

METHODOLOGY ARTICLE

Open Access

# Incorporation of covariates in simultaneous localization of two linked loci using affected relative pairs

Yen-Feng Chiu<sup>1\*</sup>, Jeng-Min Chiou<sup>2</sup>, Kung-Yee Liang<sup>3,4</sup>, Chun-Yi Lee<sup>1</sup>

## Abstract

**Background:** Many dichotomous traits for complex diseases are often involved more than one locus and/or associated with quantitative biomarkers or environmental factors. Incorporating these quantitative variables into linkage analysis as well as localizing two linked disease loci simultaneously could therefore improve the efficiency in mapping genes. We extended the robust multipoint Identity-by-Descent (IBD) approach with incorporation of covariates developed previously to simultaneously estimate two linked loci using different types of affected relative pairs (ARPs).

**Results:** We showed that the efficiency was enhanced by incorporating a quantitative covariate parametrically or non-parametrically while localizing two disease loci using ARPs. In addition to its help in identifying factors associated with the disease and in improving the efficiency in estimating disease loci, this extension also allows investigators to account for heterogeneity in risk-ratios for different ARPs. Data released from the collaborative study on the genetics of alcoholism (COGA) for Genetic Analysis Workshop 14 (GAW 14) were used to illustrate the application of this extended method.

**Conclusions:** The simulation studies and example illustrated that the efficiency in estimating disease loci was demonstratively enhanced by incorporating a quantitative covariate and by using all relative pairs while mapping two linked loci simultaneously.

## Background

With the advance of genotyping techniques, genome-wide association analysis has become the mainstream technique in genetic mapping. However, studies have shown that using information from linkage scans can improve the power of association mapping in genome scans [1]. In addition, linkage analysis could be more powerful than association analysis for some genetic mechanisms; family data can also help to estimate familial risks [2]. Hence, linkage analysis remains a useful and supplemental tool to map genes for complex diseases. As complex diseases often involve quantitative biomarkers or environmental factors, incorporating these quantitative factors into linkage mapping can improve the power to detect disease loci [3] or the

efficiency of estimating disease loci. Efficiency is defined as the inverse of the variance estimate for the disease locus estimate. Thus, smaller variance estimates have higher efficiencies. Moreover, the incorporation of covariates provides information that can be used to characterize disease loci, which is important for understanding disease etiologies and mechanisms and for identifying population subgroups that may have particularly high disease risks [4]. Methodologic work has demonstrated that failure to adequately account for gene-covariate interaction in a genetic analysis can mask the effects of both genes and covariates [5-7]. Hence, it is important to develop linkage approaches that allow the inclusion of covariates.

Thus far, several linkage analyses including covariates have been proposed to account for linkage heterogeneity or to examine biological, environmental, gene-gene or gene-environment interaction effects. Devlin (2002) [5] accounts for linkage heterogeneity by incorporating a

\* Correspondence: yfchiu@nhri.org.tw

<sup>1</sup>Division of Biostatistics and Bioinformatics, Institute of Population Health Sciences, National Health Research Institutes, 35 Keyan Rd., Zhunan, Miaoli 350, Taiwan

family-level covariate into likelihood-based mixture models; however, this approach accounts for linkage heterogeneity only. Greenwood and Greenwood (1997, 1999) [6,8] incorporated covariates into genome scanning approaches using sib-pair or relative-pair through model-based logarithms of odds (LOD) score approaches, where the generalized expected identity-by-descent (IBD) sharing was modeled as a function of some covariates through multinomial logistic regression. Rice (1999) [7] applied a novel technique to detect significant covariates in linkage analyses with a logistic regression approach using all sib pairs (concordant affected, concordant unaffected, and discordant), and Saccone et al. (2001) [9] further extended this analysis to cousin pairs. Olson (1999) [10] proposed a unified framework for model-free linkage analysis that can handle the separate inclusion of other ARPs, discordant relative pairs, covariates, or additional disease loci through a conditional-logistic parameterization. These regression-based approaches can easily be generalized to include all covariates; however, they assume either one disease locus or multiple unlinked loci and thus are not applicable to analyses of multiple linked loci. For non-regression-based approaches, Hauser et al. (2004) [11] proposed a model-free LOD scores approach that includes family-level covariate information. This approach also assumes only one disease locus and can only incorporate one covariate at a time. In addition, the problem of multiple testing may arise when researchers perform multiple tests or analyses using various combinations of multiple loci or covariates using these approaches.

On the other hand, most two-locus linkage approaches aim to detect the presence of a second susceptibility gene by accounting for the effects of a known susceptibility gene [12-14]. However, when two susceptibility loci are linked, the location of the first gene may be inaccurate because it was mapped without accounting for the effects of the linked gene. Thus, conditional analyses that rely on an inaccurate position for the first locus may result in an inaccurate second disease loci estimate as well. Biswas et al. (2003) [15] applied a Bayesian approach to simultaneously detect two linked disease genes; however, their approach was designed to detect genes under locus heterogeneity only, and this model-based approach requires the specification of unknown genetic parameters. Hence, linkage approaches that can simultaneously localize two linked disease genes are in great demand.

Rather than testing the presence of linkage, Liang et al. (2001) [16] developed a novel, robust, model-free multipoint linkage method that simultaneously estimates both the position of a disease locus as well as its effect on the disease, along with its sampling uncertainty. The

advantages of this method include: (i) It does not require specification of an underlying genetic model; hence, estimation of the parameters is robust to a wide variety of genetic mechanisms. (ii) The multiple testing issue is eliminated as a single test statistic is provided for linkage in the entire studied region; rather than testing the hypothesis for one marker at a time. (iii) While multiple markers are incorporated simultaneously in the gene mapping, there is no need to specify the phase of genotypic data with multiple markers. Many complex diseases, such as hypertension, schizophrenia, diabetes, and asthma are usually defined as dichotomous phenotypic traits; however, they are also associated with quantitative biological markers or quantitative risk factors. As a result, Glidden et al. (2003) [17] further incorporated quantitative covariates into Liang's approach [16] and estimated the genetic effect of a disease locus through a logistic-type parametric model using affected sib pairs (ASPs). Based on the same study design, Chiou et al. (2005) [18] incorporated quantitative covariates into their linkage mapping and estimated the genetic effect of a disease locus non-parametrically. This quantitative covariate could be either an environmental risk factor or itself a quantitative trait. For the quantitative trait incorporated as a covariate, its QTL (quantitative trait locus) may directly underlie a pathway of the disease or be linked to the disease locus, or the trait may be indirectly associated with the disease.

Meanwhile, Schaid et al. (2005) [19] extended the without-a-covariate approach by Liang et al. [16] to different types of ARPs. The authors' extension relaxed the limitation to ASPs only and allowed an investigator to study the risk-ratios of a disease gene estimated from multiple relative pairs; this work helped to uncover the underlying genetic mechanism of disease. To jointly localize two linked disease loci using ASP data, Biernacka et al. (2005) [20] extended this approach [16] to the localization of two linked disease-susceptibility genes. They also provided tests for the presence of two linked disease-susceptibility genes by a quasi-likelihood ratio test and a modified score test in another article [21]. Lin and Schaid (2007) [22] generalized the two-locus localization method to a variety of ARPs. Both of the unconstrained and constrained models, along with a score test and the examination of the goodness of fit of a used constrained model, were described in their generalized method. As the etiology of complex diseases often involves quantitative variables (either genetic biomarkers or environmental factors) in addition to multiple disease loci, it is helpful to incorporate a quantitative variable while localizing two linked disease loci simultaneously using ARPs. We extended Lin and Schaid's (2007) [22] approach to incorporate quantitative covariates in two-locus linkage mapping using ARPs. Generally, a statistical parametric

model is simpler and easier to interpret than a non-parametric model, while a non-parametric model has the flexibility to fit the data perfectly. To take advantages of parametric and non-parametric statistical models, we applied both models to incorporate covariates. These methods can also be applied to account for heterogeneity from quantitative covariates as well as from multiple subgroups that are stratified by categorical covariates. Systematic simulation studies under a variety of quantitative covariates were conducted to evaluate the gain in efficiency of estimating the disease loci from the proposed methods. The estimates from the proposed approaches with incorporation of covariates were compared with those from the approach without incorporating covariates. The collaborative study on the genetics of alcoholism (COGA) data released for GAW14 was used to illustrate the proposed approaches.

## Methods

To incorporate relevant covariate information while simultaneously estimate the locations of two genes using all types of relative pairs in linkage analysis, we proposed the following linkage approaches.

### Simultaneous Localization of Two Linked Disease Susceptibility Genes with Incorporation of Covariates

Consider a chromosomal region harboring two linked disease loci,  $\tau_1$  and  $\tau_2$ , with  $M$  markers genotyped at the locations  $0 = t_1 < t_2 < \dots < t_M$ . Let  $S_{ki}(t_j)$  be the identity-by-descent (IBD) sharing for the  $j^{\text{th}}$  marker of the  $i^{\text{th}}$  pair of the ARP type  $k$ ,  $j = 1, \dots, M$ ,  $i = 1, \dots, n_k$ ,  $k = 1, \dots, 5$ . The five types of relative pairs considered include full siblings (SP,  $k = 1$ ), half siblings (HS,  $k = 2$ ), first cousins (FC,  $k = 3$ ), grandparent-grandchild pairs (GP,  $k = 4$ ) and avuncular pairs (AP,  $k = 5$ ) [19]. The five affected relative pairs are abbreviated as ASP, AHS, AFC, AGP and AAP. Let  $x_{ki1}$ ,  $x_{ki2}$  be the covariates associated with relatives 1 and 2 in the  $i^{\text{th}}$  relative pair of type  $k$ , respectively. Given the covariates and assuming that the recombination fraction does not depend on the covariates, the expectation of IBD sharing at  $t_j$  for a relative pair  $ki$  [22] is

$$\mu_{ki}(t_j) = E(S_{ki}(t_j) | x_{ki1}, x_{ki2}, \Phi) = \begin{cases} a_k + b_k(d_1)C_{1k}(x_{ki1}, x_{ki2}) & \text{for } t_j < \tau_1 < \tau_2, \\ a_k + \frac{b_k(d_1)[1 - b_k^2(d_2)]}{1 - b_k^2(d_3)} C_{1k}(x_{ki1}, x_{ki2}) + \frac{b_k(d_2)[1 - b_k^2(d_1)]}{1 - b_k^2(d_3)} C_{2k}(x_{ki1}, x_{ki2}) & \text{for } \tau_1 < t_j < \tau_2, \\ a_k + b_k(d_2)C_{2k}(x_{ki1}, x_{ki2}) & \text{for } \tau_1 < \tau_2 < t_j, \end{cases} \quad (1)$$

where  $C_{lk}(x_{ki1}, x_{ki2}) = E(S_{ki}(\tau_l) | x_{ki1}, x_{ki2}, \Phi)$  -  $a_k$  is the genetic effect at locus  $l$  for a relative pair  $ki$ ;  $l = 1, 2$ ;  $\Phi$

is the event of an ARP;  $d_1 = |\tau_1 - t_j|$ ,  $d_2 = |t_j - \tau_2|$ ;  $d_3 = |\tau_2 - \tau_1|$ ;  $a_k$  is the expected count for random sharing;  $b_k(d_v)$  controls the rate of decrease of expected sharing as the distance  $d_v$  from the trait locus increases; and  $v = 1, 2, 3$ . Haldane's mapping function was used to translate recombination fraction to map distance. The values of  $b_k(d_v)$  and  $d_v$  for each relative type  $k$  and functions relating the risk ratio  $\lambda$  to  $C$  are listed in supplemental Additional file 1 Table S1 (adopted from Table 1 in Lin and Schaid (2007) [22]).

$C_1$  and  $C_2$  represent the amount of excess IBD sharing at each of the two disease gene loci, which is increased by effects due to both disease genes. The simple "effect size" interpretation does not apply to  $C_1$  and  $C_2$  in the two-locus model because the magnitude of  $C_1$  depends not only on the effect of gene 1 but also on the distance between gene 1 and gene 2.  $C_1$  and  $C_2$  can each be reparameterized to represent excess sharing at a location due to the gene at that location and thus can be considered the "effect size" of that particular gene (see Appendix of [20], page 47). They can then be used to test for the presence of linkage. We applied parametric and non-parametric methods to model the association between the excess IBD sharing ( $C_l$ ) at  $\tau_l$ ,  $l = 1, 2$  and the covariates.

### Parametric Modeling on C

In the parametric model,  $C_{1k}$  and  $C_{2k}$  can be modeled as a function of covariates [17]; an example is the postulation of a logistic regression for IBD sharing at  $\tau_1$  and  $\tau_2$ . For a relative-pair type  $k$ , assuming  $G_{lk} = (g_{lk1}, \dots, g_{lkp})^T$  is the covariate vector,  $C_{1k}$  and  $C_{2k}$  were modeled separately, where  $g_{lkr} = g_{lkr}(x_{kr1}, x_{kr2})$ ,  $r = 1, \dots, p$ , indicate covariates.

$$C_{lk}(g_{lk}) = \frac{1 \exp(\alpha_{lk} + \beta_{lk}^T G_{lk})}{f_k * (1 + \exp(\alpha_{lk} + \beta_{lk}^T G_{lk}))}, \quad (2)$$

where  $\beta_{lk}^T = (\beta_{lk1}, \dots, \beta_{lkp})$ ,  $l = 1, 2$ ,  $k = 1, \dots, 5$ ;  $f_k = 1$  for ASP,  $f_k = 4$  for AFC, and,  $f_k = 2$  for other ARPs. The gene-environment interaction for environmental variable,  $x_r$ , could be assessed by examining whether the corresponding  $\beta$ -coefficient,  $\beta_r$ , is statistically significantly different from zero. In addition, the interactions between two covariates on the genetic effects of the disease loci could also be assessed by adding an interaction term between the two covariates.

### Nonparametric Modeling on C

For the non-parametric model, given the data  $(G_{ki}, S_{ki}^*(\tilde{\tau}_l))$ , where  $G_{lki} = (g_{lki1}, \dots, g_{lkip})^T$  with  $g_{lkir} = g_{lkir}(x_{kir1}, x_{kir2})$ ,  $r = 1, \dots, p$ ,  $i = 1, \dots, n_k$ , and the imputed IBD sharing  $S_{ki}^*(\tilde{\tau}_l)$  at  $\tilde{\tau}_l$ , which is a specified or estimated value of  $\tau_l$ , the estimator of  $C_{lk}$  at an arbitrary

**Table 1 Simultaneous two-locus search incorporating quantitative traits with QTLs at  $\tau_1(X_{QTL1})$  or  $\tau_2(X_{QTL2})$**

	Disease Loci (cM)				Estimate of C								95% coverage probability (%)				$y_l$ : covariate for modeling $C_l, l = 1, 2$
	Parametric		Nonparametric		Parametric				Nonparametric				Parametric		Nonparametric		
	$\tau_1$	$\tau_2$	$\tau_1$	$\tau_2$	ASP		AGP		ASP		AGP		$\tau_1$	$\tau_2$	$\tau_1$	$\tau_2$	
					$C_{11}$	$C_{21}$	$C_{14}$	$C_{24}$	$C_{11}$	$C_{21}$	$C_{14}$	$C_{24}$					
Bias	0.1	-0.1	-0.008	1.1	0.04	0.03	-0.05	0.01	-0.02	-0.05	-0.04	0.02	95	95	93	91	$y_1 = X_{QTL1}$
Sample variance	4.0	4.0	5.4	6.2	0.003	0.003	0.001	0.001	0.002	0.002	0.001	0.001					$y_2 = X_{QTL1}$
Mean variance	4.0	4.0	4.8	5.7													
$\check{\beta}_1$					0.26	-0.25	0.16	-0.08									
p-value					0.03	0.05	0.50	0.81									
Bias	0.2	-0.05	-1.1	-0.01	0.04	0.03	-0.05	0.02	-0.04	-0.02	-0.04	0.03	94	95	91	93	$y_1 = X_{QTL2}$
Sample variance	4.9	4.2	6.7	5.0	0.003	0.003	0.001	0.001	0.002	0.002	0.001	0.001					$y_2 = X_{QTL2}$
Mean variance	4.1	3.8	5.9	4.6													
$\check{\beta}_1$					-0.25	0.26	-0.09	0.16									
p-value					0.05	0.03	0.79	0.53									
Bias	0.1	-0.1	-0.5	0.5	0.04	0.03	-0.05	0.02	-0.02	-0.02	-0.04	0.03	94	94	91	91	$y_1 = X_{QTL1}$
Sample variance	4.5	4.5	5.7	5.4	0.003	0.003	0.001	0.001	0.002	0.002	0.001	0.001					$y_2 = X_{QTL2}$
Mean variance	3.9	3.8	4.8	4.6													
$\check{\beta}_1$					0.26	0.26	0.16	0.16									
p-value					0.03	0.03	0.50	0.53									
Bias	0.2	-0.1	-0.6	0.6	0.04	0.04	-0.05	0.02	-0.04	-0.05	-0.04	0.02	94	94	91	91	$y_1 = X_{QTL2}$
Sample variance	5.5	5.6	7.6	6.9	0.003	0.003	0.001	0.001	0.002	0.002	0.001	0.001					$y_2 = X_{QTL1}$
Mean variance	4.4	4.2	5.9	5.7													
$\check{\beta}_1$					-0.25	-0.25	-0.09	-0.09									
p-value					0.05	0.05	0.79	0.81									

target  $g_{lk} = (g_{lk1}, \dots, g_{lkp})^T$  is obtained by  $C_{lk}(g_{lk}) = \beta_{lk0}$  such that  $\beta_{lk} = (\beta_{lk0}, \beta_{lk1}, \dots, \beta_{lkp})$  is the minimizer of the following kernel-weighted least squares function with respect to  $\beta_{lk} = (\beta_{lk0}, \beta_{lk1}, \dots, \beta_{lkp}), \forall l = 1, 2$ ,

$$\sum_{k=1}^5 \sum_{i=1}^{n_k} [(S_{ki}^*(\tilde{\tau}_l) - a_k) - \beta_{lk0} - \beta_{lk1}(g_{ki1} - g_{k1}) - \dots - \beta_{lkp}(g_{kip} - g_{kp})]^2 K(H^{-1})(g_{lk} - G_{lki}), \quad (3)$$

where  $K$  is a p-variate Epanechnikov kernel function,  $H$  is a nonsingular square bandwidth matrix [18], and  $a_k$  is the expected count for random sharing [19].

#### Estimating $\tau_1$ and $\tau_2$

Given the function  $C_{lk}(x_{ki1}, x_{ki2})$ , the trait locus  $\tau_l$  can be estimated by solving the estimating equation [16,18] (4)

below. Once the estimate of  $C_{lk}$  is obtained, it can be plugged into the equation (4) and the estimate of  $\tau_l$  can be updated. That is, we replace  $C_{lk}(x_{ki1}, x_{ki2})$  with the estimate  $C_{lk}(x_{ki1}, x_{ki2})$ , which then yields the following estimating equation for  $\delta = (\tau_1, \tau_2)$ :

$$\sum_{k=1}^5 \sum_{i=1}^{n_k} \left( \frac{\partial \mu_{ki}(\delta)}{\partial \delta} \right)' Cov^{-1}(S_{ki})(S_{ki} - \mu_{ki}(\delta)) = 0, \quad (4)$$

where  $S_{ki} = (S_{ki}(t_1), \dots, S_{ki}(t_M))'$ , and

$$\mu_{ki}(\delta) = (\mu_{ki}(t_1; \hat{C}_{1k}, \hat{C}_{2k}, \delta), \dots, \mu_{ki}(t_M; \hat{C}_{1k}, \hat{C}_{2k}, \delta)),$$

with  $\mu_{ki}(t_j; C_{1k}, C_{2k}, \delta) = E(S_{ki}(t_j) | C_{1k}, C_{2k})$ .

The estimates of  $C_{lk}$  and  $\delta$  were iteratively updated until the convergent criteria for  $\delta$  were met. Assuming all relative pairs share a common  $\delta$ , the estimates of  $\delta$  follows asymptotic normality (see Additional file 2, Appendix for details) with a mean vector  $\delta$  and a covariance matrix  $\Sigma^{-1}$ , where.

$$\Sigma = \sum_{k=1}^5 \sum_{i=1}^{n_k} (\partial \mu_{ki}(\delta) / \partial \delta)' Cov^{-1}(S_{ki}) (\partial \mu_{ki}(\delta) / \partial \delta).$$

### Simulation Studies

Families with three generations including eight members were simulated: The first generation (4 grandparents) included one or zero affected subjects, the second generation had no affected members, and the third generation included two affected individuals. In total, 200 independent families were simulated, each including one affected sibpair. Of the 200 families, 100 included two affected grandparent-grandchild pairs, with the others not having any affected grandparent-grandchild pairs. Hence, there were 200 ASPs and 200 AGPs per replicate. In total, 1,000 replicates were simulated for each configuration.

#### One disease locus model

First, we extended the one-locus model proposed by Schaid et al. (2005) [19] with ARP to incorporate covariates using both parametric modeling [17] and non-parametric modeling [18]. We studied the enhancement of efficiency incurred by the incorporation of a quantitative covariate and by the usage of relative pairs in place of using sib pairs alone within a one-locus model. Three sets of penetrance rates ( $f_2, f_1, f_0$ ) for the genotypes of two high-risk alleles ( $f_2$ ), one high and one low-risk alleles ( $f_1$ ), and two low-risk alleles ( $f_0$ ) at the disease locus used in the simulation study were (i) (0.67,0.05,0.007) (recessive model), (ii) (0.67,0.55,0.007) (dominant model) and (iii) (0.8,0.4,0.0) (additive model), respectively.

A covariate might be directly or indirectly associated with the disease loci, and the information from covariates under different genetic mechanisms may differentially enhance the search for the disease loci. We studied a variety of covariates correlated with the disease trait under different scenarios: (1) a quantitative trait with a pleiotropic effect (that is to say a quantitative trait that is controlled by the disease locus,  $\tau_1$ , namely, its QTL is  $\tau_1$ , yet is not directly associated with liability of the disease); (2) a quantitative trait with a coincidence effect in which the QTL is linked to a disease locus by incidence, yet does not share common genetic components from the disease locus; (3) a quantitative trait unlinked to the disease loci; (4) a covariate of age

at onset with the distribution  $\log T = -\log \lambda - \beta Z + \varepsilon/\gamma$ , where  $Z$  is the number of copies of the disease allele [17] at one disease locus. The variable  $\varepsilon$  is distributed as a standard extreme-value random variable with  $\lambda = 0.03$ ,  $\gamma = 5.0$ , and  $\beta = 0.57$ ; this distribution was built while assuming that the disease allele frequency is 0.05. The distribution of age at onset ( $T$ ) followed a Weibull distribution, and the disease allele accelerated the onset of disease by a factor of 1.78. The threshold of age at onset was 70.

The quantitative trait  $\mathbf{y}$  for scenarios (1) - (3) follows a multivariate normal distribution  $\mathbf{y}_i = \boldsymbol{\mu}_i + \mathbf{g}_i + \mathbf{e}_i$ ,  $\mathbf{e}_i \sim N(\mathbf{0}, \boldsymbol{\Sigma}_i)$ ,  $i = 1, \dots, n$ , where  $\mathbf{y}_i = (y_{1i}, y_{2i}, \dots, y_{n_i})^T$ ,  $\mathbf{g}_i = (g_{1i}, g_{2i}, \dots, g_{n_i})^T$  and  $\mathbf{e}_i = (e_{1i}, e_{2i}, \dots, e_{n_i})^T$ .  $n_i$  is the total family members in the  $i^{\text{th}}$  family;  $\boldsymbol{\mu}$  is a  $n_i \times 1$  zero vector.

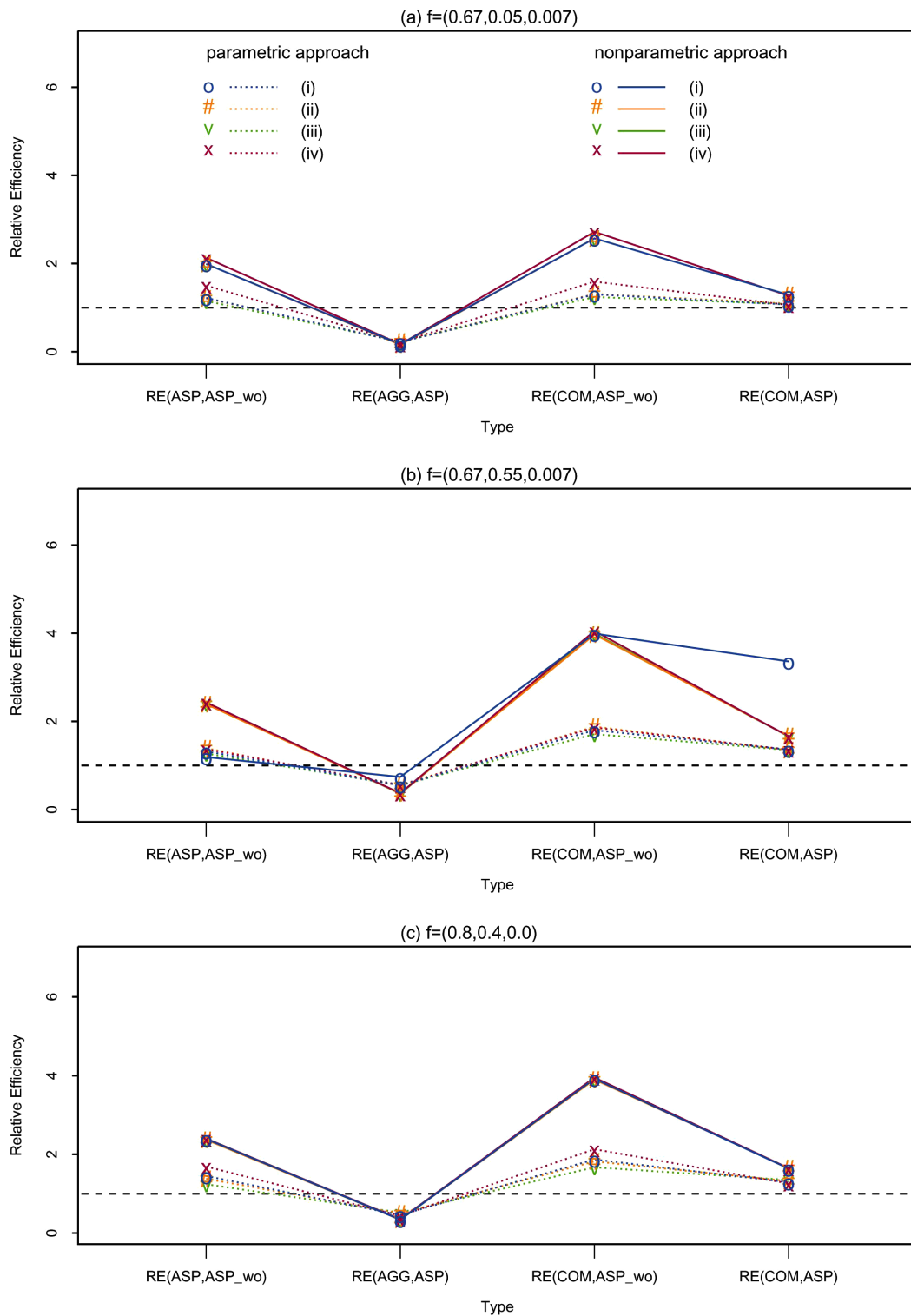
$$\boldsymbol{\Sigma}_i = \begin{bmatrix} 0.8 & 0.16 & \dots & 0.16 \\ 0.16 & 0.8 & 0.16 & \vdots \\ \vdots & \vdots & \ddots & 0.16 \\ 0.16 & \dots & 0.16 & 0.8 \end{bmatrix}_{n_i \times n_i}; \text{ and } \mathbf{g}_i \text{ is a vec-}$$

tor of genotypic effects of the QTL. The genotypic effects are 2, 0 and -2 for the genotypes of two high-risk alleles, one high-risk together with one low-risk allele and two low-risk alleles, respectively.

#### Two disease locus model

Furthermore, we simulated a two-locus disease model and compared the estimates of  $\tau_1$  and  $\tau_2$  from approaches with and without incorporating a covariate. We generated the two-locus models of model B in Bieracka et al. [20] as described in Additional file 3, Table S2 to study the impact of covariates on the estimates from the without-a-covariate approach and parametric and non-parametric with-a-covariate approaches.

For genotype data, we generated ten markers that were equally spaced at 10 cM between adjacent markers, with each marker having eight equal-frequency alleles, and the two diallelic disease loci were located at 35 and 75 cM. For scenarios (1), (2) and (3), an additive genetic model for the quantitative trait covariate was assumed. The covariate used in modeling  $C_l$  was denoted by  $y_l$ , with  $l = 1, 2$ . Assuming the quantitative traits  $X_{QTL1}$  and  $X_{QTL2}$  were controlled by  $\tau_1, \tau_2$  respectively, we examined the impact of different combinations of traits incorporated in functions of  $g_{lk}$  on estimating the two trait loci. As in the simulation for the one-locus model, four scenarios were considered for the QTL of each covariate: (1) The QTL is at 35 cM ( $\tau_1$ ) (pleiotropic effect); (2) the QTL for "age at onset" (covariate) is at 35 cM ( $\tau_1$ ); (3) the quantitative trait's QTL is at 45 cM (coincident effect); (4) the covariate's QTL is not linked to either disease locus. All



**Figure 1 Relative efficiency (RE) between two approaches in estimating the disease locus under three genetic models (a), (b) and (c).** The dotted lines are the RE for comparisons between two types of affected relative pairs in the non-parametric approach. The solid lines are the RE for comparisons between two types of affected relative pairs in the parametric approach. ASP, AGG and COM stand for affected sib pairs, affected grandparent-grandchild, and combined affected sib pairs and grandparent-grandchild pairs, respectively. ASP\_wo stands for using ASP without incorporating a covariate. The circle, pund, v and x signs refer to the relationship between the covariate's QTL and the disease locus, including (i) pleiotropy, (ii) co-incident, (iii) unlinked, and (iv) a covariate of age at onset.

covariates were determined by averaging the two individuals' covariate values in one pair, that is,  $g_{ki} = (x_{ki1} + x_{ki2})/2$ .

## Results

For the comparison under one-locus models (Figure 1, Additional file 4, Tables S3 - S5), the efficiency in estimating the disease locus was enhanced substantially when incorporating a quantitative covariate, regardless of its underlying genetic mechanisms. In the additive model using affected sibpairs, the relative efficiency (RE) ranged from 1.24 to 1.69 for the parametric approach and from 2.37 to 2.40 for the non-parametric approach. After adding affected grandparent-grandchild pairs, the RE ranged increase to 3.9-3.95 for the parametric approach and 1.67-2.13 for the non-parametric approach. The parametric approach generally had higher RE than the non-parametric approach in the simulated scenarios (Additional file 4, Tables S3 - S5). Given the same heritability of a quantitative trait, incorporating a quantitative trait with a pleiotropic effect was generally more efficient than when incorporating a linked or an unlinked trait. The variance estimate for  $\tau$  in the one-locus models was generally smaller in the parametric approach than that found in the non-parametric approach under the same scenarios. As expected, with the same sample size, the efficiency in estimating the disease locus was always higher when using affected sibpairs than when using grandparent-grandchild pairs. The efficiency in estimating the disease locus was always improved when combining both relative pairs. The 95% coverage probabilities for the disease locus were almost always slightly underestimated, as most of the variance estimates tended to be underestimated.

The smoothing parameter in (3) was set to one half of the range of the covariates, which roughly minimizes the variance estimate of the estimated loci in the analysis. The choice of bandwidth in the non-parametric approach did not have much impact on the estimation though [18]. The selection of function  $g(\cdot)$  might slightly influence bias and variance of the estimates for disease loci (these results not shown here). Results from both parametric and non-parametric approaches suggested that the efficiency in estimating disease locus was improved when combining affected sib pairs and grandparent-grandchild pairs.

Since there were two linked loci controlling the disease, we generated covariates  $X_{QTL1}$  and  $X_{QTL2}$ , controlled by  $\tau_1$  and  $\tau_2$ , respectively, and studied the impact of four different ways to incorporate  $X_{QTL1}$  or  $X_{QTL2}$  into the linkage mapping: (i) incorporating  $X_{QTL1}$  only ( $y_1 = X_{QTL1}$ ,  $y_2 = X_{QTL1}$ ); (ii) incorporating  $X_{QTL2}$  only ( $y_1 = X_{QTL2}$ ,  $y_2 = X_{QTL2}$ ); (iii) incorporating  $y_1 = X_{QTL1}$ ,  $y_2 = X_{QTL2}$  to estimate  $C_1$ ,  $C_2$ , respectively; (iv)

incorporating  $y_1 = X_{QTL2}$ ,  $y_2 = X_{QTL1}$ , to estimate  $C_1$ ,  $C_2$ , respectively. Tables 1 illustrates the impact of choosing different covariates on estimates by parametric and non-parametric approaches, respectively. In reality, we do not have information about the underlying genetic mechanism of the quantitative traits (covariates); luckily, the efficiency in estimating the disease loci was improved under any one of the above scenarios when compared to the estimates made without covariates. Since the quantitative traits were controlled by the two disease loci, incorporating both quantitative traits was helpful in estimating both loci and their 95% coverage probabilities. When incorporating only one quantitative trait, the bias and variance estimate for its corresponding disease locus (QTL) were smaller; this finding was particularly true within the parametric approach. Additionally, both of the covariates were significantly associated with the genetic effects from the two disease loci in the parametric approach ( $p$ -values = 0.029 ~ 0.050).

We also evaluated the performance of the parametric and non-parametric approaches with varying locations for covariates' QTLs (Table 2). In the parametric approach, the efficiency in estimating a disease locus was improved when the set location of the covariate's QTL was linked to the disease locus, particularly when the disease locus was also the QTL of the covariate. For example, when no covariate was incorporated, the variance estimates were 7.5 and 6.9 for the two disease loci, respectively (Additional file 5, Table S4); when a quantitative trait with a pleiotropic effect was incorporated, the variance estimates were 4.0 and 4.0 respectively (Table 2). Compared with the estimate without incorporating a covariate, the bias was slightly higher than when the covariate's locus was not the disease locus but was instead linked or unlinked to the disease locus. The biases for estimating the two loci were -0.02 and -0.2 with the pleiotropic covariate and 0.3 and -0.4 with the unlinked covariate (Table 2). In the parametric approach, the magnitude of the regression coefficient reflects the association between the disease locus and the covariate. The regression coefficient was significant only when the covariate's QTL was one of the disease loci (pleiotropy effect) (Table 2). After incorporating a covariate, the 95% coverage probabilities for  $\tau_1$  and  $\tau_2$  were both more precise than those obtained without incorporating a covariate (Tables 1 and 2; Additional file 5, Table S6). In the non-parametric approach, the efficiency in estimating both disease loci was improved when the covariate's QTL was at position  $\tau_1$  (Table 2; pleiotropic covariate or age at onset). The efficiency was lower when the covariate's QTL was linked or unlinked to position  $\tau_1$  (Tables 2). The bias was generally higher for  $\tau_2$  in the scenario where the covariate provides information for  $\tau_1$  only (Tables 2).



**Table 2 The impact of the location of the QTL for the covariate - parametric and nonparametric approaches**

	Disease Loci (cM)				Estimate of C								95% coverage probability (%)				The Location of the Covariate's QTL
	Parametric		Nonparametric		Parametric				Nonparametric				Parametric		Nonparametric		
	$\tau_1$	$\tau_2$	$\tau_1$	$\tau_2$	ASP		AGP		ASP		AGP		$\tau_1$	$\tau_2$	$\tau_1$	$\tau_2$	
					C <sub>11</sub>	C <sub>21</sub>	C <sub>14</sub>	C <sub>24</sub>	C <sub>11</sub>	C <sub>21</sub>	C <sub>14</sub>	C <sub>24</sub>					
Bias	-0.02	-0.2	-0.1	1.0	0.04	0.03	-0.05	0.02	-0.02	-0.04	-0.04	0.02	95	96	96	93	
Sample variance	4.4	3.7	4.7	5.9	0.003	0.003	0.001	0.001	0.002	0.002	0.001	0.001					
Mean variance	4.0	4.0	4.8	5.6													
$\check{\beta}_1$					0.26	-0.25	0.16	-0.08									
p-value					0.03	0.05	0.52	0.82									
Bias	0.2	-0.03	0.4	1.8	0.03	0.03	-0.05	0.02	-0.007	-0.06	-0.03	0.02	95	96	93	88	Age onset at 35 cM ( $\tau_1$ )
Sample variance	4.4	3.9	5.2	5.9	0.003	0.003	0.001	0.001	0.002	0.002	0.001	0.001					
Mean variance	4.1	4.1	4.4	6.0													
$\check{\beta}_1$					-0.04	0.04	-0.03	0.01									
p-value					0.05	0.06	0.54	0.83									
Bias	0.3	-0.3	-0.1	0.7	0.06	0.05	-0.05	0.02	-0.03	-0.04	-0.05	0.02	95	97	95	95	Co-incident 45 cM
Sample variance	6.8	5.7	9.1	8.5	0.003	0.003	0.001	0.001	0.002	0.002	0.001	0.001					
Mean variance	6.7	6.3	8.9	9.0													
$\check{\beta}_1$					-0.007	0.003	-0.002	0.006									
p-value					0.96	0.95	0.94	0.95									
Bias	0.3	-0.4	-0.8	0.6	0.003	0.05	-0.05	0.02	-0.05	-0.04	-0.06	0.010	96	96	94	95	Unlinked
Sample variance	6.8	5.5	9.6	8.6	0.005	0.003	0.001	0.001	0.002	0.002	0.001	0.001					
Mean variance	6.7	6.3	10.3	9.3													
$\check{\beta}_1$					0.96	0.002	0.006	-0.001									
p-value					0.96	0.96	0.92	0.93									

**A Data Example**

We conducted an autosome-wide scan for affected relative pairs from the COGA data [23]. Note that the disease was defined as “having psychological problems from drinking.” There are 149 affected sib pairs, 8 half sib pairs, 16 first-cousins pairs, 7 grandparent-grandchild pairs, and 71 avuncular pairs in this data set. Due to the limited sample sizes for some relative pairs, we examined the linkage peak on chromosome 1 using 149 affected sib pairs and 71 avuncular pairs, with and without incorporating the quantitative covariate “Maximum number of drinks in a 24 hour period.” Using both ASPs and AGPs, the disease locus was estimated to be

at 113.7 cM on chromosome 1 with a 95% CI: 109.5-118.0 cM. The estimate for C<sub>ASP</sub> was 0.18 with a 95% CI from 0.10-0.26 (p-value = 7.6e-6), whereas the estimate for C<sub>AAP</sub> was 0.064 with a 95% CI from -0.0001-0.13 (p = 0.051) (Table 3 and Additional file 6, Figure S1). We also applied single locus with covariate linkage mapping using ARP to locate the disease locus and assess the significance of its covariates. The disease locus estimate was 110.8 (standard error (SE) = 1.5) and 109.2 (SE = 2.3) cM in the parametric and non-parametric approaches, respectively, using all ARPs. The p-values of the covariate in the parametric approach are 0.52 and 0.20 for ASP and AAP, respectively (Table 3).

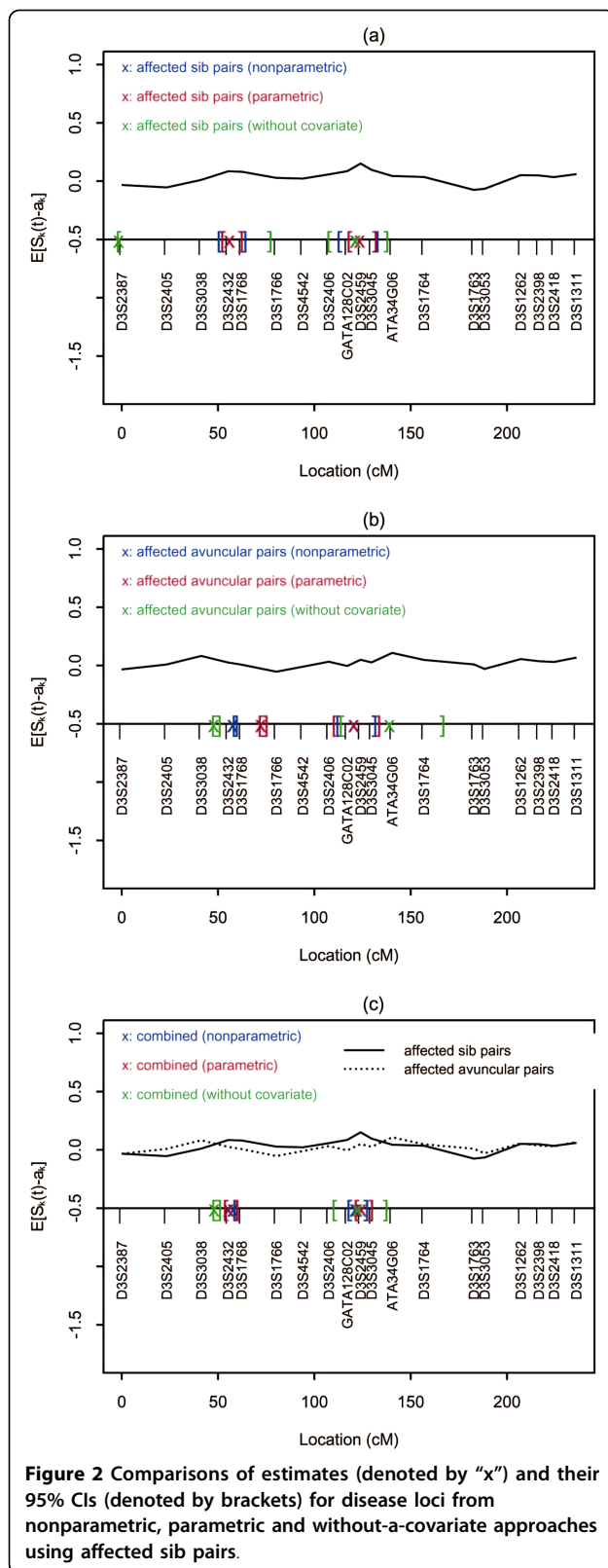


**Table 3 One-locus search on chromosome 1 with or without incorporation of “Maximum number of drinks in a 24 hour period”**

	ESTIMATE [95% CI] or (S.E.)								
	$\tau$ (cM)			C			$\lambda$		
	Without a covariate	Parametric	Nonparametric	Without a covariate	Parametric	Nonparametric	Without a covariate	Parametric	Nonparametric
Using one ARP only:									
Full siblings	112.9 (6.1)	112.8 (6.5)	110.4 (7.2)	0.18 [0.04, 0.32]	0.16 [0.001,0.32]	0.14 [0.06,0.23]	1.75 [1.14,2.84]	1.65 [1.00,2.77]	1.47 [1.13,1.87]
p-value for the covariate	0.58								
Avuncular pairs	98.8 (12.4)	105.0 (7.8)	102.8 (6.2)	0.08 [-0.07, 0.24]	0.20 [-0.06,0.28]	0.036 [-0.10,0.23]	1.46 [0.81,2.50]	1.74 [0.79,3.56]	1.37 [0.67,2.67]
p-value for the covariate	0.23								
Using both ARPs:									
Full siblings				0.18 [0.10, 0.26]	0.17 [0.009,0.32]	0.14 [0.045,0.23]	1.70 [1.08,2.74]	1.61 [1.02,2.70]	1.44 [1.10,1.87]
p-value for the covariate	0.52								
Avuncular pairs				0.064 [-0.0001, 0.13]	0.10 [-0.04,0.28]	0.034 [-0.07,0.18]	1.66 [0.69,3.72]	1.77 [0.85,3.55]	1.28 [0.74,2.09]
p-value for the covariate	0.20								
Common $\tau$	113.7 (2.2)	110.8 (1.5)	109.2 (2.3)						

To identify a region harboring two disease loci, we plotted the empirical IBD sharing of all autosomes for ASP (because the data set included mostly sib pairs). After visually reviewing all the empirical IBD sharing on autosomes, we selected chromosome 3 as a region to illustrate our approach, as there appeared to be two disease-susceptibility loci harbored within this region (Figure 2). First, we conducted the two-locus search without incorporating the covariate (Table 4) and compared the estimates to those that did incorporate covariates. The quantitative measure “maximum number of drinks in a 24-hour period” [24] was incorporated into the linkage mapping, both parametrically (Table 5) and non-parametrically (Table 6). The 95% confidence intervals (CIs) for C or  $\lambda$  were constructed with the bootstrap re-sampling approach. A total of 1,000 replicates were obtained by re-sampling. The disease loci estimates were computed for each sample and ranked. The lower and upper limits of the 95% confidence interval were the 2.5% and 97.5% percentiles of the 1,000 replicates, respectively.

The standard errors for the estimates of the disease loci were always smaller when using the entire data set with both sibpairs and avuncular pairs, compared to the estimates using sib pairs or avuncular pairs alone. Compared to the approach without the covariate, the relative efficiencies (each defined as the ratio of reversed variance estimates for the disease locus estimates) in estimating  $\tau_1$  and  $\tau_2$  are 20.25  $((0.7/0.2)^2)$  and 8.92  $((6.84/2.29)^2)$  for the non-parametric approach (Table 6) and 0.24  $((0.72/1.47)^2)$  and 11.8  $((6.84/1.99)^2)$  for the parametric approach (Table 5). The average estimated  $C_1$  and  $C_2$  were 0.084 and 0.16 for affected sibpairs in the non-parametric approach (Table 6), and were 0.16 and 0.24 in the parametric approach (Table 5). The corresponding risk ratios  $\lambda_l$  for these two loci in sib pairs within the non-parametric approach were 1.20 (95% CI: 0.99 to 1.79) and 1.45 (95% CI: 1.02 to 2.09), respectively (Table 6). The C value (or risk ratio) at  $\tau_2$  (0.237, 95% CI: 0.066 to 0.430) was higher than that at  $\tau_1$  (0.156, 95% CI: -0.014 to 0.319), and it was marginally significant after incorporation of the covariate (Table 5).



The  $C_l$  and  $\lambda_l$  values estimated from avuncular pairs were smaller than those estimated from sib pairs (Tables 4, 5, 6) with incorporation of the covariate; however, this difference was not statistically significant. Since there was no evidence of linkage at  $\tau_1$ , the estimate for  $\tau_1$  varied in the three approaches.

### Discussion and Conclusions

Many complex diseases involve multiple loci as well as multiple quantitative biological markers or quantitative risk factors. Incorporating covariates into linkage analysis is not only helpful for the identification of disease loci but is also informative with respect to disease etiology. In family-based studies, data are often available for larger pedigrees with multiple relative pairs, and therefore it is desirable to have linkage mapping approaches that can use these potentially informative data. In addition, different types of ARPs may have the potential of providing some insight into the underlying genetic mechanism [19]. Applying a one-locus model to localize a disease gene when there are actually two linked disease genes in the region is likely to estimate the two true disease gene locations inaccurately, while the corresponding effect size tends to be over-estimated [20]. Therefore, we extended a robust multipoint linkage approach in simultaneously mapping two linked disease loci while using affected relative pairs with an incorporation of quantitative covariates. A series of intensive simulation studies were conducted to examine the performance of the approach when the incorporated covariate was a quantitative trait under a variety of genetic models or when the trait was a risk factor associated with a disease locus. The simulation study suggested that incorporating a quantitative covariate, which also happened to be a quantitative trait, helped improve the efficiency of the disease-locus estimate, regardless of the genetic models that actually underlie the incorporated covariate. It seems that the underlying genetic models of the quantitative covariate (trait) did not have much impact on the efficiency in estimating  $\tau_l$ ,  $l = 1, 2$ . In addition, the inclusion of different relative pairs would make the sample size larger and improve the efficiency of the disease-locus localization when the different relative pairs share common disease loci; this would be particularly true when the genetic effect of the disease loci is small or modest. When the covariate was directly related to the liability of the disease, the efficiency improvement was greater than when it was not directly related to the disease liability; when the covariate was associated with only one disease locus, incorporating the covariate helped improve the efficiency of that locus' estimate more than

**Table 4 Simultaneous two-locus search without incorporating a covariate**

	$\tau_1$ (cM)	$\tau_2$ (cM)	ESTIMATE (S.E.) or [95% CI]			
			$C_1$	$C_2$	$\lambda_1$	$\lambda_2$
Using one ARP only:						
Full siblings	1.38 (39.44)	124.27 (7.64)	-0.04 [-0.17,0.10]	0.12 [-0.008,0.25]	1.25 [0.83,1.89]	1.44 [0.91,2.13]
Avuncular pairs	50.77 (0.73)	142.05 (13.42)	0.11 [-0.07,0.29]	0.097 [-0.03,0.22]	1.17 [0.80,1.69]	1.36 [0.83,1.95]
Using both ARPs:						
Full siblings			0.06 [-0.07,0.18]	0.12 [-0.01,0.24]	1.17 [0.86,1.70]	1.40 [0.90,2.08]
Avuncular pairs			0.11 [-0.07,0.29]	0.060 [-0.034,0.154]	1.30 [0.55,2.40]	1.45 [0.68,2.28]
Common $\tau$	50.98 (0.72)	125.43 (6.84)				

**Table 5 Simultaneous two-locus search with incorporation of "Maximum number of drinks in a 24 hour period" - parametric approach**

	$\tau_1$ (cM)	$\tau_2$ (cM)	ESTIMATE (S.E.) or [95% CI]			
			$C_1$	$C_2$	$\lambda_1$	$\lambda_2$
Using one ARP only:						
Full siblings	58.95 (2.42)	126.55 (3.43)	0.16 [-0.01,0.32]	0.24 [0.07,0.43]	1.45 [0.91,2.76]	1.90 [1.11,7.18]
Avuncular pairs	75.14 (0.72)	123.55 (5.85)	-0.07 [-0.15,0.21]	0.04 [-0.15,0.24]	0.76 [0.54,2.40]	1.17 [0.65,2.07]
Using both ARPs:						
Full siblings			0.16 [-0.08,0.32]	0.23 [0.02,0.43]	1.46 [0.86,2.70]	1.82 [1.03,6.68]
Avuncular pairs			0.005 [-0.15,0.20]	0.04 [-0.21,0.23]	1.02 [0.54,2.38]	1.16 [0.40,2.65]
Common $\tau$	58.53 (1.47)	127.41 (1.99)				

**Table 6 Simultaneous two-locus search with incorporation of "Maximum number of drinks in a 24 hour period" - nonparametric approach**

	$\tau_1$ (cM)	$\tau_2$ (cM)	ESTIMATE (S.E.) or [95% CI]			
			$C_1$	$C_2$	$\lambda_1$	$\lambda_2$
Using one ARP only:						
Full siblings	58.97 (3.37)	124.42 (4.99)	0.084 [-0.03,0.21]	0.16 [-0.004,0.27]	1.20 [0.94,1.71]	1.45 [0.99,2.21]
Avuncular pairs	60.66 (0.24)	123.46 (4.84)	0.018 [-0.11,0.11]	0.048 [-0.081,0.19]	1.07 [0.64,1.58]	1.21 [0.77,1.69]
Using both ARPs:						
Full siblings			0.083 [-0.005,0.22]	0.16 [0.011,0.26]	1.20 [0.99,1.79]	1.45 [1.02,2.09]
Avuncular pairs			0.017 [-0.11,0.12]	0.051 [-0.052,0.18]	1.07 [0.63,1.62]	1.23 [0.81,2.10]
Common $\tau$	60.81 (0.16)	124.24 (2.29)				

that of the other locus. The position of the QTL for a quantitative trait (as a covariate) might slightly affect the accuracy of the disease-loci localization; the accuracy was similar to the situation in which no covariates were incorporated given an unlinked relationship between the QTL and disease locus. Investigators can choose to incorporate covariates that improve efficiency in disease-loci estimation. Our example of an alcoholism study illustrates that incorporating a quantitative covariate into the linkage mapping helps improve the efficiency of disease-loci estimates in the two-locus models by either the parametric approach or the nonparametric approach. The assessment of associations between the disease loci and covariates helps resolve the underlying genetic mechanism of the disease. Using all affected relative pairs to estimate the common disease loci could also enhance the efficiency in estimating disease loci, and, furthermore, it could help dissect disease etiology by assessing risk ratios among different types of relative pairs.

Although the proposed approaches can be quite helpful and can also be widely applied to localize disease loci for complex diseases, they are built upon the assumption of a two-locus disease mechanism. Bias may arise when a region harboring one locus only or more than two linked loci is examined. In addition, since the relationships between the genetic effects on the two disease loci and covariates are modeled separately, the number of parameters may easily be increased when (1) several covariates are incorporated simultaneously; or (2) regression relationships between the genetic effects on the two disease loci and covariates are not assumed to be identical; or (3) several relative types are analyzed. Additionally, since fitting an incorrect model can lead to biased estimates with anti-conservative confidence intervals, it is important to decide whether a one-locus or two-locus model is more appropriate. In practice, it is always helpful to check the empirical plot (as shown in Figure 2) to determine how many "peaks" are present in the region of interest. If there is only one "peak," a one-locus model might be more appropriate than a two-locus model. If more than two peaks are present, it might be helpful to split the region into multiple smaller regions containing only two peaks each. Indeed, it is always helpful to apply both one-locus and two-locus models and evaluate which model fits the data better. In addition, the test developed by Biernacka et al. [21] can be used to help choose an appropriate model.

The proposed approaches allow gene-gene and gene-environment interactions to be assessed. As complex diseases often involve more than two disease genes, further efforts to extend this method to situations involving more than two genes are warranted. In addition, as the regions identified through linkage mapping are quite wide and may harbor numerous genes, future approaches

should be developed to identify potential causal polymorphisms by the joint modeling of linkage and association.

## Additional material

**Additional file 1: Table S1.** Expected alleles shared IBD at location  $t$  for five types of ARPs and functions relating  $\lambda$  to  $C$

**Additional file 2: Appendix.** Theoretical derivations.

**Additional file 3: Table S2.** The two-locus genetic model used in simulation studies.

**Additional file 4: Tables S3-S5.** Table S3. Comparisons of estimate for  $\tau$  with and without incorporation of a quantitative covariate under one-locus recessive model (a). Table S4. Comparisons of estimate for  $\tau$  with and without incorporation of a quantitative covariate under one-locus dominant model (b). Table S5. Comparisons of estimate for  $\tau$  with and without incorporation of a quantitative covariate under one-locus additive model (c).

**Additional file 5: Table S6.** Simultaneous two-locus search without incorporating a covariate.

**Additional file 6: Figure S1.** Comparisons of estimates (denoted by "x") and their 95% CIs (denoted by brackets) for the disease locus on chromosome one from nonparametric, parametric and without-a-covariate approaches using affected sib pairs

## Acknowledgements

We thank the data provided by the Collaborative Study on the Genetics of Alcoholism (U10AA008401). We thank the reviewers for their constructive comments, which greatly improved the quality of this manuscript. This work was supported by grant GRC 94B001-1 to J.M.C. from Academia Sinica; and, in part, by grants PH-098-pp04 and NSC98-2118-M-400-002 to Y.F.C. from National Health Research Institutes and National Science Council respectively; and a grant to K.Y.L. from National Institutes of Health, U.S.A. (HL090577).

## Author details

<sup>1</sup>Division of Biostatistics and Bioinformatics, Institute of Population Health Sciences, National Health Research Institutes, 35 Keyan Rd., Zhunan, Miaoli 350, Taiwan. <sup>2</sup>Institute of Statistical Science, Academia Sinica, Taipei, Taipei, Taiwan. <sup>3</sup>Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, USA. <sup>4</sup>National Yang-Ming University, Taipei, Taiwan.

## Authors' contributions

YFC, JMC and KYL have made contributions to the theory derivation, simulation study, statistical modeling and draft of the manuscript. CYL participated in the design of the study and performed the simulation studies and data analysis. All authors read and approved the final manuscript.

Received: 2 February 2010 Accepted: 14 July 2010

Published: 14 July 2010

## References

1. Roeder K, Bacanu SA, Wasserman L, Devlin B: Using linkage genome scans to improve power of association in genome scans. *American Journal of Human Genetics* 2006, **78**:243-252.
2. Clerget-Darpoux F, Elston RC: Are linkage analysis and the collection of family data dead? Prospects for family studies in the age of genome-wide association. *Hum Hered* 2007, **64**(2):91-96.
3. Goddard KA, Witte JS, Suarez BK, Catalona WJ, Olson JM: Model-free linkage analysis with covariates confirms linkage of prostate cancer to chromosomes 1 and 4. *Am J Hum Genet* 2001, **68**(5):1197-1206.
4. Gauderman WJ, Siegmund KD: Gene-environment interaction and affected sib pair linkage analysis. *Hum Hered* 2001, **52**(1):34-46.

5. Devlin B, Jones BL, Bacanu SA, Roeder K: **Mixture models for linkage analysis of affected sibling pairs and covariates.** *Genet Epidemiol* 2002, **22**(1):52-65.
6. Greenwood CM, Bull SB: **Incorporation of covariates into genome scanning using sib-pair analysis in bipolar affective disorder.** *Genet Epidemiol* 1997, **14**(6):635-640.
7. Rice JP, Rochberg N, Neuman RJ, Saccone NL, Liu KY, Zhang X, Culverhouse R: **Covariates in linkage analysis.** *Genet Epidemiol* 1999, **17**(Suppl 1):S691-695.
8. Greenwood CM, Bull SB: **Analysis of affected sib pairs, with covariates—with and without constraints.** *Am J Hum Genet* 1999, **64**(3):871-885.
9. Saccone NL, Rochberg N, Neuman RJ, Rice JP: **Covariates in linkage analysis using sibling and cousin pairs.** *Genet Epidemiol* 2001, **21**(Suppl 1):S540-545.
10. Olson JM: **A general conditional-logistic model for affected-relative-pair linkage studies.** *Am J Hum Genet* 1999, **65**(6):1760-1769.
11. Hauser ER, Watanabe RM, Duren WL, Bass MP, Langefeld CD, Boehnke M: **Ordered subset analysis in genetic linkage mapping of complex traits.** *Genet Epidemiol* 2004, **27**(1):53-63.
12. Farrall M: **Affected sibpair linkage tests for multiple linked susceptibility genes.** *Genet Epidemiol* 1997, **14**(2):103-115.
13. Delepine M, Pociot F, Habita C, Hashimoto L, Froguel P, Rotter J, Cambon-Thomsen A, Deschamps I, Djoulah S, Weissenbach J, et al: **Evidence of a non-MHC susceptibility locus in type I diabetes linked to HLA on chromosome 6.** *Am J Hum Genet* 1997, **60**(1):174-187.
14. Cordell HJ, Wedig GC, Jacobs KB, Elston RC: **Multilocus linkage tests based on affected relative pairs.** *Am J Hum Genet* 2000, **66**(4):1273-1286.
15. Biswas S, Papachristou C, Irwin ME, Lin S: **Linkage analysis of the simulated data - evaluations and comparisons of methods.** *BMC Genet* 2003, **4**(Suppl 1):S70.
16. Liang KY, Chiu YF, Beaty TH: **A robust identity-by-descent procedure using affected sib pairs: multipoint mapping for complex diseases.** *Hum Hered* 2001, **51**(1-2):64-78.
17. Glidden DV, Liang KY, Chiu YF, Pulver AE: **Multipoint affected sibpair linkage methods for localizing susceptibility genes of complex diseases.** *Genet Epidemiol* 2003, **24**(2):107-117.
18. Chiou JM, Liang KY, Chiu YF: **Multipoint linkage mapping using sibpairs: non-parametric estimation of trait effects with quantitative covariates.** *Genet Epidemiol* 2005, **28**(1):58-69.
19. Schaid DJ, Sinnwell JP, Thibodeau SN: **Robust multipoint identical-by-descent mapping for affected relative pairs.** *Am J Hum Genet* 2005, **76**(1):128-138.
20. Biernacka JM, Sun L, Bull SB: **Simultaneous localization of two linked disease susceptibility genes.** *Genet Epidemiol* 2005, **28**(1):33-47.
21. Biernacka JM, Cordell HJ: **Exploring causality via identification of SNPs or haplotypes responsible for a linkage signal.** *Genet Epidemiol* 2007, **31**(7):727-740.
22. Lin WY, Schaid DJ: **Robust multipoint simultaneous identical-by-descent mapping for two linked loci.** *Hum Hered* 2007, **63**(1):35-46.
23. Edenberg HJ, Bierut LJ, Boyce P, Cao M, Cawley S, Chiles R, Doherty KF, Hansen M, Hinrichs T, Jones K, et al: **Description of the data from the Collaborative Study on the Genetics of Alcoholism (COGA) and single-nucleotide polymorphism genotyping for Genetic Analysis Workshop 14.** *BMC Genet* 2005, **6**(Suppl 1):S2.
24. Bagnardi V, Zatonski W, Scotti L, La Vecchia C, Corrao G: **Does drinking pattern modify the effect of alcohol on the risk of coronary heart disease? Evidence from a meta-analysis.** *Journal of Epidemiology and Community Health* 2008, **62**(7):615-619.

doi:10.1186/1471-2156-11-67

**Cite this article as:** Chiu et al.: Incorporation of covariates in simultaneous localization of two linked loci using affected relative pairs. *BMC Genetics* 2010 **11**:67.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

