

Research

Open Access

## Functional mapping imprinted quantitative trait loci underlying developmental characteristics

Yuehua Cui\*, Shaoyu Li and Gengxin Li

Address: Department of Statistics & Probability, Michigan State University, East Lansing, MI 48824, USA

Email: Yuehua Cui\* - cui@stt.msu.edu; Shaoyu Li - lishaoyu@stt.msu.edu; Gengxin Li - ligengxi@stt.msu.edu

\* Corresponding author

Published: 17 March 2008

Received: 18 January 2008

*Theoretical Biology and Medical Modelling* 2008, **5**:6 doi:10.1186/1742-4682-5-6

Accepted: 17 March 2008

This article is available from: <http://www.tbiomed.com/content/5/1/6>

© 2008 Cui et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

**Background:** Genomic imprinting, a phenomenon referring to nonequivalent expression of alleles depending on their parental origins, has been widely observed in nature. It has been shown recently that the epigenetic modification of an imprinted gene can be detected through a genetic mapping approach. Such an approach is developed based on traditional quantitative trait loci (QTL) mapping focusing on single trait analysis. Recent studies have shown that most imprinted genes in mammals play an important role in controlling embryonic growth and post-natal development. For a developmental character such as growth, current approach is less efficient in dissecting the dynamic genetic effect of imprinted genes during individual ontology.

**Results:** Functional mapping has been emerging as a powerful framework for mapping quantitative trait loci underlying complex traits showing developmental characteristics. To understand the genetic architecture of dynamic imprinted traits, we propose a mapping strategy by integrating the functional mapping approach with genomic imprinting. We demonstrate the approach through mapping imprinted QTL controlling growth trajectories in an inbred F<sub>2</sub> population. The statistical behavior of the approach is shown through simulation studies, in which the parameters can be estimated with reasonable precision under different simulation scenarios. The utility of the approach is illustrated through real data analysis in an F<sub>2</sub> family derived from LG/J and SM/J mouse strains. Three maternally imprinted QTLs are identified as regulating the growth trajectory of mouse body weight.

**Conclusion:** The functional iQTL mapping approach developed here provides a quantitative and testable framework for assessing the interplay between imprinted genes and a developmental process, and will have important implications for elucidating the genetic architecture of imprinted traits.

### Background

Hunting for genes underlying mendelian disorders or quantitative traits has been a long-term effort in genetical research. Most current statistical approaches to gene mapping assume that the maternally and paternally derived copies of a gene in diploid organisms have a comparable

level of expression. This, however, is not necessarily true as revealed by recent studies, in which some genes show asymmetric expression, and their expression in the offspring depends on the parental origin of their alleles [1-3]. This phenomenon, termed genomic imprinting, results from the modification of DNA structure rather than

changes in the underlying DNA sequences. As one type of epigenetic phenomenon, genomic imprinting has greatly shaped modern research in genetics since its discovery. Some previously puzzling genetic phenomena can now be explained by imprinting theory. However, little is known about the size, location and functional mechanism of imprinted genes in development.

The selective control of gene imprinting is unique to placental mammals and flowering plants. There is increasing evidence that many economically important traits and human diseases are influenced by genomic imprinting [3-6]. More recent studies have shown that genomic imprinting might be even more common than previously thought [7]. Despite its importance, the study of genomic imprinting is still in its early infancy. The biological function of genomic imprinting in shaping an organism's development is still unclear. Recent publications have shown that the majority of imprinted genes in mammals play an important role in controlling embryonic growth and development [8,9], and some involve in post-natal development, affecting suckling and metabolism [9,10]. The malfunction of imprinted genes at any developmental stage could lead to substantially abnormal characters such as cancers or other genetic disorders. It is therefore of paramount importance to identify imprinted genes and to understand at which developmental stage they function, to help us explore opportunities to prevent, control and treat diseases therapeutically. With the development of new biotechnology coupled with computationally efficient statistical tools, it is now possible to map imprinted genes and understand their roles in disease susceptibility.

Several studies have shown that the effects of imprinted quantitative trait loci (iQTL) can be estimated and tested in controlled crosses of inbred or outbred lines [6,11-15]. These approaches are designed on the traditional QTL mapping framework where a phenotypic trait is measured at certain developmental stage for a mapping subject, ignoring the dynamic features of gene expression. As a highly complex process, genomic imprinting involves a number of growth axes operating coordinately at different development stages [16]. Changes in gene expression at different developmental stages reflect the dynamic changes of gene function over time. They also reflect the response of an organism to either internal or external stimuli, so it can redirect its developmental trajectory to adapt better to environmental conditions, and thereby to increase its fitness [17]. For this reason, incorporating such information into genetic mapping should provide more information about the genetic architecture of a dynamic developmental trait.

When a developmental feature of an imprinted trait is considered, traditional iQTL mapping approaches that

only consider the phenotypic trait measured at a particular time point will be inappropriate for such an analysis. In fact, for a quantitative trait of developmental behavior, the genetic effect at time  $t$  (denoted as  $G_t$ ) is composed of the genetic effect at time  $t - 1$  (denoted as  $G_{t-1}$ ) and the extra genetic effect from time  $t - 1$  to  $t$  (denoted as  $G_{\Delta t}$ ) [18]. Therefore, the phenotypic trait measured at time  $t$  reflects the cumulative gene effects from initial time to  $t$ , and is highly correlated with the trait measured at time  $t - 1$ . The correlations among traits measured at different time periods (i.e., different developmental stages) thus provide correlation information about gene expressions, and hence tell us how genes mediate to respond to internal and external stimuli. Current imprinting QTL (iQTL) mapping approaches, by ignoring the correlations among traits measured at different developmental stages, could therefore potentially overestimate the number and the effective size of iQTLs, and lead to wrong inferences.

Although conditional QTL analysis can reduce bias and increase detecting power by partitioning the genetic effect in a conditional manner [18], analysis of traits at each measurement time point is still less powerful and less attractive than analysis by considering measurements at different developmental stages jointly [19]. The recent development of functional mapping brings challenges as well as opportunities for mapping genes responsible for dynamic features of a quantitative trait [17,19,20]. Functional mapping is the integration between genetic mapping and biological principles through mathematical equations. The relative merits of functional mapping in biology lie in the strong biological relevance of QTL detection, and its statistical advantages are that it reduces data dimensions and increases the power and stability of QTL detection. By incorporating various mathematical functions into the mapping framework, functional mapping has great flexibility for mapping genes that underlie complex dynamic/longitudinal traits. It provides a quantitative framework for assessing the interplay between genetic function and developmental pattern and form.

In this article, we extend our previous work of interval iQTL mapping to functional iQTL mapping by incorporating biologically meaningful mathematical functions into a QTL mapping framework. We illustrate the idea through an inbred line  $F_2$  design, although it can be easily extended to other genetic designs. To distinguish the genetic differences between the two reciprocal heterozygous forms derived from an  $F_2$  population, information about sex-specific differences in the recombination fraction is used. Monte Carlo simulations are performed to evaluate the model performance under different scenarios considering the effect of sample size, heritability and imprinting mechanism. A real example is illustrated in which three iQTLs affecting the growth trajectory of body

weight in an  $F_2$  family derived from two different mouse strains are identified through a genome-wide linkage scan.

**Methods**

**Functional QTL Mapping**

Statistical methods for mapping QTL underlying developmental characteristics such as growth or HIV dynamics have been developed previously [19,20]. The so called functional mapping approach has been recently applied to mapping QTL underlying programmed cell death [21,22]. Functional mapping is derived under the finite mixture model-based likelihood framework. In the mixture model, each observation  $y$  is modelled as a mixture of  $J$  (known and finite) components. The distribution for each component corresponds to the genotype category depending on the underlying genetic design. For an  $F_2$  design, there are three mixture components ( $J = 3$ ). The density function for each genotype component is assumed to follow a parametric distribution ( $f$ ) such as Gaussian, which can be expressed as:

$$y \sim p(y | \bar{p}, \bar{j}, h) = p_1 f_1(y; j_1, h) + p_2 f_2(y; j_2, h) + p_3 f_3(y; j_3, h)$$

where  $\bar{\pi} = (\pi_1, \pi_2, \pi_3)$  is a vector of mixture proportions which are constrained to be non-negative and sum to unity;  $\bar{\varphi} = (\varphi_1, \varphi_2, \varphi_3)$  is a vector for the component specific parameters, with  $\varphi_j$  being specific to component  $j$ ; and  $\eta$  contains parameters (i.e., residual variance) that are common to all components.

For an  $F_2$  design initiated with two contrasting homozygous inbred lines, there are three genotypes at each locus. Suppose there is a putative segregating QTL with alleles  $Q$  and  $q$  that affects a developmental trait such as growth. In a QTL mapping study, the QTL genotype is generally considered as missing, but can be inferred from the two flanking markers. The missing QTL genotype probability  $\pi_j$  can be calculated as the conditional probability of the QTL genotype given the observed flanking marker genotypes. For a population with structured pedigree like an  $F_2$  population, it can be expressed in terms of the recombination fractions, whereas for a natural population, it can be expressed as a function of linkage disequilibrium. The derivations of the conditional probabilities of QTL genotypes can be found in the general QTL mapping literature [23].

In functional mapping, the parameters  $\bar{\varphi} = (\varphi_1, \varphi_2, \varphi_3)$  specify the underlying developmental mean function ( $\mathbf{m}$ ). For an  $F_2$  design, there are three sets of mean functions corresponding to three QTL genotypes. To reduce the number of parameters and enhance the interpretability of

functional mapping, the mean process is modelled by certain biologically meaningful mathematical functions, either parametrically or nonparametrically. Suppose that the phenotypic traits are acquired from  $n$  individuals, and that  $t$  measurements are made on each individual  $i$ . Let the response of individual  $i$  at time  $t$  be denoted by  $y_i(t)$ ,  $i = 1, \dots, n$ ;  $t = 1, \dots, \tau$ . Then the response can be modelled as

$$y_i(t) = f(t) + e_i(t),$$

where  $f(t)$  is a linear or nonlinear function evaluated at time  $t$ , depending on the underlying developmental pattern;  $e_i(t)$  is the residual error, which is assumed to be normal with mean zero and variance  $\sigma^2(t)$ . The intra-individual correlation is specified as  $\rho$ , which leads to the covariance for individual  $i$  at two different time points,  $t_1$  and  $t_2$ , expressed as  $\text{cov}(y_i(t_1), y_i(t_2)) = \rho \sigma_{t_1} \sigma_{t_2}$ . Assuming multivariate normal distribution, the density function for each progeny  $i$  who carries genotype  $j$  can be expressed as

$$f_j(y_i | \mathbf{w}_q, \mathbf{w}_r) = \frac{1}{(2\pi)^T |S|^{1/2}} \exp \left[ -\frac{1}{2} (y_i - \mathbf{m}_j) S^{-1} (y_i - \mathbf{m}_j)^T \right],$$

where  $\mathbf{m}_j = [m_j(1), \dots, m_j(\tau)]$  is the mean vector common for all individuals with genotype  $j$ , which can be evaluated through function  $f$  in Model (2). The unknown parameters that specify the position of QTL within a marker interval are arrayed in  $\Omega_r$ . The parameters that define the mean and the covariance functions are arrayed in  $\Omega_q$ .

Since we do not observe the QTL genotype, the distribution of  $y$  is modelled through a finite mixture model given in Model (1). At a particular time point (say  $t$ ), the genetic effect can be obtained by solving the following equations

$$a(t) = \frac{1}{2} (m_1(t) - m_3(t)) \text{ and } d(t) = m_2(t) - \frac{1}{2} (m_1(t) + m_3(t))$$

where  $a(t)$  and  $d(t)$  are the additive and dominant effects at time  $t$ , respectively.

**Functional iQTL Mapping**

*Modelling the imprinted mean function*

In an  $F_2$  population, three QTL genotypes are segregated. The three QTL genotypes may have different expressions which result in three different mean trajectories. Considering the imprinting property of an iQTL, we introduce the notation for the parental origin of alleles inherited from both parents. Let  $Q_M$  and  $q_M$  be two alleles inherited from the maternal parent, and  $Q_P$  and  $q_P$  be two alleles derived from the paternal parent. The subscripts  $M$  and  $P$

refer to maternal and paternal origin, respectively. These four parentally specific alleles form four distinct genotypes expressed as  $Q_M Q_P$ ,  $Q_M q_P$ ,  $q_M Q_P$ , and  $q_M q_P$ . In contrast, in a regular QTL mapping study without distinguishing the allelic parental origin, the two reciprocal heterozygotes,  $Q_M q_P$  and  $q_M Q_P$ , are collapsed to one heterozygote. When a QTL is imprinted, the four QTL genotypes show different gene expressions, which result in different developmental growth trajectories. For a maternally (or paternally) imprinted QTL, the allele inherited from the maternal (or paternal) parent is not expressed. Thus, two growth trajectories would be expected. By testing the differences of the four growth trajectories, one can test whether there is a QTL, and whether the QTL is imprinted.

For simplicity, we use numerical notation to denote the four parent-of-origin-specific genotypes, i.e.,  $Q_M Q_P = 1$ ,  $Q_M q_P = 2$ ,  $q_M Q_P = 3$ , and  $q_M q_P = 4$ . The mean functions of these genotypes are denoted as  $m_j$ , ( $j = 1, \dots, 4$ ). We know that for an imprinted gene, the expression of an allele depends on its parental origin. On a developmental scale, the two reciprocal heterozygotes,  $Q_M q_P$  and  $q_M Q_P$ , may present different mean trajectories. The degree of imprinting of an iQTL can thus be assessed by the genotype-specific parameters. Through testing the difference between the mean functions of the two reciprocal heterozygotes, we can assess the imprinting property of a QTL. An overlap of the two trajectories for the two reciprocal heterozygotes indicates no sign of imprinting.

For a developmental characteristic such as growth, it is well known that the underlying trajectory can be described by a universal growth law, which follows a logistic growth function [24]. At a developmental stage, say time  $t$ , the mean value of an individual carrying QTL genotype  $j$  can be expressed by

$$m_j(t) = \frac{a_j}{1 + b_j e^{-g_j t}}$$

where the growth parameters ( $\alpha_j$ ,  $\beta_j$ ,  $\gamma_j$ ) describe asymptotic growth, initial growth and relative growth rate, respectively [25]. With estimated growth parameters, we can easily retrieve the genotypic means at every time point by simply plugging  $t$  into Equation (5). This modelling approach can significantly reduce the number of unknown parameters to be estimated, especially when the number of measurement points is large [19].

At a particular time point (say  $t$ ), the mean expression of an individual carrying QTL genotype  $j$  can be evaluated through the three growth parameters ( $\alpha_j$ ,  $\beta_j$ ,  $\gamma_j$ ). On the basis of the univariate imprinting model given in [12], we

can partition the genetic effects at time  $t$  as the allele-specific effects, i.e.

$$\begin{aligned} a_M(t) &= \frac{1}{2}(m_1(t) + m_2(t) - m_3(t) - m_4(t)) \\ a_P(t) &= \frac{1}{2}(m_1(t) - m_2(t) + m_3(t) - m_4(t)) \\ d(t) &= m_1(t) - m_2(t) - m_3(t) + m_4(t) \end{aligned}$$

where  $a_M$  and  $a_P$  refer to the additive effects of alleles inherited from mother and father, respectively;  $d$  refers to the allele dominant effect.

To illustrate the idea, we use the growth trait to demonstrate the mapping principle. The idea can be easily extended to other developmental characteristics. For developmental characteristics other than growth, different mathematical functions should be developed. Some flexible choices include nonparametric regressions based on smoothing splines or orthogonal polynomials [21].

#### Modelling the covariance structure

To understand how QTL mediate growth, it is essential to take correlations among repeated measures into account [19]. The repeated measures provide correlation information on gene expression. Hence, dissection of the intra-individual correlation will help us to understand better how genes function over time. One commonly used model for covariance structure modelling is the first-order autoregressive (AR(1)) model [26], expressed as

$$\sigma^2(1) = \rho + \sigma^2(\tau) = \sigma^2$$

for the variance, and

$$s(t_k, t_{k'}) = s^2 r^{|t_k - t_{k'}|} (t_{k'} > t_k)$$

for the covariance between any two time points  $t_k$  and  $t_{k'}$ , where  $0 < \rho < 1$  is the proportion parameter with which the correlation decays with time lag.

For a developmental characteristic such as growth, the inter-individual variation generally increases as time increases, which leads to a nonstationary variance function. Since the AR(1) covariance model assumes stationary variance, it can not be applied directly. To stabilize the variance at different measurement time points, we apply a multivariate Box-Cox transformation to stabilize the variance [27], which has the form

$$z_i(t) = \begin{cases} \frac{y_i(t)^l - 1}{l(t)}, & \text{if } l \neq 0 \\ \log(y_i(t)), & \text{if } l = 0 \end{cases}$$

The Box-Cox transformation ensures the homoscedasticity and normality of the response  $y$ . For repeated measures or longitudinal studies, a reasonable constraint is to set  $\lambda(t) = \lambda$  for all  $t$ . Then the optimal choice of  $\lambda$  can be estimated from the data. To preserve the interpretability of the estimated mean parameters, Carroll and Ruppert [28] proposed a transform-both-sides (TBS) model in which the same transformation form is applied to both sides of Model (2). For a log-transformation, this results in  $\log y_i(t) = \log f(t) + e_i(t)$ . Wu et al. [29] later showed the favorable property of this approach in functional mapping. For the modelling purpose of stabilizing variances, we simply adopt the log-transformation in the current setting.

Alternatively, one can model the covariance structure nonstationarily without transforming the original data. Among a pool of choices, the structured antedependence (SAD) model [30] displays a number of favorable merits. The SAD model of order  $p$  for modelling the error term in Eq. (2) is given by

$$e_i(t) = \phi_1 e_i(t-1) + \cup + \phi_p e_i(t-p) + \varepsilon_i(t)$$

where  $\varepsilon_i(t)$  is the "innovation" term assumed to be independent and distributed as  $\mathcal{N}(0, \sigma_\varepsilon^2)$ . Therefore, the variance-covariance matrix can be expressed as

$$\Sigma = \mathbf{A} \Sigma_\varepsilon \mathbf{A}^T,$$

where  $\Sigma_\varepsilon$  is a diagonal matrix with diagonal elements being the innovation variance;  $\mathbf{A}$  is a lower triangular matrix which contains the antedependence coefficient  $\phi_r$ . The SAD order ( $p$ ) can be selected through an information criterion [31]. The SAD( $r$ ) model has been previously applied in functional mapping of programmed cell death [21].

**Parameter Estimation**

Assuming inter-individual independence, the joint likelihood function is given by

$$L(\mathbf{w} | \mathbf{z}, \mathcal{M}) = \prod_{i=1}^n \sum_{j=1}^4 \pi_{ji} f_j(\mathbf{z}_i | \mathbf{w}, \mathcal{M})$$

where  $\mathbf{z}_i = [z_i(1), \cup, z_i(\tau)]$  is the observed log-transformed trait vector for individual  $i$  ( $i = 1, \cup, n$ ) over  $\tau$  time points;  $f_j$  is the multivariate normal density function with log-transformed mean for QTL genotype  $j$ ;  $\pi_{ji}$  ( $j = 1, \cup, 4$ ) is the mixture proportion for individual  $i$  with genotype  $j$ , which is derived assuming a sex-specific difference in recombination rate and can be found in [12]. The

unknown parameters in  $\Omega$  comprise three sets, one defining the co-segregation between the QTL and markers and thereby the location of the QTL relative to the markers, denoted by  $\Omega_r$ , and the other defining the distribution of a growth trait for each QTL genotype, denoted by  $\Omega_q = (\Omega_m, \Omega_v)$ , where  $\mathbf{w}_m = (\mathbf{w}_{m_1}, \mathbf{w}_{m_2}, \mathbf{w}_{m_3}, \mathbf{w}_{m_4})$  defines the mean vector for different genotypes and  $\Omega_v$  defines the covariance parameters.

We implement the EM algorithm to obtain the maximum likelihood estimates (MLEs) of the unknown parameters. The first derivative of the log-likelihood function, with respect to specific parameter  $\varphi$  contained in  $\Omega$ , is given by

$$\begin{aligned} & \frac{\partial}{\partial \mathbf{w}_j} \log \ell(\mathbf{w} | \mathbf{z}, \mathcal{M}) \\ &= \sum_{i=1}^n \sum_{j=1}^4 \frac{\pi_{ji} \frac{\partial}{\partial \mathbf{w}_j} f_j(\mathbf{z}_i | \mathbf{w}, \mathcal{M})}{\sum_{j'=1}^4 \pi_{j'i} f_{j'}(\mathbf{z}_i | \mathbf{w}, \mathcal{M})} \\ &= \sum_{i=1}^n \sum_{j=1}^4 \frac{\pi_{ji} f_j(\mathbf{z}_i | \mathbf{w}, \mathcal{M})}{\sum_{j'=1}^4 \pi_{j'i} f_{j'}(\mathbf{z}_i | \mathbf{w}, \mathcal{M})} \frac{\partial}{\partial \mathbf{w}_j} \log f_j(\mathbf{z}_i | \mathbf{w}, \mathcal{M}) \\ &= \sum_{i=1}^n \sum_{j=1}^4 \Pi_{ji} \frac{\partial}{\partial \mathbf{w}_j} \log f_j(\mathbf{z}_i | \mathbf{w}, \mathcal{M}) \end{aligned}$$

where we define

$$\Pi_{ji} = \frac{\pi_{ji} f_j(\mathbf{z}_i | \mathbf{w}, \mathcal{M})}{\sum_{j'=1}^4 \pi_{j'i} f_{j'}(\mathbf{z}_i | \mathbf{w}, \mathcal{M})}$$

The MLEs of the parameters contained in  $(\Omega_m, \Omega_v)$  are obtained by solving

$$\frac{\partial}{\partial \mathbf{w}_j} \log \ell(\mathbf{w} | \mathbf{z}, \mathcal{M}) = 0$$

Direct estimation is unavailable since there is no closed form for the MLEs of parameters. The EM algorithm is applied to solve these unknowns iteratively.

**E-step:** Given initial values for  $(\Omega_m, \Omega_v)$ , calculate the posterior probability matrix  $\Pi = \{\Pi_{ji}\}$  in Eq. (8).

**M-step:** With the updated posterior probability  $\Pi$ , we can update the parameters contained in  $\Omega_q$ . The maximization can be implemented through an iteration procedure or through the Newton-Raphson or other algorithm such as simplex algorithm [32].

The above procedures are iteratively repeated between (8) and (9), until a certain convergence criterion is met. For details of the EM algorithm, one can refer to [19]. The converged values are the MLEs of the parameters. The initial values under the alternative hypothesis are generally set as the estimated values under the null. Note also that in the above algorithm, we do not directly estimate the QTL-segregating parameters ( $\Omega_r$ ). In general, we use a grid search approach to estimate the QTL location by searching for a putative QTL at every 1 or 2 cM on a map interval bracketed by two markers throughout the entire linkage map. The log-likelihood ratio test statistic for a QTL at a testing position is displayed graphically to generate a log-likelihood ratio plot called LR profile plot. The genomic position corresponding to a peak of the profile is the MLE of the QTL location.

We have found that the algorithm is sensitive to initial values, particularly the mean values of the two reciprocal heterozygotes. To make sure the parameters are converged to the "correct" ones, we normally give different initial values for the two reciprocal heterozygotes and check which one produces the highest likelihood value. The ones which produce higher likelihood value are considered as the MLEs.

**Hypothesis Testing**

*Global QTL test*

Testing whether there is a QTL affecting the developmental trajectory is the first step toward understanding of genetic architecture of an imprinted trait. Once the MLEs of the parameters are obtained, the existence of a QTL affecting the growth curve can be tested by formulating the following hypotheses

$$\begin{cases} H_0 & : \mathbf{w}_{m_1} \equiv \dots \equiv \mathbf{w}_{m_4} \\ H_1 & : \text{The equalities above do not hold,} \end{cases}$$

where  $H_0$  corresponds to the reduced model, in which the data can be fit by a single curve, and  $H_1$  corresponds to the full model, in which there exist different curves to fit the data. The above test is equivalent to test

$$\begin{cases} H_0 & : a_1 = a_2 = a_3 = a_4, b_1 = b_2 = b_3 = b_4, g_1 = g_2 = g_3 = g_4 \\ H_1 & : \text{The equalities above do not hold,} \end{cases}$$

The statistic for testing the hypotheses is calculated as the log-likelihood (LR) ratio of the reduced to the full model

$$LR = -2[\log L(\tilde{\mathbf{w}} | \mathbf{z}, \mathcal{M}) - \log L(\hat{\mathbf{w}} | \mathbf{z}, \mathcal{M})]$$

where  $\tilde{\mathbf{w}}$  and  $\hat{\mathbf{w}}$  denote the MLEs of the unknown parameters under  $H_0$  and  $H_1$ , respectively. An empirical

approach to determining the critical threshold is based on permutation tests [33].

*Imprinting test*

Rejection of the null hypothesis in Test (10) at a particular genomic position indicates evidence of a QTL at that locus. Next, we would like to know the imprinting property of a detected QTL. To test if a detected QTL is imprinted or not, we develop the following hypothesis

$$\begin{cases} H_0 & : a_2 = a_3, b_2 = b_3, g_2 = g_3 \\ H_1 & : \text{The equalities above do not hold,} \end{cases}$$

The null hypothesis states that the two reciprocal QTL genotypes have the same mean curve and hence have the same gene expression, i.e., the expressions of genotypes  $Q_M q_P$  and  $q_M Q_P$  are independent of allelic origin. Rejection of the null hypothesis indicates evidence of genomic imprinting.

Following Test (11), if the null is rejected, further tests can be done to test whether an iQTL is maternally imprinted or paternally imprinted. The following hypothesis tests can be formulated

$$\begin{cases} H_0 & : a_1 = a_2, b_1 = b_2, g_1 = g_2 \\ H_1 & : \text{The equalities above do not hold,} \end{cases}$$

for testing paternally imprinted QTL and

$$\begin{cases} H_0 & : a_1 = a_3, b_1 = b_3, g_1 = g_3 \\ H_1 & : \text{The equalities above do not hold,} \end{cases}$$

for testing maternally imprinted QTL.

The null hypothesis in Test (12) states that the two QTL genotypes  $Q_M Q_P$  and  $Q_M q_P$  have the same mean curves and hence same expressions (i.e., allele inherited from the paternal parent does not express).

The iQTL identified can then be claimed as a paternally imprinted QTL. Similarly, if one fails to reject the null in Test (13), the conclusion that there is maternal imprinting can be reached.

Note that the imprinting test (11) is only conducted at the position where a significant QTL is declared on the basis of Test (10). So Test (11) is a point test. Tests (12) and (13) are only conducted when the null in Test (11) is rejected. We can either use the likelihood ratio test or a nonparametric test based on the area under the curve (AUC). The idea of the AUC test is that if two genotypes have the same expression, the area under the developmental curve would be the same. The AUC for QTL genotype  $j$  is defined as

$$AUC_j |_1^t = \int_1^t \frac{a_j}{1+b_j e^{-g_j t}} dt = \frac{a_j}{g_j} \log\left(\frac{b_j + e^{t g_j}}{b_j + e^{g_j}}\right)$$

Similarly, Tests (11)–(13) can be defined accordingly based on the AUC. For example, to test (12), the hypothesis would be simplified to

$$\begin{cases} H_0 & : AUC_1 |_1^t = AUC_2 |_1^t \\ H_1 & : AUC_1 |_1^t \neq AUC_2 |_1^t \end{cases}$$

The significance of Tests (11)–(13) can be evaluated on the basis of permutations. In our simulation study, we found that the test based on the AUC is more sensitive and powerful than the one based on the likelihood ratio test.

**Regional test**

Even though a mean curve can be modelled throughout a continuous function, genes may not function across all the observed stages. For imprinted genes, loss of imprinting (LOI) is reported in the literature [34]. The question of how a QTL exerts its effects on an interval across a growth trajectory (say  $[t_1, t_2]$ ) can be tested using a regional test approach based on the AUC. The AUC for genotype  $j$  at a given time interval is calculated as

$$AUC_j = \int_{t_1}^{t_2} \frac{\alpha_j}{1+\beta_j e^{-\gamma_j t}} dt$$

If the AUCs of the four genotypes for a testing period  $[t_1, t_2]$  are the same, we claim there is no QTL effect at that time interval. The hypothesis test for the genetic effect over a period of growth can be formulated as

$$\begin{cases} H_0 & : AUC_1 |_{t_1}^{t_2} = \dots = AUC_4 |_{t_1}^{t_2} \\ H_1 & : \text{The equalities above do not hold} \end{cases}$$

This test can detect if a QTL exerts an early gene effect or triggers a late effect.

**Results**

**Monte Carlo Simulation**

Monte Carlo simulations are performed to evaluate the statistical behavior of the developed approach. Consider an  $F_2$  population initiated with two contrasting inbred lines, with which a 100 cM long linkage group composed of 6 equidistant markers is constructed. A putative QTL that affects the imprinted growth process is located at 46 cM from the first marker on the linkage group. The marker genotypes in the  $F_2$  family are simulated by mimicking sex-specific recombination fractions in mice, i.e.,  $r_M = 1.25r_p$ . The Haldane map function is used to convert the

map distance into the recombination fraction. Data are simulated with different specifications, namely different heritability levels ( $H^2 = 0.1$  vs  $0.4$ ) and different sample sizes ( $n = 200$  vs  $500$ ). For each  $F_2$  progeny, its phenotype is simulated with 10 equally spaced time points. The covariance structure is simulated assuming the first-order AR(1) model. Note that the variance parameter ( $\sigma^2$ ) is calculated on the basis of the log-transformed data.

Several data sets are simulated assuming no imprinting, partial imprinting, complete maternal and paternal imprinting. The simulation results are summarized in Tables 1, 2, 3, 4. As we expected, the precision of parameter estimates is increasing with the increase of the sample size and heritability under different imprinting scenarios. For example, when a QTL is not imprinted (Table 4), the RMSE of the parameter  $a$  for genotype  $Q_M Q_P$  decreases from 0.397 to 0.327, an 18% increase in precision when the sample size increases from 200 to 500 with fixed heritability level (0.1). For the same parameter, when a QTL is completely maternally imprinted, a reduction in RMSE from 0.478 to 0.305 is observed (Table 1). When we fix the sample size and increase the heritability level, the reduction in RMSE is even more noteworthy. For example, under fixed sample size ( $n = 200$ ), the RMSE of the parameter  $a$  for QTL genotype  $Q_M Q_P$  is reduced from 0.397 to 0.137, a 65% increase in precision compared to an 18% increase when sample size increases from 200 to 500 with fixed heritability (Table 4). Large heritability infers high genetic variability and low environmental variation [35]. Therefore, to increase the precision of parameter estimation, well managed experiments in which environmental variation is reduced is more important than just simply increasing sample sizes.

Under different simulation scenarios, another general trend is that the estimation for the genetic parameters of the two homozygotes performs better than that for the two reciprocal heterozygotes. For example, the RMSE of the growth parameter  $a$  for  $Q_M q_P$  is 0.765, while it decreases to 0.397 for genotype  $Q_M Q_P$  with fixed sample size 200 and heritability level 0.1 (Table 4). This is what we expected since partitioning the heterozygote into two parts may cause information loss. As the sample size or the heritability level increases, the RMSEs are greatly reduced for the two reciprocal heterozygous genotypes. For example, the RMSE (for parameter  $a$ ) is reduced from 0.765 to 0.304 when the heritability level increases from 0.1 to 0.4 under fixed sample size 200 (Table 4). Overall, the QTL position estimation is reasonably good under different simulation scenarios, even though the precision is reduced a little with completely imprinted models (Tables 1 and 2), compared with the non-imprinting and partial imprinting models (Tables 3 and 4).

**Table 1: The MLEs of the model parameters and the QTL position derived from 200 simulation replicates assuming complete maternal imprinting. The square root of the mean square errors (RMSEs) of the MLEs are given in parentheses.**

$H^2$	$n$	Position (cM)	$Q_M Q_P$			$Q_M q_P$			$q_M Q_P$			$q_M q_P$			Residual	
			$\alpha_1$ 36.5	$\beta_1$ 6.5	$\gamma_1$ 0.75	$\alpha_2$ 33.5	$\beta_2$ 5.5	$\gamma_2$ 0.75	$\alpha_3$ 36.5	$\beta_3$ 6.5	$\gamma_3$ 0.75	$\alpha_4$ 33.5	$\beta_4$ 5.5	$\gamma_4$ 0.75	$\sigma^2$	$\rho$ 0.8
0.1	200	45.31 (7.506)	36.52 (0.478)	6.50 (0.135)	0.75 (0.010)	33.74 (0.992)	5.60 (0.333)	0.75 (0.014)	36.22 (0.998)	6.38 (0.341)	0.75 (0.015)	33.47 (0.438)	5.50 (0.117)	0.75 (0.011)	0.0086 (0.001)	0.79 (0.014)
	NI	45.54 (8.267)	36.5685 (0.515)	6.53 (0.147)	0.75 (0.010)	35.03 (1.579)	5.99 (0.504)	0.75 (0.007)	35.03 (1.554)	5.99 (0.523)	0.75 (0.007)	33.42 (0.464)	5.48 (0.115)	0.75 (0.011)	0.009 (0.001)	0.81 (0.013)
0.1	500	45.88 (3.615)	36.54 (0.305)	6.51 (0.087)	0.75 (0.006)	33.72 (0.866)	5.57 (0.285)	0.75 (0.009)	36.32 (0.832)	6.42 (0.272)	0.75 (0.009)	33.49 (0.289)	5.50 (0.071)	0.75 (0.006)	0.0089 (0.0005)	0.80 (0.008)
	NI	46.12 (4.483)	36.59 (0.323)	6.53 (0.089)	0.75 (0.006)	34.97 (1.486)	5.97 (0.477)	0.75 (0.005)	34.97 (1.541)	5.97 (0.501)	0.75 (0.005)	33.41 (0.304)	5.48 (0.079)	0.75 (0.006)	0.009 (0.0006)	0.81 (0.01)
0.4	200	46.33 (2.671)	36.51 (0.206)	6.50 (0.057)	0.75 (0.004)	33.54 (0.352)	5.51 (0.111)	0.75 (0.004)	36.46 (0.361)	6.49 (0.112)	0.75 (0.004)	33.49 (0.186)	5.50 (0.045)	0.75 (0.004)	0.0015 (0.0003)	0.80 (0.014)
	NI	47.63 (4.625)	36.67 (0.251)	6.56 (0.080)	0.75 (0.004)	34.96 (1.502)	5.98 (0.495)	0.75 (0.005)	34.96 (1.582)	5.98 (0.532)	0.75 (0.004)	33.38 (0.212)	5.45 (0.065)	0.75 (0.004)	0.0018 (0.0003)	0.82 (0.027)
0.4	500	46.07 (1.684)	36.48 (0.119)	6.50 (0.031)	0.75 (0.002)	33.50 (0.127)	5.50 (0.034)	0.75 (0.003)	36.49 (0.145)	6.50 (0.037)	0.75 (0.003)	33.49 (0.123)	5.50 (0.030)	0.75 (0.002)	0.0015 (0.0001)	0.80 (0.007)
	NI	48.21 (2.644)	36.67 (0.215)	6.56 (0.072)	0.75 (0.002)	34.97 (1.487)	5.98 (0.486)	0.75 (0.002)	34.97 (1.551)	5.98 (0.526)	0.75 (0.002)	33.36 (0.182)	5.45 (0.057)	0.75 (0.003)	0.0018 (0.0003)	0.83 (0.029)

The location of the simulated QTL is described by the map distances (in cM) from the first marker of the linkage group (100 cM long). The hypothesized  $\sigma^2$  value is 0.009 for  $H^2 = 0.10$  and 0.0015 for  $H^2 = 0.4$ . The analysis results by non-imprinting model are indicated by "NI".

**Table 2: The MLEs of the model parameters and the QTL position derived from 200 simulation replicates assuming complete paternal imprinting. The square root of the mean square errors (RMSEs) of the MLEs are given in parentheses.**

$H^2$	$n$	Position (cM)	$Q_M Q_P$			$Q_M q_P$			$q_M Q_P$			$q_M q_P$			Residual	
			$\alpha_1$ 36.5	$\beta_1$ 6.5	$\gamma_1$ 0.75	$\alpha_2$ 36.5	$\beta_2$ 6.5	$\gamma_2$ 0.75	$\alpha_3$ 33.5	$\beta_3$ 5.5	$\gamma_3$ 0.75	$\alpha_4$ 33.5	$\beta_4$ 5.5	$\gamma_4$ 0.75	$\sigma^2$	$\rho$ 0.8
0.1	200	43.56 (8.519)	36.84 (0.567)	6.54 (0.141)	0.75 (0.010)	36.84 (0.974)	6.51 (0.297)	0.75 (0.016)	33.89 (0.904)	5.61 (0.293)	0.75 (0.017)	33.87 (0.585)	5.57 (0.140)	0.75 (0.012)	0.01 (0.0012)	0.81 (0.019)
0.1	500	45.27 (4.683)	36.90 (0.508)	6.56 (0.107)	0.75 (0.006)	36.88 (0.731)	6.55 (0.204)	0.75 (0.010)	33.87 (0.732)	5.56 (0.210)	0.75 (0.011)	33.82 (0.437)	5.55 (0.090)	0.75 (0.007)	0.009 (0.0005)	0.81 (0.011)
0.4	200	45.67 (3.217)	36.51 (0.193)	6.50 (0.050)	0.75 (0.003)	36.47 (0.374)	6.49 (0.114)	0.75 (0.004)	33.52 (0.348)	5.51 (0.108)	0.75 (0.004)	33.52 (0.168)	5.51 (0.046)	0.75 (0.004)	0.0015 (0.0004)	0.80 (0.012)
0.4	500	46.04 (1.579)	36.48 (0.118)	6.50 (0.035)	0.75 (0.002)	36.49 (0.129)	6.50 (0.039)	0.75 (0.002)	33.50 (0.121)	5.50 (0.033)	0.75 (0.003)	33.50 (0.111)	5.50 (0.028)	0.75 (0.003)	0.0015 (0.0001)	0.80 (0.008)

The location of the simulated QTL is described by the map distances (in cM) from the first marker of the linkage group (100 cM long). The hypothesized  $\sigma^2$  value is 0.009 for  $H^2 = 0.10$  and 0.0015 for  $H^2 = 0.4$ .



**Table 3: The MLEs of the model parameters and the QTL position derived from 200 simulation replicates assuming partial imprinting. The square root of the mean square errors (RMSEs) of the MLEs are given in parentheses.**

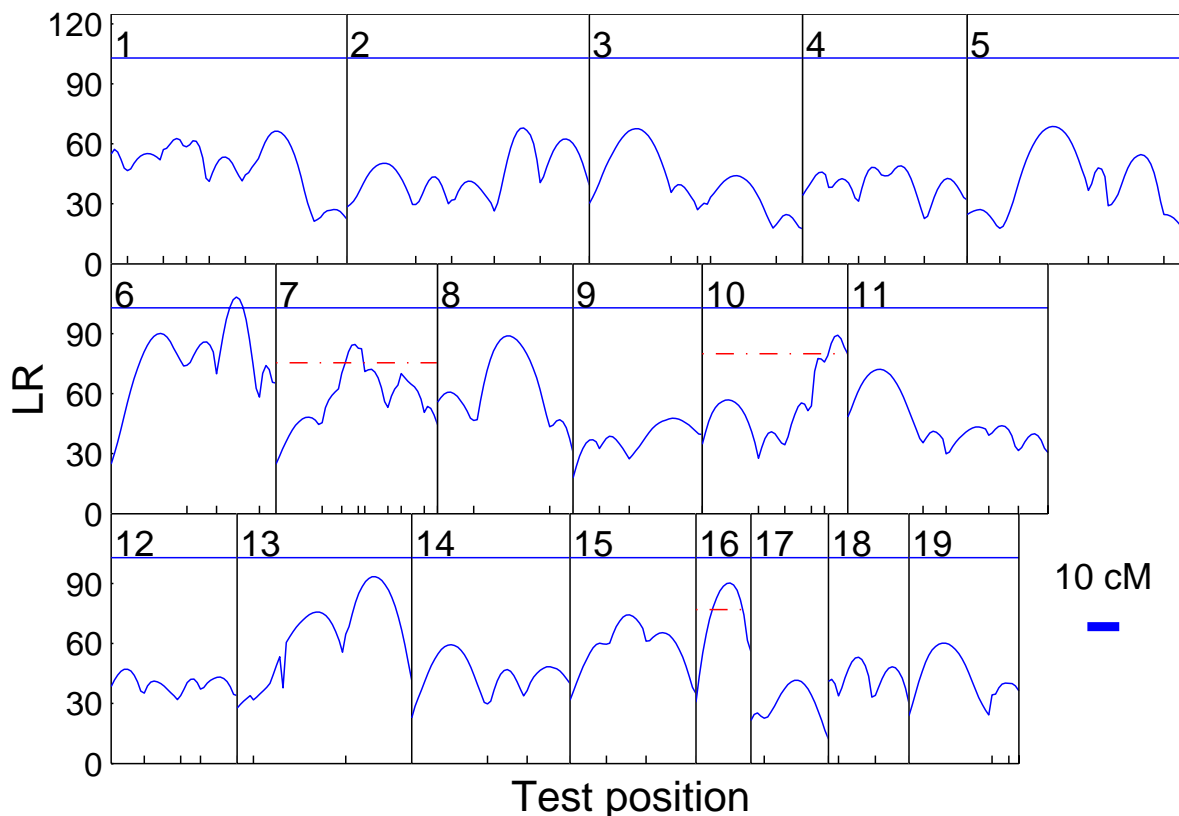
$H^2$	$n$	Position (cM)	$Q_M Q_P$			$Q_M q_P$			$q_M Q_P$			$q_M q_P$			Residual	
			$\alpha_1$ 36.5	$\beta_1$ 6.5	$\gamma_1$ 0.7	$\alpha_2$ 35.5	$\beta_2$ 6.5	$\gamma_2$ 0.7	$\alpha_3$ 34.5	$\beta_3$ 6	$\gamma_3$ 0.7	$\alpha_4$ 33.5	$\beta_4$ 5.5	$\gamma_4$ 0.7	$\sigma^2$	$\rho$ 0.8
0.1	200	45.37 (3.932)	36.51 (0.324)	6.49 (0.096)	0.70 (0.006)	35.18 (0.891)	6.27 (0.363)	0.70 (0.011)	34.85 (0.863)	6.23 (0.356)	0.70 (0.012)	33.51 (0.316)	5.51 (0.084)	0.70 (0.007)	0.0043 (0.001)	0.79 (0.014)
0.1	500	45.96 (2.206)	36.52 (0.225)	6.50 (0.060)	0.70 (0.004)	35.15 (0.686)	6.30 (0.321)	0.70 (0.008)	34.88 (0.716)	6.20 (0.323)	0.70 (0.008)	33.49 (0.201)	5.50 (0.052)	0.70 (0.004)	0.0044 (0.0008)	0.80 (0.011)
0.4	200	46.21 (1.787)	36.51 (0.134)	6.50 (0.038)	0.70 (0.002)	35.21 (0.566)	6.36 (0.268)	0.70 (0.003)	34.78 (0.572)	6.14 (0.271)	0.70 (0.003)	33.50 (0.123)	5.50 (0.032)	0.70 (0.0003)	0.0008 (0.0005)	0.79 (0.013)
0.4	500	46.17 (1.09)	36.51 (0.093)	6.50 (0.024)	0.70 (0.002)	35.29 (0.461)	6.40 (0.229)	0.70 (0.002)	34.72 (0.476)	6.11 (0.234)	0.70 (0.002)	33.50 (0.076)	5.50 (0.021)	0.70 (0.002)	0.0007 (0.0002)	0.80 (0.005)

The location of the simulated QTL is described by the map distances (in cM) from the first marker of the linkage group (100 cM long). The hypothesized  $\sigma^2$  value is 0.0045 for  $H^2 = 0.10$  and 0.00075 for  $H^2 = 0.4$ .

**Table 4: The MLEs of the model parameters and the QTL position derived from 200 simulation replicates assuming no imprinting. The square root of the mean square errors (RMSEs) of the MLEs are given in parentheses.**

$H^2$	$n$	Position (cM)	$Q_M Q_P$			$Q_M q_P$			$q_M Q_P$			$q_M q_P$			Residual	
			$\alpha_1$ 36.5	$\beta_1$ 6.5	$\gamma_1$ 0.7	$\alpha_2$ 35	$\beta_2$ 6	$\gamma_2$ 0.7	$\alpha_3$ 35	$\beta_3$ 6	$\gamma_3$ 0.7	$\alpha_4$ 33.5	$\beta_4$ 5.5	$\gamma_4$ 0.7	$\sigma^2$	$\rho$ 0.8
0.1	200	45.24 (4.351)	36.74 (0.397)	6.53 (0.096)	0.698 (0.006)	35.09 (0.765)	5.99 (0.188)	0.70 (0.013)	35.35 (0.855)	6.09 (0.204)	0.70 (0.014)	33.71 (0.379)	5.54 (0.090)	0.70 (0.007)	0.004 (0.0009)	0.80 (0.02)
0.1	500	46.01 (2.196)	36.74 (0.327)	6.54 (0.070)	0.70 (0.004)	35.17 (0.567)	6.00 (0.138)	0.70 (0.012)	35.28 (0.640)	6.07 (0.160)	0.70 (0.012)	33.69 (0.273)	5.53 (0.058)	0.70 (0.004)	0.004 (0.0004)	0.80 (0.01)
0.4	200	46.07 (1.731)	36.56 (0.137)	6.51 (0.037)	0.70 (0.002)	34.99 (0.304)	6.00 (0.076)	0.70 (0.005)	35.11 (0.318)	6.02 (0.076)	0.70 (0.005)	33.55 (0.127)	5.51 (0.033)	0.70 (0.003)	0.001 (0.0005)	0.80 (0.011)
0.4	500	46.17 (1.07)	36.56 (0.106)	6.51 (0.026)	0.70 (0.002)	35.03 (0.208)	6.00 (0.053)	0.70 (0.004)	35.07 (0.222)	6.01 (0.056)	0.70 (0.004)	33.54 (0.088)	5.51 (0.021)	0.70 (0.002)	0.0007 (0.0001)	0.80 (0.005)

The location of the simulated QTL is described by the map distances (in cM) from the first marker of the linkage group (100 cM long). The hypothesized  $\sigma^2$  value is 0.0041 for  $H^2 = 0.10$  and 0.0007 for  $H^2 = 0.4$ .



**Figure 1**  
**Genomewide likelihood ratio profile plot.** The profiles of the log-likelihood ratios (LR) between the full and reduced (no QTL) model estimated from the functional imprinting model for body mass growth trajectories across chromosome 1 to 19 using the linkage map constructed from microsatellite markers [36]. The threshold value for claiming the existence of QTLs is given as the horizontal dotted line for the genome-wide level and dashed line for the chromosome-wide level. The genomic positions above the threshold line and corresponding to the peaks of the curves are the MLEs of the QTL positions. The positions of markers on the linkage groups [36] are indicated at ticks.

Table 1 also summarizes the results of comparison between the imprinting and non-imprinting models, in which the regular non-imprinting functional mapping model is indicated by "NI". Data are simulated assuming complete maternal imprinting, and are then subject to analysis using the imprinting (four QTL genotypes) and non-imprinting (three QTL genotypes) models. It can be seen that the non-imprinting model produces poorer estimation than the imprinting model. The RMSE is generally large when data are analyzed with the non-imprinting model, especially the mean parameters for the two reciprocal heterozygotes. We observed similar results under other imprinting mechanisms (e.g., partial or complete

paternal imprinting) and the results are omitted. When data are simulated assuming no imprinting, the non-imprinting model, however, outperforms the imprinting model, in which the standard errors of the mean parameters fitted with the imprinting model are slightly higher than those fitted with the non-imprinting model (data not shown). Similar results were also obtained in our previous univariate imprinting analysis [22]. Therefore, caution is needed about the interpretation of the results. One should try both imprinting and non-imprinting models and report the union of QTLs that are shown in both analyses.

### A Case Study

We apply the developed model to a published data set [36] to show the utility of the approach. The data contain 502  $F_2$  mice derived from two inbred strains differing greatly in body weight, the Large (LG/J) and Small (SM/J). Each  $F_2$  progeny was measured for its body mass at 10 equally spaced weeks starting at day 7 after birth. Ninety-six codominant markers were obtained with an average length of  $\sim 23$  cM spanning the 19 autosomes. For more information about the data, readers are referred to the original paper [36].

The sex-specific recombination rate is reconstructed on the basis of the marker data according to the average 1.25:1 recombination rate between female and male chromosomes [37]. The data were analyzed previously with a univariate imprinting model [12], in which three iQTL and one Mendelian QTL were detected. Considering the limitation of the univariate analysis as introduced in the background section, we apply the data to the newly developed functional mapping model. As clearly shown by the genomewide LR profile plot in Fig 1, the model detects four major QTLs. One QTL passes the genomewide threshold and is located on chromosome 6. The other three QTLs, located on chromosomes 7, 10 and 16, respectively, are only significant at the chromosome-wide level. In the plot, the solid curve corresponds to the likelihood ratio statistics at every testing position. The 5% significance threshold value for claiming the existence of QTLs at the genome-wide level is marked with the horizontal solid line based on permutation tests. The dashed line represents the 5% chromosome-wide threshold, which is obtained by only scanning the targeted chromosome.

Table 5 tabulates the estimated QTL positions, the marker intervals of the QTL belonging to each, the MLEs of curve parameters that specify the developmental pattern, and the asymptotic standard errors of the estimators (in parentheses). It can be seen that all parameters can be reasonably estimated with small sampling errors. The four QTLs detected are located between markers Nds5 and Mit15 on chromosome 6, between markers Nds1 and Mit148 on chromosome 7, between markers Mit133 and Mit14 on chromosome 10, and between markers Mit2 and Mit5 on chromosome 16. The data are also analyzed with the regular functional mapping approach and four QTLs are detected, among which only two QTLs, located on chromosomes 6 and 7, agree in both models. The other two, located on chromosomes 11 and 15, do not show significance with the imprinting analysis. The miss-detection of these two QTLs by the imprinting model may partially be due to the limitation of the imprinting approach as revealed by the simulation study.

To dissect the imprinting property of the detected QTLs, we further conduct hypothesis tests based on Tests (11)–(13). Among the four QTLs detected, three show signs of imprinting, i.e., the ones located on chromosomes 6, 10 and 16. The one located on chromosome 7 shows no sign of imprinting. The developmental trajectories of the identified QTLs are plotted and are shown in Fig. 2 with the growth trajectories for all individuals indicated as gray in the background. The imprinting property of the four QTLs can also be inferred from the plot. For example, for the iQTL detected on chromosome 6, the solid and the dashed blue curves for QTL genotypes  $Q_M Q_P$  and  $q_M Q_P$  are almost merged together and the solid and dashed red curves for QTL genotypes  $Q_M q_P$  and  $q_M q_P$  are almost merged together. This implies that only the allele inherited from the paternal parent is expressed and the QTL is maternally imprinted. A similar result is obtained for this QTL with our previous univariate imprinting model [12]. A similar pattern can be seen for other two imprinted QTLs, while for the QTL detected on chromosome 7, four separate developmental curves can be clearly seen, which indicates that the QTL is not imprinted. The trajectory plots confirm the testing results based on Tests (12) and (13).

### Discussion

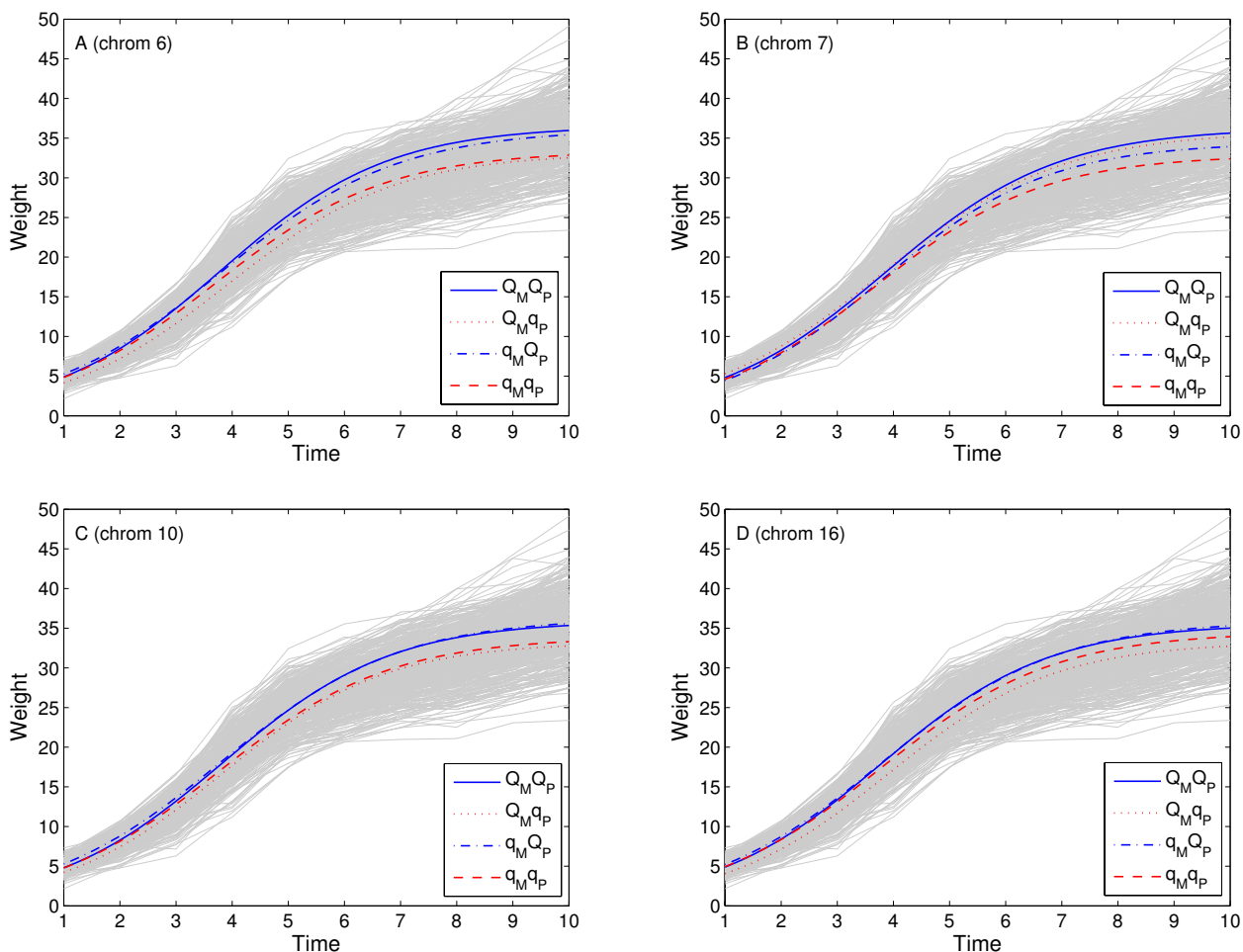
We all know that the diversity of offspring genes inherited from parents is due to the perturbation and reshuffling of parental genetic information during meiosis. Maintaining a functional expression balance between the paternal and maternal alleles is crucial for an organism's normal development. Breakage of the balance (e.g., the same allele expresses differently depending on its parental origin) often results in a phenomenon called genomic imprinting. In extreme cases, when one allele is completely silent, deletion of the functional allele may have serious consequences. For example, a deletion of the paternal copies of the imprinted SNRPN gene and neccin gene on chromosome 15 located in the region 15q11–13 may result in a genetic disorder called Prader-Willis syndrome [38].

Genomic imprinting has been ubiquitously observed in nature. The cause of imprinting has been thought to be related to DNA methylation in which certain parts of the DNA sequence are methylated and hence are silent in expression [39]. Although the functional mechanism of genomic imprinting is not totally clear, scientists thought that genomic imprinting has the one-generation dependence property. At the early stage of meiosis, the imprinting mark is erased and is subsequently reset at the end of meiosis, and imprinting is passed from one generation to another [40]. Offspring expression of an imprinted gene only depends on the allelic origin and is independent of the offspring sex. Therefore, any population in which the

**Table 5: The QTL location, MLEs of the estimated parameters and their asymptotic standard errors in the parentheses with the AR(1) covariance structure.**

CH*	QTL position	$Q_M Q_P$			$Q_M q_P$			$q_M Q_P$			$q_M q_P$			Residual	
		$\alpha_1$	$\beta_1$	$\gamma_1$	$\alpha_2$	$\beta_2$	$\gamma_2$	$\alpha_3$	$\beta_3$	$\gamma_3$	$\alpha_4$	$\beta_4$	$\gamma_4$	$\rho$	$\sigma^2$
6(M)	<i>Nds5-Mit15</i> ( <i>Nds5</i> +12 cM)	36.53 (0.362)	6.47 (0.104)	0.67 (0.007)	33.09 (0.408)	7.01 (0.136)	0.67 (0.008)	36.11 (0.392)	5.87 (0.097)	0.66 (0.007)	33.36 (0.352)	5.90 (0.106)	0.66 (0.008)	0.77 (0.011)	0.012 (0.0005)
7(NI)	<i>Nds1-Mit148</i> ( <i>Nds1</i> +7.5 cM)	36.28 (0.651)	6.53 (0.098)	0.65 (0.006)	35.88 (0.434)	5.79 (0.153)	0.63 (0.009)	34.44 (0.357)	6.72 (0.091)	0.68 (0.007)	32.86 (0.317)	6.13 (0.086)	0.67 (0.006)	0.81 (0.009)	0.013 (0.0004)
10(M)	<i>Mit133-Mit14</i> ( <i>Mit133</i> +6.7 cM)	35.93 (0.335)	6.49 (0.094)	0.66 (0.006)	33.24 (0.436)	6.90 (0.139)	0.69 (0.009)	36.32 (0.398)	5.87 (0.097)	0.63 (0.007)	33.87 (0.329)	6.08 (0.090)	0.65 (0.006)	0.79 (0.008)	0.012 (0.0005)
16(M)	<i>Mit2-Mit5</i> ( <i>Mit2</i> +20 cM)	35.58 (0.369)	6.28 (0.103)	0.67 (0.007)	33.27 (0.439)	7.18 (0.145)	0.68 (0.009)	35.97 (0.441)	5.90 (0.106)	0.64 (0.008)	34.52 (0.367)	5.98 (0.099)	0.65 (0.007)	0.78 (0.013)	0.011 (0.0007)

\*CH refers to chromosome, M refers to maternal imprinting, and NI refers to no imprinting.



**Figure 2**

**Growth trajectory plot.** Four curves for the dynamic changes of mouse growth trajectory, each representing one of the four groups of genotypes,  $Q_M Q_P$ ,  $Q_M q_P$ ,  $q_M Q_P$ , and  $q_M q_P$ , at each of the four significant QTLs. Mouse growth trajectories for all observed individuals are indicated in gray background.

allelic parental origin can be traced can be used for mapping purposes theoretically.

Experimental line crossing has been widely used in QTL mapping studies. Several studies have been reported based on inbred line crosses for iQTL mapping [11-13]. These approaches are developed for univariate QTL analysis. Considering the functional dynamics of a gene, a powerful approach for understanding the genetic architecture of a dynamic trait would be to incorporate the dynamic feature of gene function into a mapping model. Statistical functional mapping, as revealed by empirical studies [20,22], shows its unique merits in mapping QTL underlying developmental character or reaction norm. To understand the genetic architecture of a dynamic imprinted trait, we have extended the functional mapping approach to map iQTL responsible for a dynamic

imprinted trait. The model is a natural extension of our previous single trait imprinting model [12]. Simulation studies have shown the small sample properties of the approach under different simulation scenarios, considering the effect of sample size, heritability levels and different imprinting mechanisms.

As revealed by the real data analysis, the current functional iQTL mapping approach and our previous single trait iQTL mapping approach [12] agree on most significant QTLs, i.e., the ones from chromosomes 6,7 and 10. The imprinting property of these QTLs also agrees with the previous results, i.e., the QTLs on chromosome 6 and 10 are maternally imprinted and the QTL on chromosome 7 is not imprinted. However, the new model identifies one new iQTL located on chromosome 16. This iQTL is maternally imprinted. The one located on chromosome

15 detected by our previous single trait iQTL model does not show significance in the current analysis. Since the single trait analysis is based on the marginal distribution of the data at certain developmental stage, it may amplify the QTL effect by not adjusting the cumulative genetic effects up to time  $t$  [18]. Further experimental evidence is needed to confirm the results.

It is interesting to note that the detected iQTL for this data set are all maternally imprinted. The results, however, are not unusual, as explained by the parental genetic conflict theory [41]. Based on the theory, paternally derived alleles always trigger a favorable effect on the offspring's growth, whereas maternally derived alleles tend to trigger the opposite effect. As a result, imprinted genes that are preferentially expressed from the paternal alleles are enriched in genes that promote offspring growth. Therefore, for the growth trait in the current study, it is not surprising to see many maternally imprinted QTLs.

As a first attempt of its kind, we construct our functional iQTL mapping idea on a tractable one-QTL interval mapping framework. A one-QTL model does not consider the effects of background markers and is very limited to precisely elucidate the complex genetic architecture of a dynamic imprinted trait. The incorporation of ideas from more advanced mapping approaches such as composite interval mapping [42] and multiple interval mapping [43] can greatly enhance the utility of the developed method. More recently, Yang et al. [44] developed a composite functional mapping approach, which adopted a similar idea as the composite interval mapping [42] and shows improved features against the one-QTL functional mapping model. To make our work more useful in practice, modelling of multiple QTLs by composite or multiple interval mapping will be considered in future work.

The developed functional iQTL mapping is illustrated using an inbred  $F_2$  design. Since the allele parental origin for the two reciprocal heterozygotes is not distinguishable, we incorporate the sex specific recombination rate information to discern the difference in distribution of the two reciprocal heterozygotes. We assume an average 1.25:1 female-to-male recombination rate in mouse to illustrate our method. Additional simulations are conducted by varying the recombination ratio and similar results are obtained (data not shown). The sex specific recombination rates are commonly observed in nature. For example, averaged over the entire genome, the female-to-male recombination rate is 1.6:1 in human [45], 1.4:1 in dog [46], 1.4:1 in pig [47]. However, we also expect local variation in sex specific recombination rate. For example, Marklund et al. [47] reported that the chromosome specific female-to-male recombination rate varies from 2.2:1 (Pig chromosome 5) to 1:1 (Pig chromosome

1), an approximately two-fold difference in female-to-male recombination rates. This kind of local variation in sex specific recombination rate may affect the analysis results, and concomitantly the inference of genomic imprinting. The use of a sex-specific linkage map would greatly improve the inference of a local imprinting property of a QTL. With the availability of sex specific linkage information in experimental species, we expect the method be more robust and to provide more informative results. Meanwhile, owing to the limited information about the sex specific recombination rate in other species, this may limit the application of the developed approach to other populations. An alternative genetic design that can trace the allelic parental origin using experimental line crosses is the backcross design. Statistical approaches focusing on the backcross design have been developed for iQTL mapping in univariate trait analysis [11,13]. The idea can be extended to functional iQTL mapping with little modification.

Functional mapping and its extension like the one presented in this article provide a stimulating way to map complex biological processes by incorporating curve fitting into a mapping framework. Regardless of the limitations mentioned above, the integration of imprinting information into the functional mapping framework provides a testable quantitative platform for understanding the genetic basis of imprinted genes accounting for quantitative variation of a dynamic trait. The incorporation of more flexible mean function modelling approaches, such as non-parametric regression [48], would greatly enhance the flexibility of the current approach.

### Competing interests

The author(s) declare that they have no competing interests.

### Authors' contributions

YC conceived the idea, developed the model, performed the analysis, and drafted the manuscript. SL and GL assisted in the analysis and preparation of the manuscript.

### Acknowledgements

The authors wish to thank the executive editor Paul Agutter and the three anonymous referees for their valuable comments. The authors are also grateful to James Cheverud for providing the mouse data. This work was supported by research grant DMS-0707031 from the National Science Foundation, NIH grant RR015116 and an intramural research grant from Michigan State University.

### References

1. Barlow DP: **Gametic imprinting in mammals.** *Science* 1995, **270**:1610-1613.
2. Reik W, Walter J: **Genomic imprinting: parental influence on the genome.** *Nat Rev Genet* 2001, **2**:21-32.
3. Reik W, Walter J: **Evolution of imprinting mechanisms: the battle of the sexes begins in the zygote.** *Nat Genet* 2001, **27**:255-256.

4. Dilkes BP, Dante RA, Coelho C, Larkins BA: **Genetic analysis of endoreduplication in Zea mays endosperm: evidence of sporophytic and zygotic maternal control.** *Genetics* 2002, **160**:1163-1177.
5. Nezer C, Moreau L, Brouwers B, Coppieters W, Detilleux J, Hanset R, Karim L, Kvasz A, Leroy P, Georges M: **An imprinted QTL with major effect on muscle mass and fat deposition maps to the IGF2 locus in pigs.** *Nat Genet* 1999, **21**:155-156.
6. de Koning D-J, Rattink AP, Harlizius B, van Arendonk JAM, Brascamp EW, Groenen MA: **Genome-wide scan for body composition in pigs reveals important role of imprinting.** *Proc Natl Acad Sci USA* 2000, **97**:7947-7950.
7. Gimelbrant A, Hucchinson JN, Thompson R, Chess A: **Widespread monoallelic expression on human autosomes.** *Nature* 2007, **318**:1136-1140.
8. Isles AR, Holland AJ: **Imprinted genes and mother-offspring interactions.** *Early Hum Dev* 2005, **81**(1):73-77.
9. Tycko B, Morison IM: **Physiological functions of imprinted genes.** *J Cell Physiol* 2002, **192**(3):245-258.
10. Constanca M, Kelsey G, Reik W: **Resourceful imprinting.** *Nature* 2004, **432**(7013):53-57.
11. Cui YH: **A Statistical Framework for Genome-wide Scanning and Testing Imprinted Quantitative Trait Loci.** *J Theo Biol* 2007, **244**:115-126.
12. Cui YH, Lu Q, Cheverud JM, Littel RL, Wu RL: **Model for mapping imprinted quantitative trait loci in an inbred F<sub>2</sub> design.** *Genomics* 2006, **87**:543-551.
13. Cui YH, Cheverud JM, Wu RL: **A Statistical Model for Dissecting Genomic Imprinting through Genetic Mapping.** *Genetica* 2007, **130**:227-239.
14. Knott SA, Marklund L, Haley CS, Andersson K, Davies W, Ellegren H, Fredholm M, Hansson I, Hoyheim B, Lundstrom K, Moller M, Andersson L: **Multiple marker mapping of quantitative trait loci in a cross between outbred wild boar and Large White pigs.** *Genetics* 1998, **149**:1069-1080.
15. de Koning D-J, Bovenhuis H, van Arendonk JAM: **On the detection of imprinted quantitative trait loci in experimental crosses of outbred species.** *Genetics* 2002, **161**:931-938.
16. Bartolomei MS, Tilghman SM: **Genomic imprinting in mammals.** *Annu Rev Genet* 1997, **31**:493-525.
17. Wu RL, Ma C-X, Lin M, Casella G: **A general framework for analyzing the genetic architecture of developmental characteristics.** *Genetics* 2004, **166**:1541-1551.
18. Zhu J: **Analysis of conditional effects and variance components in developmental genetics.** *Genetics* 1995, **141**:1633-1639.
19. Ma C-X, Casella G, Wu RL: **Functional mapping of quantitative trait loci underlying the character process: A theoretical framework.** *Genetics* 2002, **161**:1751-1762.
20. Wu RL, Lin M: **Functional mapping – How to map and study the genetic architecture of dynamic complex traits.** *Nat Rev Genet* 2006, **7**:229-237.
21. Cui Y, Wu R, Casella G, Zhu J: **Non-parametric functional mapping of quantitative trait loci underlying programmed cell death.** *Stat Appl Genet Mol Biol* 2008 in press.
22. Cui Y, Zhu J, Wu R: **Functional Mapping for Genetic Control of Programmed Cell Death.** *Physiol Genomics* 2006, **25**:458-469.
23. Wu RL, Casella G, Ma C-X: *Statistical Genetics of Quantitative Traits: Linkage, Maps and Qtl* New York: Springer; 2007.
24. West GB, Brown JH, Enquist BJ: **A general model for ontogenetic growth.** *Nature* 2001, **413**:628-631.
25. Niklas KL: *Plant Allometry: The Scaling of Form and Process* Chicago: University of Chicago; 1994.
26. Diggle PJ, Heagerty P, Liang KY, Zeger SL: *Analysis of Longitudinal Data* Oxford, UK: Oxford University Press; 2002.
27. Box GEP, Cox DR: **An analysis of transformations.** *Journal of the Royal Statistical Society Series B* 1964, **26**:211-252.
28. Carroll RJ, Ruppert D: **Power-transformations when fitting theoretical models to data.** *J Am Stat Assoc* 1984, **79**:321-328.
29. Wu RL, Ma CX, Lin M, Wang ZH, Casella G: **Functional mapping of growth quantitative trait loci using a transform-both-sides logistic model.** *Biometrics* 2004, **60**:729-38.
30. Jaffrézic F, Thompson R, Hill WG: **Structured antedependence models for genetic analysis of repeated measures on multiple quantitative traits.** *Genet Res* 2003, **82**:55-65.
31. Zhao W, Chen YQ, Casella G, Cheverud JM, Wu RL: **A non-stationary model for functional mapping of complex traits.** *Bioinformatics* 2005, **21**:2469-2477.
32. Nelder JA, Mead R: **A simplex method for function minimization.** *Computer J* 1965, **7**:308-313.
33. Churchill GA, Doerge RW: **Empirical threshold values for quantitative trait mapping.** *Genetics* 1994, **138**:963-971.
34. Feinberg AP: *Genomic imprinting and cancer, in The Metabolic and Molecular Bases of Inherited Disease* Edited by: Scriver CR, Beaudet AL, Sly WS, Valle D. New York: McGraw-Hill; 2001:525-537.
35. Lynch M, Walsh B: **Genetics and Analysis of Quantitative Traits.** Sunderland, MA: Sinauer; 1998.
36. Vaughn TT, Pletscher LS, Peripato A, King-Ellison K, Adams E, Erikson C, Cheverud JM: **Mapping quantitative trait loci for murine growth: A closer look at genetic architecture.** *Genet Res* 1999, **74**:313-322.
37. Dietrich WF, Miller J, Steen R, Merchant MA, Damron-Boles D, Husain Z, Dredge R, Daly MJ, Ingalls KA, O'Connor TJ: **A comprehensive genetic map of the mouse genome.** *Nature* 1996, **380**:149-152.
38. Cassidy SB, Lai LW, Erickson RP, Magnuson L, Thomas E, Gendron R, Herrmann J: **Trisomy 15 with loss of the paternal 15 as a cause of Prader-Willi syndrome due to maternal disomy.** *Am J Hum Genet* 1992, **51**:701-708.
39. Gold JD, Pedersen RA: **Mechanisms of Genomic Imprinting in Mammals.** *Current Topics in Developmental Biology* 1994, **29**:227-247.
40. Walter J, Paulsen M: **Imprinting and disease.** *Seminars in Cell & Dev Biol* 2003, **14**:101-110.
41. Wilkins JF, Haig D: **What good is genomic imprinting: the function of parent specific gene expression.** *Nat Rev Genet* 2003, **4**:359-368.
42. Zeng Z-B: **Precision mapping of quantitative trait loci.** *Genetics* 1994, **136**:1457-1468.
43. Kao CH, Zeng Z-B, Teasdale RD: **Multiple interval mapping for quantitative trait loci.** *Genetics* 1999, **152**:1203-1216.
44. Yang R, Gao H, Wang X, Zhang J, Zeng Z-B, Wu RL: **A Semiparametric Approach for Composite Functional Mapping of Dynamic Quantitative Traits.** *Genetics* 2007, **177**:1859-1870.
45. Dib C, Faure S, Fizames C, Samson S, Drouot N, Vignal A, Millasseau P, Marc S, Hazan J, Seboun E, Lathrop M, Gyapay G, Morissette J, Weissbach J: **A comprehensive genetic map of the human genome based on 5,264 microsatellites.** *Nature* 1996, **380**:152-154.
46. Neff MW, Broman KW, Mellersh CS, Ray K, Acland GM, Aguirre GD, Ziegler JS, Ostrander EA, Rine J: **A second-generation genetic linkage map of the domestic dog, *Canis familiaris*.** *Genetics* 1999, **151**:803-820.
47. Marklund L, Moller MJ, Hoyheim B, Davies W, Fredholm M, Juneja RK, Mariani P, Coppieters W, Ellegren H, Andersson L: **A comprehensive linkage map of the pig based on a wild pig-Large White intercross.** *Anim Genet* 1996, **27**:255-269.
48. Boularan J, Ferre L, Vieu P: **Growth curves: a two-stage nonparametric approach.** *J Stat Plan Infer* 1994, **38**:327-350.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

