

# A statistical model for functional mapping of quantitative trait loci regulating drug response

Y Gong<sup>1,3</sup>  
Z Wang<sup>1,3</sup>  
T Liu<sup>1</sup>  
W Zhao<sup>1</sup>  
Y Zhu<sup>1</sup>  
JA Johnson<sup>2</sup>  
R Wu<sup>1</sup>

<sup>1</sup>Department of Statistics, University of Florida, Gainesville, FL, USA; <sup>2</sup>Department of Pharmacy Practice, University of Florida, Gainesville, FL, USA

**Correspondence:**

Dr R Wu, Department of Statistics, 533  
McCarty Hall C, University of Florida,  
Gainesville, FL 32611, USA.  
Tel.: +1 352 392 3806  
Fax: +1 352 392 8555  
E-mail: [rwu@stat.ufl.edu](mailto:rwu@stat.ufl.edu)

## ABSTRACT

Differential drug response, that is, pharmacodynamics, is most often likely to be a complex trait, controlled by the combined influences of multiple genes and environmental influences. Genetic mapping has proven to be a powerful tool for detecting and identifying specific genes affecting complex traits, that is, quantitative trait loci (QTL), based on polymorphic markers. In this article, we present a novel statistical model for genetic mapping of QTL governing pharmacodynamic processes. In principle, this model is a combination of *functional mapping* proposed to map function-valued traits and *linkage disequilibrium mapping* designed to provide high-resolution mapping of QTL by making use of recombination events created at a historic time. We implement a closed-form solution for the Expectation-Maximization algorithm to estimate the population genetic parameters of QTL and the simplex algorithm to estimate the curve parameters describing the pharmacodynamic changes of different QTL genotypes in response to drug dose or concentrations. Extensive simulations are performed to investigate the statistical properties of our model. The implications of our model in pharmacogenetic and pharmacogenomic research are discussed.

*The Pharmacogenomics Journal* (2004) 4, 315–321. doi:10.1038/sj.tpj.6500262  
Published online 20 July 2004

**Keywords:** drug response; functional mapping; linkage disequilibrium; quantitative trait locus

## INTRODUCTION

There is tremendous interindividual variation in pharmacological response to medications. Although such variability in drug effects may be attributed to the pathogenesis of the disease being treated, drug interactions, and the individual's age, nutritional status, renal or liver function, increasing evidence has been observed for influences of genetic differences in the metabolism and disposition of drugs and the targets of drug therapy (such as receptors) on the efficacy and toxicity of medications (reviewed in Nebert,<sup>1</sup> Evans and Relling<sup>2</sup> and Evans and Johnson<sup>3</sup>). As such, drug response is typically a complex trait, with multiple genes and various biochemical, developmental and environmental factors contributing differently to the overall phenotype.<sup>4</sup> To understand more fully the genetic basis of drug response, approaches are needed in which new specific genes or quantitative trait loci (QTL) can be identified.

The way individuals respond to varying drug dosages or concentrations presents a pharmacodynamic problem that can be regarded as function-valued traits in genetic mapping research. The genetic architecture of function-valued traits can be studied using the marker-based *functional mapping* model, developed

<sup>3</sup>These authors contributed equally to this work.

by Wu and colleagues.<sup>5–10</sup> Different from traditional methods for mapping a complex trait described by a single value, functional mapping has power to map dynamic QTL responsible for a biological process that need be measured at a finite number of time points. In modeling functional mapping, fundamental principles behind biological or biochemical networks described by mathematical functions are incorporated into a QTL mapping framework. Functional mapping estimates parameters determining shapes and functions of a particular biological network, rather than directly estimating gene effects at all possible points within the network. Because of the connection of these points through mathematical functions, functional mapping strikingly reduces the number of parameters to be estimated and, hence, displays increased statistical power.<sup>11</sup>

Functional mapping is also advantageous in terms of biological relevance because biological principles are embedded into the estimation process of QTL parameters. The results derived from functional mapping will be closer to biological realms. For example, using functional mapping, Wu *et al*<sup>5</sup> detected a QTL responsible for stem diameter growth trajectories in an experimental plantation of forest trees. This QTL was detected to trigger phenomenal effects only after tree canopies close in the stand. Such a dynamic pattern of QTL genetic effect is broadly in agreement with the ecological theory for asymmetric competition of growing trees. Lastly, but not least, functional mapping provides an organizing framework to understand the coherent behavior of a whole biological system from the integration and coordination of its parts.

In current pharmacogenetic research, increasing attempts have been made to identify candidate genes that influence pharmacological responses.<sup>1</sup> These include genes involved in drug transport (eg polymorphisms in the gene encoding P-glycoprotein 1 and the plasma concentration of digoxin), genes involved in drug metabolism (eg polymorphisms in the gene encoding thiopurine S-methyltransferase and thiopurine toxicity) and genes encoding drug targets (eg polymorphisms in the gene encoding the  $\beta$ -adrenoceptor and response to  $\beta$ -adrenoceptor agonists). With advanced molecular genotyping technologies, a number of polymorphic sites (such as single-nucleotide polymorphisms or SNPs) within or near these candidate genes can be genotyped. SNPs, especially SNPs that occur in gene regulatory or coding regions (cSNPs), can be associated with phenotypic traits to detect genetic variants causing pharmacological response variability. Linkage disequilibrium mapping based on the nonrandom association between different genes in a population has proven to be a powerful means for high-resolution mapping of genes for complex traits.<sup>11</sup> More recently, we derived a closed-form solution based on the EM algorithm for estimating the allele frequencies of functional genetic variants and their disequilibria with SNPs.<sup>12</sup> It is possible that this algorithm can be used to map QTL affecting the extent to which an individual responds to a particular pharmacological action.

The motivation of this study is to extend the idea of functional mapping to map QTL governing pharmacody-

namic (PD) processes by incorporating our linkage disequilibrium analysis algorithm. In the next sections, we first introduce basic models for PD models and then formulate a likelihood function for the functional mapping of dynamic QTL that determines pharmacodynamic processes. These will be followed by the investigation of the statistical properties of this model through extensive simulation studies.

## THE PD MODEL

The mainstay of modeling PD is the Hill, or sigmoid  $E_{\max}$ , equation, which postulates the following relationship between drug concentration ( $C$ ) and drug effect ( $E$ )<sup>13</sup>

$$E = E_0 + \frac{E_{\max} C^H}{EC_{50}^H + C^H} \quad (1)$$

where  $E_0$  is the baseline value for the drug response parameter,  $E_{\max}$  is the asymptotic (limiting) effect,  $EC_{50}$  is the drug concentration that results in 50% of the maximal effect, and  $H$  is the slope parameter that determines the slope of the concentration–response curve. The larger  $H$ , the steeper the linear phase of the log–concentration–effect curve (equation (1)). When the effect is a continuous variable, estimates of  $E_{\max}$ ,  $EC_{50}$  and  $H$  are usually obtained by extended least squares or iteratively reweighted least squares when there is sufficient data for analysis of individual subjects. When sparse data are pooled from multiple patients, then a population analysis is a better approach. Different from such a traditional treatment, we will estimate these curve parameters separately for different genotypes at a latent QTL.

There is a standard clinical procedure to measure effect–dose relationships. Patients receive escalating doses of a drug at a particular time interval. The goal of the test is to increase the dose until a person reaches his or her target response or predetermined maximum dose. During this test, the person's heart rate and blood pressure are continually monitored.

## THE STATISTICAL MODEL

Suppose there is a random sample of size  $n$  drawn from a natural human population at Hardy–Weinberg equilibrium. In this sample, multiple SNP markers are genotyped, aimed at the identification of QTL affecting the PD process. For a complete trial, drug effects are measured at six hallmark dose–concentration levels, so we have a finite set of data on each individual  $i$ , which can be regarded as a multivariate trait vector,  $y_i(1), \dots, y_i(6)$ . This finite set of data can be modeled by the  $E_{\max}$  model (equation (1)).

Assume that a pleiotropic QTL,  $A$  (of alleles  $A$  and  $a$ ) affecting the PD process is segregating in the human population. The allele frequencies of  $A$  and  $a$  are expressed as  $q$  and  $1-q$ , respectively. For a particular genotype  $j$  of this QTL ( $j=0$  for  $aa$ , 1 for  $Aa$  and 2 for  $AA$ ), the parameters describing its PD process are denoted by  $\Theta_j = (E_{0j}, E_{\max j}, EC_{50j}, H_j)$  for the  $E_{\max}$  model. The comparisons of these parameters between the three different genotypes can

determine whether and how this putative QTL affects the PD process.

The trait phenotype of individual  $i$  for drug effects measured at concentration  $C$  due to the QTL can be expressed by a linear statistical model,<sup>14</sup>

$$y_i(C) = \sum_{j=0}^2 [\xi_{ij} g_j(C)] + e_i(C) \quad (2)$$

where  $\xi_{ij}$  is an indicator variable for the possible genotypes of the QTL for individual  $i$  and defined as 1 if a particular QTL genotype  $j$  is indicated and 0 otherwise,  $g_j(C)$  is the genotypic value of the QTL for the trait at concentration  $C$ , which can be fit using the PD model expressed in equation (1), and  $e_i(C)$  is the residual effect of individual  $i$ , including the aggregate effect of polygenes and error effect, and distributed as  $N(0, \sigma_e^2(C))$ .

To attempt to make the errors in model (2) homoscedastic and normal, a transformation approach of the effect phenotypes  $y$  can be used. As an example, we consider a log transformation in which we transform both sides of equation (2) to obtain

$$\log[y_i(C)] = \sum_{j=0}^2 \{ \xi_{ij} \log[g_j(C)] \} + \varepsilon_i(C),$$

or equivalently, defining  $z_i(C) = \log[y_i(C)]$ ,

$$z_i(C) = \sum_{j=0}^2 [\xi_{ij} h_j(C)] + \varepsilon_i(C),$$

where

$$h_j(C) = \log[g_j(C)] \quad (3)$$

and  $\varepsilon_i(C)$  is the residual error for the log-transformed data, distributed as  $N(0, \sigma_e^2(C))$ .

By transforming both sides of equation (2), we maintain the functional relationship between dose and response. Carroll and Ruppert<sup>16</sup> investigated a similar approach, although they allow the transformation used to be estimated by the data. In contrast, we suggest the data analyst consider a number of transformations until one is found that appropriately accounts for the particular features of the data being analyzed.

The likelihood of the human samples receiving six dose levels from  $0 \rightarrow 5 \rightarrow 10 \rightarrow 20 \rightarrow 30 \rightarrow 40$ ,  $\mathbf{z}_i = [z_i(1), \dots, z_i(6)]$ , for individual  $i$ , can be represented by a multivariate mixture model

$$L(\mathbf{z}) = \prod_{i=1}^n \left[ \sum_{j=0}^2 \pi_{ij} f_j(\mathbf{z}_i) \right] \quad (4)$$

where  $\pi_{ij}$  is the frequency of individual  $i$  to carry a particular QTL genotype  $j$ , and

$$f_j(\mathbf{z}_i) = \frac{1}{(2\pi)^3 |\Sigma_e|^{1/2}} \exp \left[ -\frac{1}{2} (\mathbf{z}_i - \mathbf{h}_j)^T \Sigma_e^{-1} (\mathbf{z}_i - \mathbf{h}_j) \right]$$

with  $\mathbf{h}_j = [h_j(1), \dots, h_j(6)]^T$  being the vector of the genotypic values of the log-transformed effects for QTL genotype  $j$  measured at six different concentrations, and  $\Sigma_e$  is the residual variance-covariance matrix of the log-transformed

phenotypes measured at these concentrations. Analogous to the repeated measurement problem,  $\Sigma_e$  can be fitted by the AR(1) model,

$$\Sigma_e = \sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \rho^4 & \rho^6 & \rho^8 \\ \rho & 1 & \rho & \rho^3 & \rho^5 & \rho^7 \\ \rho^2 & \rho & 1 & \rho^2 & \rho^4 & \rho^6 \\ \rho^4 & \rho^3 & \rho^2 & 1 & \rho^2 & \rho^4 \\ \rho^6 & \rho^5 & \rho^4 & \rho^2 & 1 & \rho^2 \\ \rho^8 & \rho^7 & \rho^6 & \rho^4 & \rho^2 & 1 \end{bmatrix} \quad (5)$$

which states that the residual variance ( $\sigma^2$ ) is constant over different concentrations, and that the residual correlation of response between different concentrations decreases proportionally (in  $\rho$ ) with increased concentration interval.

Suppose that this QTL is genetically associated with a codominant SNP marker,  $M$ , with three genotypes  $MM$ ,  $Mm$  and  $mm$ . Let  $p$  and  $1-p$  be the allele frequencies of alternative marker alleles  $M$  and  $m$ , respectively, and  $D$  be the coefficient of (gametic) linkage disequilibrium between the marker and QTL. According to the linkage disequilibrium-based mapping theory,<sup>15</sup> the detection of significant linkage disequilibrium between the marker and QTL implies that the QTL may be linked with and, therefore, can be genetically manipulated by the marker. The four haplotypes for the marker and QTL are  $MA$ ,  $Ma$ ,  $mA$  and  $ma$ , with respective frequencies expressed as  $p_{11} = pq + D$ ,  $p_{10} = p(1-q) - D$ ,  $p_{01} = (1-p)q - D$  and  $p_{00} = (1-p)(1-q) + D$ . Thus, the population genetic parameters  $p$ ,  $q$ ,  $D$  can be estimated by solving a group of regular equations if we can estimate the four haplotype frequencies.

The four haplotypes randomly unite to generate 16 cells. Some of these cells are collapsed into nine distinguishable genotypes whose frequencies are tabulated as:

	AA	A a + a A	aa	Total
MM	$p_{11}^2(n_1)$	$2p_{11}p_{10}(n_2)$	$p_{10}^2(n_3)$	$m_1 = n_1 + n_2 + n_3$
Mm	$2p_{11}p_{01}(n_4)$	$2p_{11}p_{00}(n_{51}) + 2p_{10}p_{01}(n_{52})$	$2p_{10}p_{00}(n_6)$	$m_2 = n_4 + n_5 + n_6$
mm	$p_{01}^2(n_7)$	$2p_{01}p_{00}(n_8)$	$p_{00}^2(n_9)$	$m_3 = n_7 + n_8 + n_9$

Of these genotypes, double heterozygote  $MmAa$  includes two different diplotypes (a set of haplotype pairs) each with a different frequency of formation. If we can observe nine genotypes ( $n_1$ – $n_9$ ) and two diplotypes for  $MmAa$  ( $n_{51}$  and  $n_{52}$ , summing to  $n_5$ ), haplotype frequencies can be estimated from these observations based on explicit estimators.

However, what we can observe in practice is the three marker genotypes with the number of  $m_1$  to  $m_3$ . Each of these marker genotypes is a mixture of the three latent QTL genotypes. Thus, the complete data set permitting explicit estimators of haplotype frequencies include the observable marker data and missing marker-QTL configurations. Here, we implement a closed-form EM algorithm for the estimation of the haplotype frequencies.<sup>12</sup> Again, let  $\pi_{ij}$  denote the probability of individual  $i$  to carry QTL genotype  $j$ . This

**Table 1** Maximum likelihood estimates of the parameters describing the three PD curves, each corresponding to a QTL, and marker allele frequency, QTL allele frequency and marker-QTL linkage disequilibrium

Parameter	Genotype	True value	Heritability ( $H^2$ )	
			0.1	0.4
$E_0$	AA	72	73.32 (2.1517)	72.10 (0.2876)
	Aa	70	72.09 (2.4300)	70.07 (0.2081)
	aa	60	62.04 (2.6165)	60.10 (0.2783)
$E_{\max}$	AA	79	76.61 (2.9550)	78.89 (0.2945)
	Aa	90	88.04 (4.1818)	89.77 (0.6288)
	aa	70	72.39 (9.7975)	71.19 (0.5890)
$EC_{50}$	AA	5	5.24 (0.2641)	4.79 (0.2101)
	Aa	20	19.66 (0.9228)	19.17 (0.8533)
	aa	30	35.08 (6.2243)	29.90 (0.6116)
$H$	AA	2	2.03 (0.0508)	2.01 (0.0120)
	Aa	1	1.07 (0.0857)	1.00 (0.0076)
	aa	0.5	0.51 (0.0582)	0.48 (0.0314)
$\rho$ $\sigma^2$		0.6	0.62 (0.0225)	0.60 (0.0212)
		60.12	52.69 (8.8831)	
		1.20		1.29 (0.1342)
$p$		0.6	0.6017 (0.0172)	0.6003 (0.0171)
$q$		0.6	0.5982 (0.0198)	0.5953 (0.0203)
$D$		0.08	0.0793 (0.0109)	0.0803 (0.0109)

The numbers in parentheses are the squared roots of the mean squared errors of the estimates.

(prior) probability takes the values of the above table depending on the (known) marker genotype of individual  $i$  to carry. In the E-step, the posterior probability that individual  $i$  have QTL genotype  $j$ ,

$$\Pi_{ij} = \frac{\pi_{ij}f_j(Z_i)}{\sum_{j=0}^2 \pi_{ij}f_j(Z_i)} \quad (6)$$

is calculated. In the M-step, the calculated posterior probability is used to solve the haplotype frequencies<sup>11</sup> expressed as

$$\begin{aligned} \hat{p}_{11} &= \frac{1}{2n} \left[ \sum_{i=1}^{m_1} (2\Pi_{i2} + \Pi_{i1}) + \sum_{i=1}^{m_2} (\Pi_{i2} + \psi\Pi_{i1}) \right] \\ \hat{p}_{10} &= \frac{1}{2n} \left\{ \sum_{i=1}^{m_1} (\Pi_{i1} + 2\Pi_{i0}) + \sum_{i=1}^{m_2} [\Pi_{i0} + (1 - \psi)\Pi_{i1}] \right\} \\ \hat{p}_{01} &= \frac{1}{2n} \left\{ \sum_{i=1}^{m_3} (2\Pi_{i2} + \Pi_{i1}) + \sum_{i=1}^{m_2} [\Pi_{i2} + (1 - \psi)\Pi_{i1}] \right\} \\ \hat{p}_{00} &= \frac{1}{2n} \left[ \sum_{i=1}^{m_3} (\Pi_{i1} + 2\Pi_{i0}) + \sum_{i=1}^{m_2} (\Pi_{i0} + \psi\Pi_{i1}) \right] \end{aligned} \quad (7-10)$$

where  $\psi = p_{11}p_{00} = (p_{11}p_{00} + p_{10}p_{01})$ .

To iterate between the E- and M steps (equations (7)–(10)), we need to estimate the parameters for response curves of different QTL genotypes  $\Theta_j$  and model parameters ( $\rho$ ,  $\sigma^2$ ). As shown in Ma *et al*<sup>10</sup> and Wu *et al*,<sup>7,9</sup> these parameters can be estimated with the EM algorithm.<sup>16</sup> But, this will encounter a considerable difficulty in equation derivations. Zhao *et al*<sup>17</sup> implemented the simplex method as advocated by Nelder and Mead<sup>18</sup> to the estimation process of functional mapping, which can strikingly increase computational efficiency. In this article, the simplex algorithm is embedded in the EM algorithm above to provide simultaneous estimation of haplotype frequencies and curve parameters based on the TBS-based model.

## HYPOTHESIS TESTS

Different from traditional mapping approaches, our functional mapping for function-valued traits allows for the tests of a number of biologically or clinically meaningful hypotheses. These hypothesis tests can be a test for the existence of significant QTL, or a test for the genetic effect on the maximal (asymptotic) effect ( $E_{\max}$ ), the drug concentration that results 50% of the maximal effect, and the slope that determines the steepness of the concentration–response curve.

Testing whether specific QTL exist to affect the shape of the  $E_{\max}$  model is a first step toward the understanding of the genetic architecture of the PD process. The genetic control over the entire PD process can be tested by formulating the following hypotheses:

$$\begin{cases} H_0: E_{02} = E_{01} = E_{00}, E_{\max 2} = E_{\max 1} = E_{\max 0}, \\ EC_{50(2)} = EC_{50(1)} = EC_{50(0)}, H_2 = H_1 = H_0 \\ H_1: \text{Not all these equalities above hold} \end{cases} \quad (11)$$

The  $H_0$  states that there are no QTL affecting the PD process (the reduced model), whereas the  $H_1$  proposes that such QTL do exist (the full model). The test statistic for testing the hypotheses (11) is calculated as the log-likelihood ratio of the reduced to the full model:

$$LR = -2 [\ln L(\tilde{\Omega}|\mathbf{z}) - \ln L(\hat{\Omega}|\mathbf{z})] \quad (12)$$

where  $\tilde{\Omega}$  and  $\hat{\Omega}$  denote the MLEs of the unknown parameters including haplotype under  $H_0$  and  $H_1$ , respectively. The LR is asymptotically  $\chi^2$ -distributed with 10 degrees of freedom. An empirical approach for determining the critical threshold is based on permutation tests, as advocated by Churchill and Doerge.<sup>19</sup> By repeatedly shuffling the relationships between marker genotypes and phenotypes, a series of the maximum log-likelihood ratios are calculated, from the distribution of which the critical threshold is determined.

We can also test for the significance of the genetic effect of the QTL on the response at a particular concentration level ( $C^*$ ) of interest, expressed as

$$\begin{cases} H_0: E_2(C^*) = E_1(C^*) = E_0(C^*) \\ H_1: \text{At least one of the equalities above does not hold} \end{cases} \quad (13)$$

which is equivalent to testing the difference of the full model with no restriction and the reduced model with a

restriction  $E_2(C^*) = E_1(C^*) = E_0(C^*)$ . Similar restrictions can be taken to test the genetic effect of the QTL on individual curve parameters, such as  $E_{\max}$ ,  $EC_{50}$  and  $H$ . The tests of these parameters are important for the design of personalized drugs to control particular diseases.

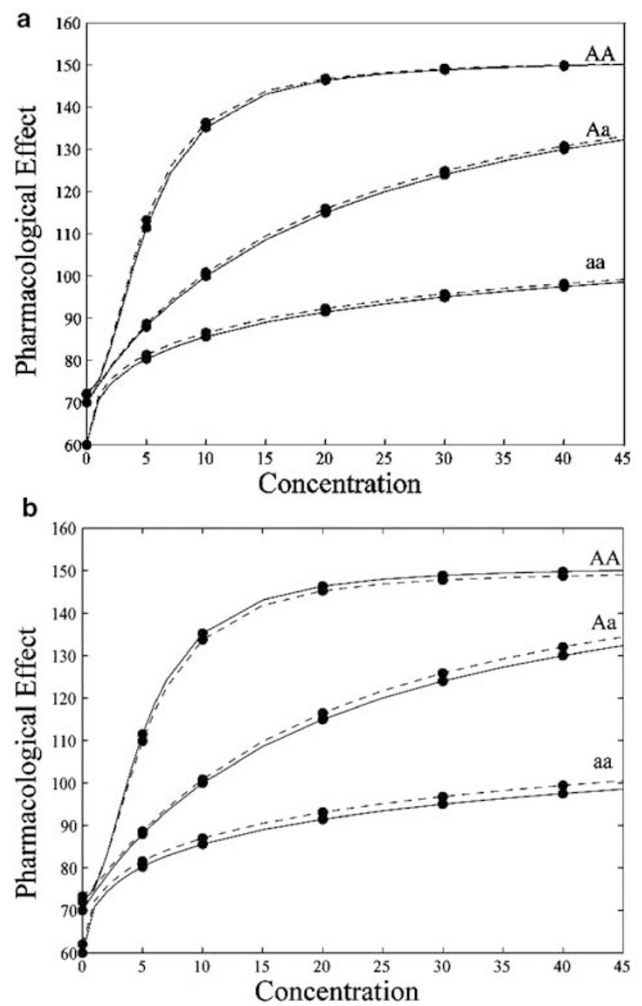
## RESULTS

We performed Monte Carlo simulation experiments to examine the statistical properties of the model proposed for genetic mapping of the PD process. We randomly choose 400 individuals from a human population at Hardy-Weinberg equilibrium. Let us first suppose one marker of two alleles  $M$  and  $m$ . This marker is used to infer a QTL of two alleles  $A$  and  $a$  for the PD process based on the nonrandom association between the marker and QTL. The allele frequencies are assumed as  $P=0.6$  for allele  $M$  and  $q=0.6$  for allele  $A$ . A positive value of linkage disequilibrium ( $D=0.08$ ) between alleles  $M$  and  $A$  is assumed, suggesting that these two more common alleles are in coupling phase.<sup>20</sup>

The three QTL genotypes,  $AA$ ,  $Aa$  and  $aa$ , are each hypothesized to have a different  $E_{\max}$  curve described by equation (1). The curve parameters ( $E_0$ ,  $E_{\max}$ ,  $EC_{50}$ ,  $H$ ) for the three genotypes, given in Table 1, are determined in the ranges of empirical estimates of these parameters from a pharmacological study.<sup>21</sup> We use a standard design, with drug effects measured at six different concentrations (Figure 1), for this simulation study. Using the genetic variance due to the QTL for the response at the last measurement point, we calculated the residual variances under different heritability levels ( $H^2=0.1$  and  $0.4$ ). These residual variances, plus a given residual correlation ( $\rho=0.7$ ), form a residual (co)-variance matrix according to equation (5). The phenotypic values of drug effect for 400 random patients are simulated by the summations of genotypic values predicted by the curves and residual errors following multivariate normal distributions, with  $MV N(0, \Sigma_e)$ .

Figure 1 illustrates different forms of the  $E_{\max}$  curves simulated from three hypothesized QTL genotypes and estimated from the simulated data using our model proposed above. The results suggest that the QTL responsible for drug response can be detected using the marker in association with the QTL. The parameters for the  $E_{\max}$  model of each QTL genotype can be estimated accurately (Table 1), with the estimated curves being broadly consistent with the hypothesized curves. The accuracy of curve estimates is better under higher (Figures 1b) than lower heritability (Figure 1a).

The estimates of the four parameters ( $E_0$ ,  $E_{\max}$ ,  $EC_{50}$ ,  $H$ ) for each PD curve also display reasonable precision, as assessed by the squared roots of mean square errors over 100 repeated simulations. As expected, the estimation precision increases remarkably when the heritability ( $H^2$ ) increases from 0.1 to 0.4. The population genetic parameters of the QTL can be estimated with reasonably high precision using our closed-form solution approach. We compare the estimation of marker allele frequencies, QTL allele frequencies and



**Figure 1** Estimated response curves (broken) for each of the three QTL genotypes,  $AA$ ,  $Aa$  and  $aa$ , in a comparison with the hypothesized curves (solid) used to simulate individual curves. Six different concentrations at which responses are measured are indicated with dots. The consistency between the estimated and hypothesized curves at two different heritability levels  $H^2=0.1$ (a) and  $0.4$ (b) suggests that our model can provide the precise estimation of the genetic control over response curves in human patients.

marker-QTL linkage disequilibrium under different heritability levels. The precision of these parameters is not much affected by differences in heritability (Table 1).

In each of 100 simulations, we calculate the log-likelihood ratios (LR) for the hypothesis test of the presence of a QTL affecting the entire PD process. The LR values average 1667 for  $H^2=0.1$  and 8380 for  $H^2=0.4$ , strikingly greater than the critical threshold estimated from 100 replicates of simulations under the null hypothesis that there is no QTL. This suggests that our model has 100% power to detect the QTL under given curve and population genetic parameters for our simulation.

Given the differences among these three response–dosage curves, our hypothesis tests are extended to test the effects of the curve QTL on two critical parameters,  $EC_{50}$  and  $H$ . The test results suggest that this QTL would significantly determine the difference in half-concentration and the slope of the curve even when the heritability is at a lower level (0.1).

## DISCUSSION

There has been increased evidence that genetic polymorphisms in drug-metabolizing enzymes, transporters, receptors, and other drug targets are associated with interindividual variation in the efficacy and toxicity of many medications.<sup>2,4,22</sup> The inherited nature of such variation in drug disposition and effects must be clearly elucidated to provide theoretical principles for optimizing drug therapy on the basis of individual patients' genetic constitutions (reviewed in Evans and Johnson<sup>3</sup>). A greater understanding of the genetic determinants of drug response has the potential to revolutionize the use of many medications.

In this article, we develop a novel statistical model for mapping quantitative trait loci (QTL) determining the discrepancies in the extent and pattern of different persons to respond to drugs. By increasing our ability to prospectively identify patients at risk for severe toxicity, or those likely to benefit from a particular treatment, our model promises to help us move towards the ultimate goal of individualized therapy. From statistical, clinical, and genetic standpoints, our model displays three unique advantages compared to traditional approaches for QTL mapping of single traits. First, we make simultaneous use of drug effects measured at all possible concentrations, so that the power to detect genetic determinants of drug response and the precision of genetic parameter estimation can be increased. Second, our model incorporates the commonly used  $E_{max}$  model<sup>13</sup> for characterizing drug–response curves into the QTL mapping framework and permit the tests of a number of clinically meaningful variables. These variables include the concentration at which persons exhibit differences in drug response, the slope of response curve and the concentration corresponding to half the maximal effect.

Third, our model is robust and flexible to different genetic settings. The results from a simulation study indicate that the QTL can be well detected when it accounts for a modest proportion of the observed variation in drug response. For human natural populations, as assumed in this article, the model is implemented with high-resolution linkage disequilibrium mapping of QTL.<sup>11</sup> However, for mice used as a model system for pharmacogenetic discovery,<sup>4</sup> this model allows for genome-wide scanning for response QTL based on linkage analysis. Given the inheritance properties of multiple genes involved in drug response, the model can be extended to detect epistatic QTL and QTL interacting with environmental and developmental signals (see Wu et al<sup>8</sup>). Currently, increased efforts are underway to construct comprehensive high-density maps of the human and mouse genome using high-throughput, single-nucleotide poly-

morphism-based genotyping technologies.<sup>23</sup> With the discovery of an increased number of candidate genes for drug response, our model will create a unique opportunity to contribute to unraveling the genetic mechanisms of drug response.

In this article, we assume that all subjects can be measured at the same ranges of dose concentrations. However, this may not be true in practical pharmacological trials because some persons reach their thresholds more quickly than others, leading to different response curves. The choice of a qualitatively different curve by individual persons is likely to be under control of specific QTL. Such QTL need to be modeled in our model, and their role in regulating individualized drug response to be identified. Wilson et al<sup>22</sup> detected considerable population genetic differentiation in metabolizing enzyme loci. There are interethnic differences in metabolizing enzyme allele frequencies. The knowledge about ethnicity (or geographic origin) need be integrated into our functional mapping model in order to shed better light on the genetic control mechanisms of response to a particular drug or group of drugs.

## REFERENCES

- 1 Nebert DW. Polymorphisms in drug-metabolizing enzymes: what is their clinical relevance and why do they exist? *Am J Hum Genet* 1997; **60**: 265–271.
- 2 Evans WE, Relling MV. Pharmacogenomics: translating functional genomics into rational therapeutics. *Science* 1999; **286**: 487–491.
- 3 Evans WE, Johnson JA. Pharmacogenomics: the inherited basis for interindividual differences in drug response. *Annu Rev Genomics Hum Genet* 2001; **2**: 9–39.
- 4 Watters JW, McLeod HL. Using genome-wide mapping in the mouse to identify genes that influence drug response. *Trends Pharmacol Sci* 2003; **24**: 55–58.
- 5 Wu RL, Ma C-X, Chang M, Littell RC, Wu SS, Yin TM et al. A logistic mixture model for characterizing genetic determinants causing differentiation in growth trajectories. *Genet Res* 2002; **19**: 235–245.
- 6 Wu RL, Ma C-X, Yang MCK, Chang M, Santra U, Wu SS et al. Quantitative trait loci for growth in *Populus*. *Genet Res* 2003; **81**: 51–64.
- 7 Wu RL, Ma C-X, Zhao W, Casella G. Functional mapping of quantitative trait loci underlying growth rates: a parametric model. *Physiol Genomics* 2003; **14**: 241–249.
- 8 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.
- 9 Wu RL, Ma C-X, Lin M, Wang Z, Casella G. Functional mapping of quantitative trait loci underlying growth trajectories using a transform-both-sides logistic model. *Biometrics* 2004; **60**: 729–738.
- 10 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.
- 11 Wu RL, Casella G. *Statistical Genomics of Complex Traits: A Quantitative Trait Loci Perspective*. Springer: New York 2005; (in press).
- 12 Lou X-Y, Casella G, Littell RC, Yang MKC, Wu RL. A haplotype-based algorithm for multilocus linkage disequilibrium mapping of quantitative trait loci with epistasis in natural populations. *Genetics* 2003; **163**: 1533–1548.
- 13 Giraldo J. Empirical models and Hill coefficients. *Trends Pharmacol Sci* 2003; **24**: 63–65.
- 14 Lander ES, Botstein D. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 1989; **121**: 185–199.
- 15 Wu RL, Ma C-X, Casella G. Joint linkage and linkage disequilibrium mapping of quantitative trait loci in natural populations. *Genetics* 2002; **160**: 779–792.

- 16 Carroll RJ, Ruppert D. Power-transformations when fitting theoretical models to data. *J Am Stat Assoc* 1984; **79**: 321–328.
- 17 Zhao W, Wu RL, Ma C-X, Casella G. A fast algorithm for functional mapping of complex traits. *Genetics* 2004; (accepted).
- 18 Nelder JA, Mead R. A simplex method for function minimization. *Computer J* 1965; **7**: 308–313.
- 19 Churchill GA, Doerge RW. Empirical threshold values for quantitative trait mapping. *Genetics* 1994; **138**: 963–971.
- 20 Lynch M, Walsh B. *Genetics and Analysis of Quantitative Traits*. Sinauer: Sunderland, MA 1998.
- 21 Sowinski KM, Burlew BS, Johnson JA. Racial differences in sensitivity to the negative chronotropic effects of propranolol in healthy men. *Clin Pharm Therap* 1995; **57**: 678–683.
- 22 Wilson JF, Weale ME, Smith AC, Gratrix F, Fletcher B, Thomas MG *et al*. Population genetic structure of variable drug response. *Nat Genet* 2001; **29**: 265–269.
- 23 Kruglyak L. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nat Genet* 1999; **22**: 139–144.