**Modeling sequence–sequence interactions for drug response**

Min Lin¹,⁴, Hongying Li¹, Wei Hou², Julie A. Johnson³ and Rongling Wu¹,*

¹Department of Statistics, ²Department of Epidemiology and Health Policy Research, ³Department of Pharmacy Practice, University of Florida, Gainesville, FL 32611 and ⁴Department of Biostatistics and Bioinformatics, Duke University, Durham, NC 27715, USA

Received on April 20, 2006; revised on March 1, 2007; accepted on March 14, 2007

Advance Access publication March 28, 2007

Associate Editor: Martin Bishop

**ABSTRACT**

**Motivation:** Genetic interactions or epistasis may play an important role in the genetic etiology of drug response. With the availability of large-scale, high-density single nucleotide polymorphism markers, a great challenge is how to associate haplotype structures and complex drug response through its underlying pharmacodynamic mechanisms.

**Results:** We have derived a general statistical model for detecting an interactive network of DNA sequence variants that encode pharmacodynamic processes based on the haplotype map constructed by single nucleotide polymorphisms. The model was validated by a pharmacogenetic study for two predominant beta-adrenergic receptor (βAR) subtypes expressed in the heart, β1AR and β2AR. Haplotypes from these two receptors trigger significant interaction effects on the response of heart rate to different dose levels of dobutamine. This model will have implications for pharmacogenetic and pharmacogenomic research and drug discovery.

**Availability:** A computer program written in Matlab can be downloaded from the webpage of statistical genetics group at the University of Florida.

**Contact:** rwu@mail.ifas.ufl.edu

**Supplementary information:** Supplementary data are available at Bioinformatics online.

**1 INTRODUCTION**

The increasing number of genetic studies for complex biological characteristics in humans requires more advanced techniques of statistical analysis. This is due to two reasons. First, genes interact in a complex network to determine the final phenotype of a trait. Second, the phenotypic formation of every trait can be partitioned into its multiple continuous developmental steps on a time scale. Any statistical approaches that attempt to draw a picture of the genetic architecture of a complex phenotype should simultaneously consider the interactions among different genes and the dynamic changes of these interactions during the process of phenotypic formation.

It has been well recognized that a complex trait is controlled by a number of Mendelian-inherited, environmentally sensitive genes, of which some act additively, whereas many others are operational in a multiplicative or compensatory way (Frankel and Schork, 1996; Moore, 2003). Many current statistical techniques used in genetic research assume the additive control of genes (Lynch and Walsh, 1998), aimed to facilitate data analysis and modeling, which may provide misleading results when genetic interactions or epistasis actually occur. The mapping strategy, pioneered by Lander and Botstein (1989), has proven to be powerful for detecting individual genes, known as quantitative trait loci (QTL), that affect a complex trait based on the cosegregation of genotyped markers and the underlying QTL that are bracketed by the markers. QTL identified from this strategy presents a hypothesized chromosomal segment whose DNA sequence is unknown. More recently, Liu et al. (2004) have developed a new strategy that can identify specific DNA sequence variants for a complex trait. This strategy that relies on the recent advent of high-throughput single nucleotide polymorphism (SNP) technologies allows for the genome-wide scan of causal DNA sequences, called quantitative trait nucleotides or QTNs. Lin and Wu (2006) have developed a conceptual framework for detecting interaction effects between different QTNs that contribute to quantitative variation.

Considering the dynamic feature of many complex traits, a series of new statistical models, called functional mapping, have been recently developed to characterize QTL that are responsible for genetic variation in these dynamic traits (Ma et al., 2002; Wu et al., 2004a, b, c). Functional mapping capitalizes on the mathematical functions that describe biological processes and embeds them within the context of genetic mapping theory. By estimating the mathematical parameters that determine the patterns of longitudinal trajectories, functional mapping has proven efficient and effective for unveiling the genetic architecture of complex traits (Wu et al., 2003; Wu and Lin, 2006).

A different but statistically similar example of these so-called time series or longitudinal traits is drug response, a field that has gained much attention due to possible clinical applications, ranging from individualized therapy to new drug development (Arranz et al., 2003). Recent studies have suggested that a variable number of polymorphisms in various genes are involved in modulating the response and/or side effects to drugs (Serretti and Artioli, 2003; Tate et al., 2005). Lin et al. (2005) extended the idea of functional mapping to map and detect individually acting genetic variants for drug response by integrating the underlying pharmacodynamic mechanisms into functional mapping. But the model that can address how...
epistatic interactions between different QTNs affect the pattern and form of drug response and timing of various events in the response process has not been developed.

The motivation of this study is to develop a statistical method for detecting genetic interactions that control drug response. The new method will incorporate Lin and Wu (2006) QTN–QTN interaction model into the functional mapping framework to investigate the roles of epistatic interactions between different DNA sequence variants in encoding drug response. The epistatic model will take into account the autocorrelation of drug response measured at multiple dosage levels and provide a quantitative framework for a series of hypothesis tests regarding the interplay between genetic actions/interactions and the shape and pattern of drug response. The model can be used to estimate and test a network of genetic interactions, thus enabling clinical geneticists to draw a detailed picture of genetic control and regulation for the pharmacodynamic process of drug response. We will use an example of a pharmacogenetic study for two beta-adrenergic receptor (βAR) subtypes expressed in the heart to validate the usefulness of the epistasis model.

2 SYSTEMS AND METHODS

2.1 The normal mixture model

Approaches for QTL and QTN mapping are statistically similar in the conceptual formation of a mixture model. For QTL mapping, each observation must arise from one of multiple QTL genotypes, although the QTL genotype for this individual is unknown (Lander and Botstein, 1989). The normal mixture model is constructed to contain the possible impact of these QTL genotypes each of which is assumed to follow a normal distribution. As shown in Liu et al. (2004), QTN mapping is also based on a normal mixture model but in which the phenotypic value of a heterozygous individual for two or more SNPs is thought to arise as one of multiple diplotype constructs by a series of SNPs.

Diploype difference can be assumed to result from the composition of different haplotypes. We define the haplotype that is different from the rest of haplotypes as reference or risk haplotype (Liu et al., 2004). Those remaining haplotypes are thus defined as non-reference or non-risk haplotype. Consider a QTN (A) that is composed of a set of SNPs. Let A and B be the risk and non-risk haplotypes for this QTN, respectively, which thus form three different composite diplotypes, AA, A and A. In the mixture model, longitudinal observations for each composite diplotype, j, are characterized by a different multivariate normal distribution with mean vector μj and covariance matrix Σj.

For QTN mapping, only diplotypes that are heterozygous for two or more SNPs contain different composite diplotypes because these diplotypes are not consistent with their phenotypically observable genotypes. Therefore, the mixture model is formulated only for those diplotypes.

2.2 Epistatic effects

We developed an interactive model aimed to detect sequence–sequence epistasis. Let B and C be the risk and non-risk diplotypes at a second QTN (B), respectively. The two QTNs, A and B, generate nine different composite diplotypes expressed as ABB, ABB, ABB, ABB, ABB, ABB, ABB, ABB and ABB. Traditional quantitative genetic theories can be used to model the genetic effects of the composite diplotypes (Lynch and Walsh, 1998). The time or dose-dependent vector for the genotypic value (μAB) of a joint composite diplotype at the two QTNs can be decomposed into nine different components as follows:

\[
\begin{align*}
\mu_{AB} &= \mu + j_1a + j_2b + j_3d + j_4a_b + j_5a_d + j_6b_d + j_7a_b_d + j_8a_d_b + j_9a_d_b_d \\
&= \mu + j_1a + j_2b + j_3d + j_4(a_b + a_d + b_d) + j_5(a_b_d + a_d_b + a_d_b_d)
\end{align*}
\]

where

\[
j_{AB} = \left\{ \begin{array}{ll}
1 & \text{for } AA \text{ or } BB \\
0 & \text{for } A \text{ or } B \\
-1 & \text{for } A \text{ and } B
\end{array} \right.
\]

stand for the composite diplotypes at QTN A and B, respectively, a and d are the additive and dominant effect vectors at the corresponding QTN, respectively, and i, j, k and l are the additive × additive, additive × dominant, dominant × additive and dominant × dominant epistatic effect vectors between the two QTNs, respectively.

If the maximum likelihood estimates (MLEs) of the genotypic value vectors at the left side of Equation (1) is observed, we can solve the vectors for the overall mean, additive, dominant and four kinds of epistatic effects between two QTNs by

\[
\begin{align*}
u &= \frac{1}{4}(u_{AABB} + u_{AABB} + u_{AABB} + u_{AABB}) \\
a &= \frac{1}{4}(u_{AABB} - u_{AABB} + u_{AABB} - u_{AABB}) \\
b &= \frac{1}{4}(u_{AABB} - u_{AABB} - u_{AABB} + u_{AABB}) \\
d &= \frac{1}{4}(2u_{AABB} - u_{AABB} - u_{AABB} + 2u_{AABB}) \\
i_{aa} &= \frac{1}{4}(u_{AABB} - u_{AABB} - u_{AABB} + u_{AABB}) \\
i_{ab} &= \frac{1}{4}(u_{AABB} - u_{AABB} - 2u_{AABB} + u_{AABB}) \\
i_{ad} &= \frac{1}{4}(u_{AABB} - 2u_{AABB} + u_{AABB}) \\
i_{dd} &= \frac{1}{4}(u_{AABB} + u_{AABB} + u_{AABB} + u_{AABB}) \\
i_{ab} &= \frac{1}{4}(u_{AABB} - 2u_{AABB} + u_{AABB} + u_{AABB}) \\
i_{ad} &= \frac{1}{4}(u_{AABB} - 2u_{AABB} + 2u_{AABB} + 2u_{AABB}) \\
i_{dd} &= \frac{1}{4}(2u_{AABB} - 2u_{AABB} + u_{AABB} + u_{AABB}) \\
i_{ad} &= \frac{1}{4}(2u_{AABB} - u_{AABB} + 2u_{AABB} + u_{AABB}) \\
i_{dd} &= \frac{1}{4}(4u_{AABB} + u_{AABB} + u_{AABB} + u_{AABB})
\end{align*}
\]
phenotypically the same, and they can be expressed as a function of haploid frequencies.

Suppose there are $R$ and $S$ SNPs for QTNA and QTNB, respectively. The two alleles, 1 and 0, at each of these SNPs are symbolized by $k_1, \ldots, k_R$ and $l_1, \ldots, l_S$, respectively. A haplotype frequency is denoted as $p_{i_1 \ldots i_R}^A$ for QTNA and $p_{j_1 \ldots j_S}^B$ for QTNB. In this article, two QTNs are assumed to be independent in their frequency. Thus, across-QTN haplotype frequencies can be calculated as the product of the corresponding haplotype frequencies from a different QTN, expressed as

$$p_{i_1 \ldots i_R, j_1 \ldots j_S} = p_{i_1 \ldots i_R}^A p_{j_1 \ldots j_S}^B$$

where, the parentheses are used to separate two different QTNs for a given across-QTN haplotype. With these across-QTN haplotype frequencies, expected across-QTN diplotype frequencies and across-QTN genotype frequencies can be calculated, respectively, under Hardy–Weinberg equilibrium (Lynch and Walsh, 1998; Wu et al., 2002).

With across-QTN diplotype and genotype frequencies, the likelihood function based on across-QTN genotype observations, $n = \{n_{i_1 \ldots i_R, j_1 \ldots j_S}(i_1 \ldots i_R, j_1 \ldots j_S)\}$, can be constructed. We simplify the presentation of our model by assuming two SNPs for each QTN. In Supplementary Table 1, all possible genotypes and diplotypes are given as well as their frequencies at two SNPs genotyped from a QTN (say A). Each diplotype is composed of two haplotypes, one from the mother and the other from the father. The diplotype frequencies can be expressed in terms of the haplotype frequencies. The same genotype may contain two different diplotypes, depending on its heterozygosity. Supplementary Table 1 also provides the relative frequencies with which a genotype carries a particular haplotype. Such relative frequencies will be useful for partitioning observed genotypes into underlying diplotypes in the construction of likelihood function based on the phenotype.

Let $\Omega_A = \{p_{i_1 \ldots i_R}^A, p_{j_1 \ldots j_S}^B\}$ be the population genetic parameters to be estimated for QTNA and QTNB. Lin and Wu (2006) formulated a log-likelihood function of these unknown haplotype frequencies given observed genotypes (n) in a multinomial form. They also provided a series of closed-form solutions for the expectation-maximization (EM) algorithm to estimate these haplotype frequencies.

Haplotype frequencies can be expressed as a function of allelic frequencies and linkage disequilibria (LD). For a two-SNP haplotype for QTNA, we have

$$p_{i_1 \ldots i_R}^A = \frac{n_{i_1 \ldots i_R}^A}{n_{i_1 \ldots i_R}^A} = \left(1 - \sum_{i_{R+1} \ldots i_R} D_{i_{R+1} \ldots i_R} \right)^{-1} \frac{n_{i_1 \ldots i_R}^A}{n_{i_1 \ldots i_R}^A},$$

where, $D_{i_1 \ldots i_R}$ is the linkage disequilibrium between the two SNPs at QTNA. Thus, once haplotype frequencies are estimated, we can estimate allelic frequencies and LD by solving Equation (4). Similar calculations are also done for QTNB.

While traditional models assume the association between the genotype and drug response (Gong et al., 2004), our model can estimate the effects of different diplotypes on the pharmacodynamic response of drugs. A particular four-SNP genotype from two QTNs, expressed as $(k_1 k_2 k_3 k_4) (l_1, l_2 l_3 l_4)$, can be partitioned into four possible diplotypes, $\{kJ k_3 k_4 l_1 \l_2 l_3 l_4\}$, $\{k_1 k_2 \{kJ k_3 k_4\} l_1 \l_2 l_3 l_4\}$, $\{kJ k_3 k_4\} l_1 \l_2 l_3 l_4\}$ and $\{k_1 k_2 k_3 k_4\}$ $\{kJ k_3 k_4\} l_1 \l_2 l_3 l_4\}$. Let $\Omega_A$ be quantitative genetic parameters that specify the mean vectors of different diplotypes and the residual covariance matrix. The likelihood of unknown population ($\Omega_A$) and quantitative genetic parameters ($\Omega_A$) given a drug response measured at C dosage levels, $y = \{y_1, \ldots, y_C\}$, and SNP genotype observations, n, can be constructed (see Supplementary Material—Likelihood).

Assume that haplotype [11] (denoted by A) is the risk haplotype for QTNA, and haplotype [10] (denoted B) is the risk haplotype for QTNB, respectively. As shown above, this leads to nine different across-QTN composite diplotypes. Equation (1) provides the structure of an arbitrary across-QTN composite diplotype formed by the risk and non-risk haplotypes. For a given composite diplotype $j_A j_B$, we have a multivariate normal distribution expressed as

$$f_{j_A j_B}(y; \Omega_A) = \frac{1}{(2\pi)^{C/2} |\Sigma|^{1/2}} \exp \left[-\frac{1}{2} (y - u_{j_A j_B}) \Sigma^{-1} (y - u_{j_A j_B}) \right].$$

At a particular dosage $c$, the relationship between the observation and genotypic value can be described by a regression model,

$$y(c) = \sum_{j_A j_B = 1}^{2} \xi_{j_A j_B} u_{j_A j_B}(c) + e(c),$$

where, $\xi_{j_A j_B}$ is the indicator variable denoted as 1 if a composite diplotype $j_A j_B$ is considered for subject $i$ and 0 otherwise and $e(c)$ is the residual error (i.e. the accumulative effect of polygenes and errors). The error term is specified by a structured antedependence (SAD) model (Zhao et al., 2005), in which the error at a particular dosage level is determined by the error at a previous dosage level and innovative error that occurs at the current dosage level (Gabriel, 1962). Ultimately, the covariance matrix in Equation (5) is described by two parameters, antedependence parameter and innovation variance.

### 2.4 Modeling the mean-covariance structure

Traditional genetic mapping for multiple traits attempts to estimate each element in the mean vector and the covariance matrix. This may not be efficient and effective for two reasons. First, when dosage level $c$ is high, an exponentially increasing number of parameters need to be estimated. Second, this approach does not consider biological principles underlying pharmacodynamic responses. The mainstay of modeling drug response is the Hill, or sigmoid Emax, equation, which postulates the following relationship between drug concentration ($c$) and drug effect (E) (Giraldo, 2003)

$$E = \frac{E_{max}c^H}{EC_{50}^{E} + c^H},$$

where $E_{max}$ 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. 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 composite diplotypes.

The composite diplotype-specific mean vectors in Equation (5) will be modeled by the $E_{max}$ model, expressed as

$$u_{j_A j_B} = \{u_{j_A j_B}(1), \ldots, u_{j_A j_B}(C)\} = \left\{ \frac{E_{max}}{E_{max} + EC_{50}^{E} + c^H}, \ldots, \frac{E_{max}}{E_{max} + EC_{50}^{E} + c^H} \right\},$$

where the mathematical parameters (($E_{max}$, $EC_{50}$, $H_{j_A j_B}$)) are described by $\Theta_{j_A j_B}$, describe the drug response profile for composite diplotype $j_A j_B$. Thus, based on Equation (8), our estimates will be concentrated on $\Theta_{j_A j_B}$ rather than on $u_{j_A j_B}$. This modeling of the mean vectors has two advantages: (1) clinically meaningful curves are used in genetic mapping.
so that the results will be closer to biological reality, and (2) the number of parameters to be estimated is reduced, thus increasing the power of the model to detect significant QTN and their interactions.

It is not parsimonious to estimate all the elements in the within-subject covariance matrix among different dose levels because some structure exists for dosage-dependent variances and correlations. The structure of the residual covariance matrix in (5) can be modeled by Zimmerman and Núñez-Antón’s (2001) SAD model. Using matrix notation, the error term in (6) can be expressed as

\[ e \sim \mathcal{N}(0, \mathbf{A}) \]

where \( e = [\epsilon_1, \ldots, \epsilon(C)] \), \( \epsilon = [\epsilon(1), \ldots, \epsilon(C)] \) and for the first-order SAD or SAD(1) model

\[
\mathbf{A} = \begin{pmatrix}
1 & 0 & 0 & 0 \\
\phi_1 & 1 & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots \\
\phi_1^{C-1} & \phi_1^{C-2} & \cdots & \phi_1^1
\end{pmatrix}
\]

(10)

The variance-covariance matrix of the longitudinal trait is then expressed as

\[ \mathbf{\Sigma} = \mathbf{A} \mathbf{\Sigma}_c \mathbf{A}^T \]

(11)

where, \( \mathbf{\Sigma}_c \) is the innovation variance-covariance matrix and is expressed as:

\[
\mathbf{\Sigma}_c = \begin{pmatrix}
\sigma^2(1) & 0 & 0 & \cdots & 0 \\
0 & \sigma^2(2) & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & \sigma^2(C)
\end{pmatrix}
\]

The closed forms for the inverse and determinant of matrix \( \mathbf{\Sigma} \) help to estimate the parameters, \( \mathbf{\Theta}_c = (\phi_1, \ldots, \phi_C, \sigma^2) \), that model the matrix. The vector \( \mathbf{\Theta}_{\text{iA}} \) and \( \mathbf{\Theta}_c \) form the quantitative genetic parameters \( \mathbf{\Theta}_c \).

3 HYPOTHESIS TESTS

With our epistatic model, we can make a number of hypothesis tests regarding the genetic control of overall drug response to a spectrum of dosages and other clinically important events. Two major hypotheses can be made in the following sequence: (1) the association between different SNPs within each QTN by testing their LD, and (2) the significance of an assumed across-QTN risk haplotype for its effect on drug response. The LD between two given SNPs within QT A can be tested using the following hypotheses:

\[ H_0: D_A = 0 \quad \text{versus} \quad H_1: D_A \neq 0 \]

(12)

The log-likelihood ratio (LR) test statistic for the significance of LD is calculated by comparing the likelihood values under the \( H_1 \) (full model) and \( H_0 \) (reduced model) using

\[
\text{LR}_{D_A}^H = -2\log \left( \frac{L(\hat{\mathbf{p}}_{D_A}^H, \hat{\mathbf{p}}_{E}^H; D_A = 0, \hat{\mathbf{\Theta}}_{\text{iA}}, \hat{\mathbf{\Theta}}_c; \mathbf{n})}{L(\hat{\mathbf{\Theta}}_{\text{iA}}, \hat{\mathbf{\Theta}}_c; \mathbf{n})} \right)
\]

where, the tilde and hat denote the MLEs of unknown parameters under \( H_0 \) and \( H_1 \), respectively. The \( \text{LR}_{D_A}^H \) is considered to asymptotically follow a \( \chi^2 \) distribution with one degree of freedom. The MLEs of allelic frequencies under \( H_0 \) can be estimated using the EM algorithm described above, but with the constraint \( p_{10}^H p_{01}^H = p_{00}^H p_{11}^H \). A similar test can be made for QTN B.

Diplotype or haplotype effects on a complex disease can be tested using null and alternative hypotheses expressed as

\[
\begin{align*}
H_0 : & \mathbf{\Theta}_{\text{iA}} = \mathbf{\Theta}(\lambda_A, \beta_B = -1, 0, 1) \\
H_1 : & \text{at least one equality in } H_0 \text{ does not hold}
\end{align*}
\]

(13)

The log-likelihood ratio test statistic (LR) under these two hypotheses can be similarly calculated. Unlike the significance test used in traditional QTL mapping (Lander and Botstein, 1989), the \( H_0 \) of hypothesis (13) for QTN haplotyping does not contain a nonidentifiable parameter and, therefore, the distribution of the LR calculated can be thought to asymptotically follow a \( \chi^2 \) distribution with 24 degrees of freedom. In practice, if the sample size used is not adequately large, the permutation test approach proposed by Churchill and Doerge (1994), which does not rely upon the distribution of the \( LR_e \), may be used to determine the critical threshold for determining the existence of a QTN.

Different genetic effects, such as the additive (\( a_A \) and \( a_B \)), dominant (\( d_A \) and \( d_B \)) and additive × additive (\( i_{aa} \)), additive × dominant (\( i_{ad} \)), dominant × additive (\( i_{da} \)) and dominant × dominant (\( i_{dd} \)) between QTN \( \mathbf{A} \) and \( \mathbf{B} \) can also be tested individually. The critical thresholds for these individual effects can be obtained from a \( \chi^2 \)-distribution table with one degree of freedom. But they can be empirically determined on the basis of simulation studies if the sample size used is not adequately large.

Our model allows for the tests of different outcomes for drug response. For overall drug response curves, hypothesis tests based on mathematical parameters can be informative for the characterization of genetic control. One alternative test can be performed on the area under curve (AUC). The AUC can be calculated by taking integral for each composite haplotype, expressed as

\[
\text{AUC}_{\lambda_A} = \int_a^c \frac{E_{\lambda_A}^c}{E_{\lambda_A}^c + c^{\lambda_A} \mathcal{N}} \, dc.
\]

The null hypothesis based on the AUC can be formulated as \( \text{AUC}_{\lambda_A} = \text{AUC} \). The permutation tests can be used to determine the critical value for AUC-related hypothesis tests.

4 DISCUSSION AND CONCLUSION

4.1 Worked example

To show how our model works in practice, we use it to analyze a real example from a pharmacogenetic study for drug response. Numerous genes have been investigated to affect heart function (Chagnon et al., 2003; Mason et al., 1999). The \( \beta_1 \)AR and \( \beta_2 \)AR genes are two such examples (Green et al., 1995; Large et al., 1997) in each of which there are several polymorphisms common in the population. Two common polymorphisms are identified at codons 49 (Ser49Gly) and 389 (Arg389Gly) for the \( \beta_1 \)AR gene and at codons 16 (Arg16Gly) and 27 (Gln27Glu) for the \( \beta_2 \)AR gene, respectively. The polymorphisms in each of these two receptor genes are in linkage disequilibrium, which suggests the importance of taking...
into account haplotypes, rather than a single polymorphism, when defining biological function. This study attempts to detect haplotype variants within these candidate genes which determine drug response.

To determine whether sequence variants at the two polymorphisms from one gene interact with those from the other gene to affect drug response, a group of 155 men and women were investigated with ages from 32 to 86 years old with a large variation in heart rates in response to different dosage levels of dobutamine. Dobutamine was injected into these subjects to investigate their response in heart rate to this drug. The subjects received increasing doses of dobutamine, until they achieved target heart rate response or predetermined maximum dose. The dose levels used were 0 (baseline), 5, 10, 20, 30 and 40 mcg/min, at each of which heart rate was measured. The time interval of 3 min is allowed between two successive doses for subjects to reach a plateau in response to that dose. Of the subjects, 57 reached the thresholds of their heart rates at the maximum dose. The dose levels used were 0 (baseline), 5, 10, 20, 30 and 40 mcg/min, at each of which heart rate was measured.

Each of these patients was determined for their genotypes at codon 49 with two allele Ser49 (A) and Gly49 (G) and codon 389 with two alleles Arg389 (C) and Gly389 (G) within the β1AR gene on chromosome 10 as well as at codon 16 with two alleles Arg16 (A) and Gly16 (G) and codon 27 with two alleles Gln27 (C) and Glu27 (G) within the β2AR gene on chromosome five, and measured for the response in heart rate to dobutamine. Two SNPs from each gene theoretically form 81 composite genotypes drawn with the estimated haplotype combinations. The maximal LR value (56.32) with all the nine composite genotypes was obtained when GC as the risk haplotype of the β1AR gene is combined with GG as the risk haplotype of the β2AR gene. The LR values for the other combinations range from 4.62 to 40.52. The optimal risk-haplotype combination for drug response is statistically tested with 1000 permutation tests which obtained the critical threshold value of 54.08 at the 5% significance level. This test suggests that there exist significant haplotype effects at these two candidate genes on heart rate curves.

Statistical tests for individual genetic effects for DNA sequence variants based on Equation (1) suggest that the additive and dominant effects at each of the candidate genes are highly significant. Four kinds of epistasis between the two genes are all significant at the 0.0001 significance level (Table 2). Figure 1 displays the profiles of heart rate to increasing concentration levels of dobutamine for nine across-QTN composite genotypes drawn with the estimated response parameters given in Supplementary Table 2. Considerable over-crossing among different curves suggests sequence–sequence interactions within β1AR and β2AR.

### 4.2 Simulation

This model has been examined through simulation studies (Supplementary Material—Simulation Result). The results suggest that the model provides reasonably precise estimates of the population genetic parameters, and the parameters that model drug response for different diplotypes and the covariance matrix of drug response measured at different dosage levels. The model is quite stable and robust in which reasonable power for QTN detection can be obtained for a small sample size (100) or a small heritability (0.1).
4.3 Concluding remarks

Given a recent upsurge of interest in pharmacogenetics and pharmacogenomics, there is a pressing need for the development of statistical models for unraveling the genetic etiology of pharmacovariation. QTL mapping, by superimposing real biological phenotypes on genome sequence, structural polymorphisms and gene expression data, can provide an unbiased view of the network of gene actions and interactions that build a complex phenotype-like drug response. Through an integrated approach, studies can move to characterize the best drug, the best dosage of the drug and the best injection time for individual patients based on their genetic structure. By incorporating genetic tests into the prescription process, physicians might improve outcomes for patients.

Genetically, drug response is a complex trait, involving multiple biochemical pathways each controlled by different interacting genes (Marchini et al., 2005). Traditional QTL mapping can probe the important chromosomal segments for drug response, but is difficult to identify the causal DNA sequences of these segments. The completion of the human genome project makes it possible to genome-wide associate DNA sequence variants with complex phenotypes. The model proposed in this article has capacity to characterize the effects of genetic factors and their epistatic effects on drug response at the DNA sequence level.

There are several ways in which our model may be extended. First, the model can be modified to consider different interacting QTNs that are in linkage disequilibrium by estimating more disequilibrium parameters. Nourreddine et al. (2005) identified the linkage disequilibrium between modifier genes and genes for Parkinson disease. Second, with the principle of two-SNP interactions, we can model an arbitrary number of SNPs whose sequences are associated with the phenotypic variation. As shown in Lin and Wu (2006), the multi-SNP sequencing model will encounter the problem of many haplotypes and haplotype pairs. An AIC- or BIC-based model selection strategy (Burnham and Anderson, 1998) has been framed to determine the haplotype that is most distinct from the rest of haplotypes in explaining quantitative variation. Third, our model is based on the data set without genotyping errors. It has been well observed that even small genotyping errors can substantially reduce power to genetic association (Gordon et al. 2002; Kang et al., 2004). The impact of genotyping errors and misclassification on the detection and estimation of sequence–sequence interactions deserves in-depth investigations.

ACKNOWLEDGEMENTS

We thank the three anonymous referees for their constructive comments that have improved the presentation of the manuscript. The preparation of this manuscript is partially supported by NSF grant (No. 0540745).

Conflict of Interest: none declared.

REFERENCES

Arranz, M.J. et al. (2003) Pharmacogenetic and pharmacogenomic research in psychiatry: current advances and clinical applications. Curr. Pharmacogenomics, 1, 151–158.

Burnham, K.P. and Anderson, D. (1998) Model Selection and Inference. A Practical Information-Theoretic Approach. Springer, New York.

Chagnon, Y.C. et al. (2003) The human obesity gene map: the 2002 update. Obes. Res., 11, 313–367.

Churchill, G.A. and Doerge, W.R. (1994) Empirical threshold values for quantitative trait mapping. Genetics, 138, 963–971.

Frankel, W.N. and Schork, N.J. (1996) Who’s afraid of epistasis? Nat. Genet., 14, 371–373.

Gabriel, K.R. (1962) Ante-dependence analysis of an ordered set of variables. Ann. Math. Stat., 33, 201–212.

Giraldo, J. (2003) Empirical models and Hill coefficients. Trends Pharmacol. Sci., 24, 63–65.

Gong, Y. et al. (2004) A statistical model for functional mapping of quantitative trait loci regulating drug response. Pharmacogenomics J., 4, 315–321.

Gordon, D. et al. (2002) Power and sample size calculations for case-control genetic association tests when errors are present: application to single nucleotide polymorphisms. Hum. Hered., 54, 22–33.

Green, S.A. et al. (1995) Implications of genetic variability of human 2-adrenergic receptor structure. Palm. Pharmacol., 8, 1–10.

Kang, S.J. et al. (2004) Quantifying the percent increase in minimum sample size for SNP genotyping errors in genetic model-based association studies. Hum. Hered., 58, 139–144.

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

Large, V. et al. (1997) Human beta-2 adrenoceptor gene polymorphisms are highly frequent in obesity and associate with altered adipocyte beta-2 adrenoceptor function. J. Clin. Invest., 100, 3005–3013.

Lin, M. et al. (2005) Sequencing drug response with HapMap. Pharmacogenomics J., 5, 149–156.

Lin, M. and Wu, R.L. (2006) Detecting sequence–sequence interactions for complex diseases. Curr. Genomics, 7, 59–72.

Liu, T. et al. (2004) Sequencing complex diseases with HapMap. Genetics, 168, 503–511.

Lynch, M. and Walsh, B. (1998) Genetics and Analysis of Quantitative Traits. Sinauer, Sunderland, MA, USA.

Ma, C.-X. et al. (2002) Functional mapping of quantitative trait loci underlying the character process: a theoretical framework. Genetics, 161, 1751–1762.

Marchini, J. et al. (2005) Genome-wide strategies for detecting multiple loci that influence complex diseases. Nat. Genet., 37, 413–417.

Fig. 1. Nine profiles of heart rate in response to different dosages of dobutamine, each representing a different across-QTN composite diploig (foreground), identified for SNPs within two genes. The profiles of 98 studied subjects are also shown (background).
Mason, D.A. et al. (1999) A gain-of-function polymorphism in a G-protein coupling domain of the human β1-adrenergic receptor. J. Biol. Chem., 274, 12670–12674.

Moore, J.H. (2003) The ubiquitous nature of epistasis in determining susceptibility to common human diseases. Hum. Hered., 56, 73–82.

Noureddine, M.A. et al. (2005) Association between the neuron-specific RNA-binding protein ELAVL4 and Parkinson disease. Hum. Genet., 117, 27–33.

Serretti, A. and Artioli, P. (2003) Predicting response to lithium in mood disorders: role of genetic polymorphisms. Am. J. Pharmacogenomics, 3, 17–30.

Tate, S.K. et al. (2005) Genetic predictors of the maximum doses patients receive during clinical use of the anti-epileptic drugs carbamazepine and phenytoin. Proc. Natl Acad. Sci. USA, 102, 5507–5512.

Wu, R.L. et al. (2002) Joint linkage and linkage disequilibrium mapping of quantitative trait loci in natural populations. Genetics, 160, 779–792.

Wu, R.L. et al. (2003) Molecular dissection of allometry, ontogeny and plasticity: A genomic view of developmental biology. BioScience, 53, 1041–1047.

Wu, R.L. et al. (2004a) A general framework for analyzing the genetic architecture of developmental characteristics. Genetics, 166, 1541–1551.

Wu, R.L. et al. (2004b) Functional mapping of quantitative trait loci underlying growth trajectories using a transform-both-sides logistic model. Biometrics, 60, 729–738.

Wu, R.L. et al. (2004c) A mechanistic model for genetic machinery of ontogenetic growth. Genetics, 168, 2383–2394.

Wu, R.L. and Lin, M. (2006) Functional mapping – how to map and study the genetic architecture of dynamic complex traits. Nat. Rev. Genet., 7, 229–237.

Zhao, W. et al. (2005) A non-stationary model for functional mapping of longitudinal quantitative traits. Bioinformatics, 21, 2469–2477.

Zimmerman, D.L. and Núñez-Antón, V. (2001) Parametric modeling of growth curve data: an overview (with discussion). Test, 10, 1–73.