mendelian	23	1.82	-0.80	-1.20	-0.80								
2	Paternal	148	0.00	1.65	2.52	1.20					0.00	1.65	2.52	1.20
3	Maternal	256	2.55	1.36	-2.02	-1.27					-2.55	-1.36	2.02	1.27
4	Bipolar	332									2.55	1.36	-2.02	-1.27
5	Polar	522					2.00	-1.25	0.00	-1.28	2.00	-1.25	0.00	-1.28

Table 3: Mean estimates and standard deviations (in parentheses) of iQTL regression effects

Sample size	QTL No.	a_0	a_1	a_2	a_3	d_0	d_1	d_2	d_3	i_0	i_1	i_2	i_3
250	1	1.77(0.23)	-0.95(0.35)	-1.17(0.40)	-0.92(0.33)								
	2	-0.07(0.24)	1.68(0.36)	2.47(0.45)	1.16(0.39)					-0.03(0.31)	1.63(0.28)	2.42(0.49)	1.17(0.33)
	3	2.58(0.43)	1.42(0.30)	-1.97(0.41)	-1.30(0.43)					2.53(0.44)	1.39(0.35)	-1.99(0.48)	-1.32(0.51)
	4									3.01(0.49)	0.04(0.18)	-0.98(0.33)	1.79(0.31)
	5					1.96(0.32)	-1.31(0.31)	0.05(0.25)	-1.19(0.40)	1.91(0.31)	-1.36(0.40)	0.02(0.21)	-1.15(0.47)
500	1	1.86(0.21)	-0.85(0.33)	-1.18(0.34)	-0.84(0.25)								
	2	0.02(0.19)	1.67(0.22)	2.49(0.38)	1.18(0.30)					0.01(0.22)	1.64(0.23)	2.45(0.34)	1.15(0.28)
	3	2.56(0.25)	1.40(0.24)	-2.04(0.32)	-1.30(0.29)					2.50(0.23)	1.44(0.28)	-2.00(0.35)	-1.34(0.26)
	4									2.96(0.28)	0.01(0.11)	-1.07(0.24)	1.74(0.25)
	5					2.03(0.19)	-1.27(0.24)	0.03(0.18)	-1.25(0.24)	1.97(0.21)	-1.16(0.20)	0.01(0.15)	-1.30(0.20)

Table 4: Mean estimates and standard deviations (in parentheses) of iQTL positions and statistical power of iQTL detection

Sample Size		QTL No.				
		1	2	3	4	5
250	Position	21.3(3.6)	147.3(4.2)	258.7(5.4)	334.1(5.8)	524.7(6.1)
	Power (%)	84	72	80	90	84
500	Position	22.9(3.3)	148.5(3.7)	257.9(4.1)	333.7(4.6)	520.5(5.0)
	Power (%)	100	96	100	100	98

Table 5: Estimates for positions and regression coefficients of iQTLs for body weight growth in mice

QTL	Position (Chr.-cM)	a_0	a_1	a_2	a_3	d_0	d_1	d_2	d_3	i_0	i_1	i_2	i_3
1	2-104.3	0.254	0.633	0.166	0.033					0.074	0.180	0.061	0.013
2	4-38.1					-0.070	-0.360	-0.132	0.036	-0.050	-0.118	-0.028	0.028
3	6-73.2	0.425	0.525	0.075	0.055					0.162	0.210	0.060	0.032
4	7-63.1	0.787	0.851	-0.044	0.103					0.185	0.210	0.016	0.141
5	10-72.7					0.027	-0.165	0.018	-0.018	0.023	-0.061	0.004	-0.006
6	15-12.7	0.896	0.311	-0.314	0.062					0.117	0.071	-0.035	0.034

Table 6: 2logBFs for iQTL effects and imprinting types for detected iQTLs for body weight growth in mice

QTL	a	d	i	Imprinting Type
1	3.29	0.64	2.20	Additive-Paternal
2	1.01	3.34	4.47	Dominance-Over
3	4.72	1.83	3.24	Additive-Paternal
4	2.95	0.54	2.97	Additive-Paternal
5	0.87	4.32	3.68	Dominance-Over
6	4.46	1.70	4.23	Additive-Paternal