

The Value of Relatives With Phenotypes But Missing Genotypes in Association Studies for Quantitative Traits

Peter M. Visscher* and David L. Duffy

Genetic Epidemiology, Queensland Institute of Medical Research, Brisbane, Australia

The additional statistical power of association studies for quantitative traits was derived when ungenotyped relatives with phenotypes are included in the analysis. It was shown that the extra power is a simple function of the coefficient of additive genetic relationship and the phenotypic correlation coefficient between the genotyped and ungenotyped relatives. For close relatives, such as pairs of fullsibs and identical twin pairs, gains in power in the range of 10 to 30% are achieved if only one of the pair is genotyped. The theoretical results were verified by simulations. It was shown that ignoring the error in estimating the genotype of the ungenotyped relative has little impact on the estimates and on statistical power, consistent with results from quantitative trait loci (QTL) linkage studies. For genome-wide association studies in which not all relatives with phenotypes can be genotyped, our study provides a prediction of the additional power of an analysis that includes phenotypes on ungenotyped individuals, and can be used in experimental design. We show that a two-step procedure, in which missing genotypes are imputed and subsequently an association analysis is performed, is efficient and powerful. *Genet. Epidemiol.* 30:30–36, 2006. © 2005 Wiley-Liss, Inc.

Key words: association; quantitative trait; twins; statistical power

Contract grant sponsor: NIAAA; Contract grant number: AA07728; Contract grant sponsor: European Commission (GenomEUtwin).

*Correspondence to: Dr. Peter Visscher, Queensland Institute of Medical Research, Brisbane 4029, Australia.

E-mail: peter.visscher@qimr.edu.au

Received 17 July 2005; Accepted 24 September 2005

Published online 14 December 2005 in Wiley InterScience (www.interscience.wiley.com).

DOI: 10.1002/gepi.20124

INTRODUCTION

With the advent of high-density genotyping platforms, the era of genome-wide association studies has arrived. In addition to genome-wide association studies for (binary) disease traits, in, for example, a case-control setting, investigators will also perform whole genome association studies for quantitative traits. Many researchers have measured quantitative phenotypes on a large number of individuals for genetic linkage studies, in nuclear families, deep pedigrees, and twin families, and it makes sense to use the same individuals for genome-wide association studies. However, the cost of whole genome association analysis is such that usually only a subset of those individuals with quantitative phenotypes can be genotyped. When there are more individuals with phenotypes than individuals that can be genotyped, there are a number of questions of scientific interest: (1) which individuals should be selected for genotyping, (2) how should the data be

analysed when some genotypes may be missing, and (3) what is the value, in terms of additional statistical power, of the relatives that have phenotypes but have not been genotyped. The present note deals mainly with the third question, but has implications for the first two.

The analysis of gene–disease association using family data with missing genotypes on relatives has been dealt with extensively in the literature [Boerwinkle et al., 1986; Hasstedt and Moll, 1989; Li and Thompson, 1997; Vitezica et al., 2005; Wacholder et al., 1998]. In likelihood based analysis approaches, the missing genotypes are modeled explicitly, and inference on association takes into account the most likely genotypes of ungenotyped individuals. The information on the likely genotypes of the ungenotyped individuals comes from three sources: the distribution of genotypes in the population, the observed genotypes on relatives, and the phenotypes of relatives and the ungenotyped individuals.

For quantitative traits, regression and likelihood methods in which putative trait loci are explicitly modelled in terms of allele frequency and penetrance have been developed for “major genes” [Hasstedt and Moll, 1989; Kerr and Kinghorn, 1996; Kinghorn et al., 1993] but are not widely used in practice because the underlying quantitative trait loci (QTL) are likely to explain only a small proportion of the total variation, so that the phenotype is not informative to predict the genotype. For QTL, inference on the genotypes of ungenotyped individuals has been reported using likelihood and Markov Chain Monte Carlo (MCMC) segregation analysis methods [Hasstedt and Moll, 1989; Kerr and Kinghorn, 1996; Kinghorn et al., 1993] and MCMC methods that infer the missing genotypes only from genotypes of relatives [Hernandez-Sanchez et al., 2003]. For the actual association analysis, the inferred genotypes (or genotype probabilities) are usually treated as if they were known without error. This is, for example, how the computer program Mendel performs an association analysis when there are relatives with phenotypes but no genotypes [Lange et al., 1976, 2005]. A number of computer programs can estimate genotype probabilities from observed genotypes on relatives [Hasstedt, 2002; Lange et al., 2005].

In this study, we quantify the value, in terms of statistical power, of having relatives with measured phenotypes. The study is performed assuming that pairs of relatives with quantitative phenotypes are available and fitting an additive genetic association model, but the results are relevant for arbitrary pedigrees and non-additive models.

METHODS

GENOTYPES

Ungenotyped individuals with phenotypes can provide information for association because their genotypes can be predicted from the observed genotypes of relatives. We consider pairs of relatives, both of whom have been phenotyped but only one of whom has been genotyped. How much extra information is provided for association depends on two key population parameters that describe the genetic and phenotypic relationship between individuals i and j : the additive genetic coefficient of relationship (r_{ij}) and the phenotypic correlation coefficient (ρ_{ij}). The additive coefficient of relationship is twice the

coefficient of kinship. Where possible we drop the subscript ij .

We consider a bi-allelic locus (for example, a SNP) with genotypes BB, Bb, and bb that have effects of $-a$, d , and a on the phenotype, respectively. Genotypes are assumed to be in Hardy-Weinberg equilibrium (HWE), with genotype frequencies of p^2 , $2p(1-p)$ and $(1-p)^2$, respectively. We define an indicator variable, x , to denote an individual's measured genotype, with values of 1, 0, and -1 for genotypes BB, Bb, and bb, respectively. A model to estimate and test the additive effect of the genotype on a quantitative phenotype is to regress the phenotypes on the indicator variables.

Although a phenotype of the ungenotyped relative is available, its genotype has not been measured. However, the genotype, and hence the indicator variable, of the ungenotyped relative can be predicted when the genotype proportions are in HWE. At the locus under consideration, the additive genetic variance ($\text{var}(A)$) among the ungenotyped individuals can be partitioned as

$$\text{var}(A_j) = r_{ij}^2 \text{var}(A_j) + (1 - r_{ij}^2) \text{var}(A_j).$$

The first of these terms is the variance due to the observed genotype of the genotyped relative, i.e., $\text{var}(\hat{A}_j) = r_{ij}^2 \text{var}(A)$, the QTL variance among the ungenotyped individuals that is captured by the observed genotypes of their relatives. Similarly,

$$\text{var}(E(x_j|x_i)) = \text{var}(\hat{x}_j) = r_{ij}^2 \text{var}(x_i).$$

Hence, a proportion of r^2 , e.g., $1/4$ for fullsibs, of the additive genetic variance at the locus under consideration among the ungenotyped individuals can be “recovered” from the measured genotypes of their relatives. The conditional means of the genotypic values among the ungenotyped relatives is easily derived from the allele frequency and genotype probabilities at the locus. We derived these means for fullsibs, and summarise the means and variances of the genotypes and indicator variables in Table I [Falconer and Mackay, 1996; Lynch and Walsh, 1998; Visscher and Hopper, 2001].

Although we do not consider a dominance term in the analysis of data, the relevant terms for dominance variance ($\text{var}(D)$) are straightforward. In particular, the total genetic variance at the QTL that is explained by the observed genotype in one relative is,

$$\text{var}(\hat{G}_j) = r_{ij}^2 \text{var}(A) + d_{ij}^2 \text{var}(D),$$

TABLE I. Genotypes and indicator variables for fullsibs when only one is genotyped

Genotype	Genotyped sib			Ungenotyped sib	
	Prob.	$E(y_i)$	x_1	$E(y_2 x_1)$	$x_2 = E(x_2 x_1)$
BB	p^2	a	1	$pa + 1/2(1 - p^2)d$	p
Bb	$2p(1-p) = h$	d	0	$(p - 1/2)a + 1/2[1 + p(1 - p)]d$	$p - 1/2$
bb	$(1-p)^2$	$-a$	-1	$-(1 - p)a + 1/2[1 - (1 - p)^2]d$	$-(1-p)$
Mean		$(2p-1)a + hd$	$2p-1$	$(2p-1)a + hd$	$2p-1$
Variance		$h[a - (2p-1)d]^2 + (hd)^2$	h	$1/4h[a - (2p-1)d]^2 + 1/16(hd)^2$	$1/4h$

with d_{ij} the coefficient of fraternity, the probability that both alleles at a locus of individuals i and j are identical-by-descent [Falconer and Mackay, 1996; Lynch and Walsh, 1998]. For example, for fullsibs or dizygotic twin pairs, a proportion of $(1/4)^2 = 1/16$ of the dominance variance at the QTL in the population of ungenotyped sibs is explained by the observed genotypes in the population of genotyped sibs.

STATISTICAL POWER

We use a simple linear regression model to test for association,

$$y = \mu + \beta x + e.$$

For n unrelated individuals, the variance of the estimate of the regression coefficient is, approximately,

$$\text{var}(\hat{\beta}) = \sigma_e^2 / (n \text{var}(x)).$$

We assume that the proportion of variance explained by the QTL is small relative to the error variance and, without loss of generality, that the effect of the QTL is expressed in residual standard deviation units. The non-centrality-parameter (NCP) for a test of additive association between a genotype and phenotypes is,

$$NCP = \beta^2 / \text{var}(\hat{\beta}) = n2p(1-p)a^2 = nq^2 \quad (1)$$

with q^2 the proportion of phenotypic variance due to the QTL. Now consider n pairs of relatives, with relationship coefficient r , both with phenotypes but only one with a measured genotype. We perform a linear regression for the n individuals with a measured genotype, a separate regression for the n individuals with an inferred genotype, and combine the regression coefficients using weights that depend on r and the phenotypic correlation coefficient (ρ),

$$\beta = w_1\beta_1 + w_2\beta_2.$$

For the individuals with a measured genotype (subscript 1), the variance of the regression coefficient is proportional to $1/\text{var}(x_1)$. For the individuals with an inferred genotype (subscript 2), the variance of the regression coefficient is proportional to $1/(r^2\text{var}(x_1))$. The covariance of the two regression coefficients is proportional to $\rho/(r\text{var}(x_1))$. It follows, from straightforward least squares theory, that the optimum weights for the two regression coefficients are

$$w_1 = (1 - \rho r) / (1 + r^2 - 2\rho r) \text{ and}$$

$$w_2 = r(r - \rho) / (1 + r^2 - 2\rho r).$$

The ratio of the NCP for the combined analysis relative to the NCP for the analysis when only the genotyped individuals are used (or a random sample of n unrelated individuals) is,

$$\frac{NCP(x_1, y_1, x_2, y_2)}{NCP(x_1, y_1)} = \frac{(1 - 2\rho r + r^2)}{(1 - \rho^2)}. \quad (2)$$

By combining equations (1) and (2) we obtain the expected contribution to the test statistic of a single phenotyped but ungenotyped relative as,

$$NCP(x_2, y_2) = q^2(\rho - r^2) / (1 - \rho^2). \quad (3)$$

Finally, we consider the case of m ungenotyped but phenotyped relatives, all with the same relationship (r) to the genotyped individual and assuming that the same phenotypic correlation coefficient (ρ) is the same for all pairs of relatives. In this case, the ungenotyped relatives all have the same inferred genotype indicator variable (x_2) and the mean phenotype of the ungenotyped relatives can be used in the regression analysis. The ratio of the NCP for m relatives with mean phenotype y_m to the NCP without any additional ungenotyped relatives is,

$$\frac{NCP(x_1, y_1, x_2, y_m)}{NCP(x_1, y_1)} = \frac{(c - 2\rho r + r^2)}{(c - \rho^2)} \quad (4)$$

with $c = [1+(m-1)\rho]/m$. The expected contribution of the m relatives to the test statistic for association is

$$NCP(x_2, y_m) = q^2(\rho - r)^2 / (c - \rho)^2 = mq^2(\rho - r)^2 / [(1 - \rho)(1 + m\rho)]. \tag{5}$$

For $m = 1$ equation (5) reduces to equation (3), whereas for a very large number of relatives the expected contribution is $q^2(\rho - r)^2 / [(1 - \rho)\rho]$. Equations (3) and (5) can be used to decide on the most efficient experimental design.

RESULTS

In Figure 1, we plot the relative increase in power for MZ twin pairs and for pairs of fullsibs as a function of the phenotypic correlation coefficient. For MZ twin pairs ($r = 1$), the ratio of the NCP when all data are analysed compared to the NCP when only the genotyped individuals are analysed, is largest when the phenotypic correlation is small. For an MZ phenotypic correlation in the range of 0.4 to 0.8, the gain in power is between 10 and 40%. For pairs of fullsibs, the increase in power is largest when their phenotypic correlation is either very small or very large. For example, the gain in power is ~10% and ~25% when the sib correlation is 0.2 and 0.8, respectively.

In Figure 2, we plot the relative gain in power for fullsibs under the assumption of an additive model of family resemblance. Including ungenotyped fullsibs (including DZ co-twins) into the analysis will effectively increase sample size by 5 to 20% for a realistic range of sib correlations in the range of 0.1 to 0.5.

More general results can be obtained from equation (2), the standardised ratio of the NCP for pairs of related individuals relative to the NCP of using genotyped individuals only:

1. For the limiting case of $\rho = 1$, there is infinitely more information in the combined analysis, essentially because the phenotype of the ungenotyped individual is the same yet its genotype is different.
2. For the case of monozygotic (MZ) twin pairs ($r = 1$), the ratio is $2/(1+\rho)$ and the largest increase of information is for a small phenotypic correlation. Intuitively, this makes sense, because if the phenotypic correlation is low, we obtain a phenotypically independent sample with the same genotype. At the limit ($\rho = 0$), we would effectively be doubling the sample size.
3. For fullsibs (including dizygotic [DZ] twin pairs), the extra information from the combined analysis is $(5/4 - \rho)/(1 - \rho^2)$. This ratio is maximised for $\rho = 1$.
4. There is no gain in power if $\rho = r$. Again this makes sense because when these parameters

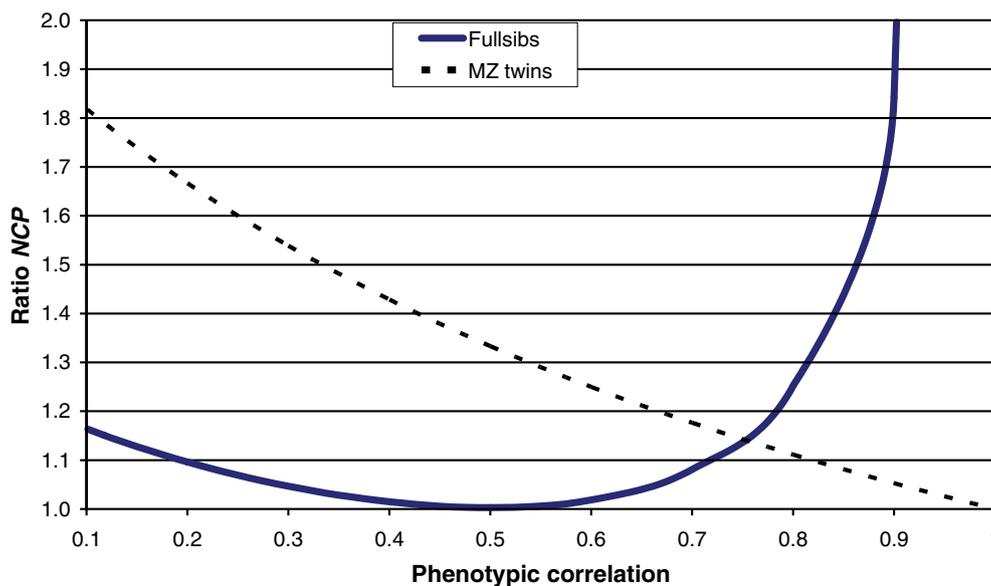


Fig. 1. Ratio of the non-centrality parameter (NCP) for pairs of MZ twins or pairs of fullsibs when only one individual per pair is genotyped relative to the NCP for a random sample of unrelated individuals.

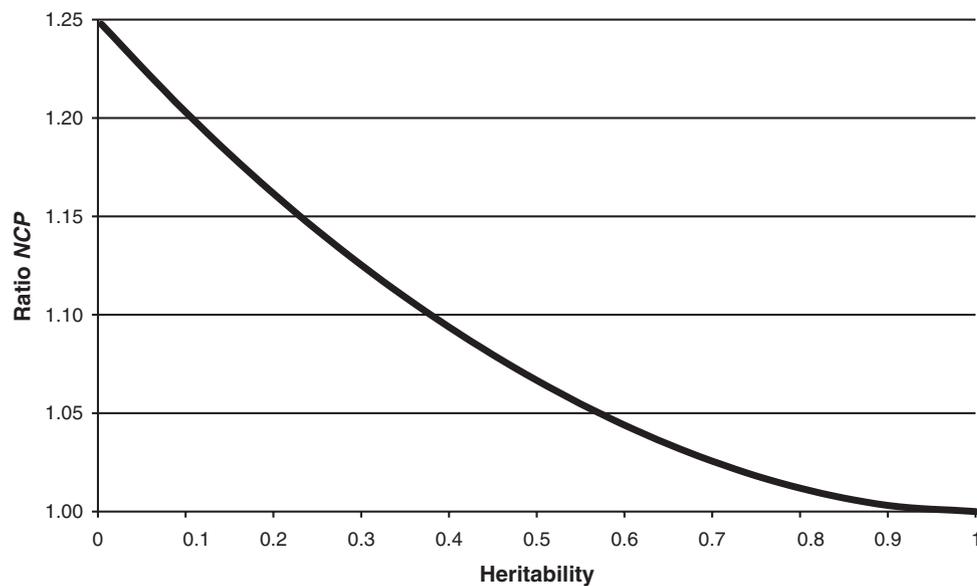


Fig. 2. Power (ratio of *NCP*) to detect association for fullsibs in which only one individuals has been genotyped relative to the power for a random sample of unrelated individuals for an additive genetic model of family resemblance.

are equal, the accuracy of prediction of the phenotype of the ungenotyped individual from the phenotype of the genotyped relative is the same as the accuracy of prediction of the genotype of the ungenotyped individual.

- For a pure additive genetic model of resemblance between relatives, $\rho = rh^2$, with h^2 the heritability of the trait. The relative value of including ungenotyped individuals in the analysis is then $[1+r^2(2-h^2)]/[1-r^2h^4]$ or, approximately, $1+r^2(1-2h^2)$ when h^2 is small. Hence, the increase in power is greatest for close relatives and a low to moderate heritability.

Equations (2) to (5) were used to predict the contribution of ungenotyped but phenotyped relatives for a number of examples. In all examples, we assume that a causal variant has been genotyped on a sample of unrelated individuals, that the QTL explains 2% of the phenotypic variance ($q^2 = 0.02$), that the heritability of the quantitative trait is 0.5, and that the phenotypic correlation between relatives is rh^2 .

EXAMPLE 1: 500 MZ TWIN PAIRS AND 500 DZ TWIN PAIRS, WHERE ONLY ONE OF THE TWINS IS GENOTYPED

The *NCP* for each of the 500 genotyped MZ and 500 genotyped DZ individuals is 10.0 (i.e., an expected contribution to the LOD score of 2.17). The value of the ungenotyped MZ co-twins is 3.33

and that of the ungenotyped DZ co-twins 0.67. Hence, in total the *NCP* of the experiment is 24.0 (expected LOD of 5.22), an increase of 20% relative to an analysis of the genotyped individuals only.

EXAMPLE 2: 1,000 UNRELATED INDIVIDUALS, EACH WITH 4 UNGENOTYPED SIBLINGS

The *NCP* for the genotyped individuals is 20.0 and the extra value of the siblings 3.33 (using equation (5)), giving a total of 23.33, an increase in power of 17%.

EXAMPLE 3: 1,000 UNRELATED INDIVIDUALS, EACH WITH 12 UNGENOTYPED COUSINS

The *NCP* for the genotyped individuals is 20.0 and the extra value of the siblings 0.57 (using equation (5)), giving a total of 20.57, an increase in power of 3%.

SIMULATION

We verified the theoretical results using simulation. Genotypes and phenotypes of pairs of fullsibs were generated using a bi-allelic locus with genotypes in HWE [for details, see Visscher and Hopper, 2001]. We considered a number of different analyses to estimate association between the genotype and phenotype, and these are summarised in Table II. In all but the last one of these analyses, we have ignored the fact that the conditional genotype of the ungenotyped sibling is not known without error. That is, the indicator

TABLE II. Analysis methods^a

Analysis number	Description	Model/estimation
I	Genotyped individuals only	$y_1 = \mu + \beta_1 x_1$
II	Ungenotyped individuals only	$y_2 = \mu + \beta_2 x_2$
III	All, ignoring phenotypic correlation structure	$y = \mu + \beta_3 x$
IV	Weighted regression of I and II	$\beta_4 = w_1 \beta_1 + w_2 \beta_2$
V	Mean phenotypes and genotypes	$1/2(y_1 + y_2) = \mu + \beta_5 1/2(x_1 + x_2)$
VI	All, fitting correlation structure	$\beta = (X'V^{-1}X)^{-1}X'V^{-1}y$
VII	Full maximum likelihood	PAP

^aPAP = Pedigree Analysis Package.

variable (x) used in the regression analysis is assumed to have been observed when in truth it has been estimated. This is a case analogous to QTL linkage mapping, where proportions of alleles shared between individuals are estimated from marker data and used in a regression or variance component analysis. For QTL linkage mapping, the difference, in terms of statistical power and estimates of effects, for methods that do and do not take account of the uncertainty in the estimation of allele sharing, has been shown to be small.

In all but the maximum likelihood analysis using Pedigree Analysis Package (PAP) [Hasstedt and Moll, 1989], we used simple linear regression. The test statistic was calculated as the ratio of the square of the estimated regression coefficient and the estimated variance of the regression coefficient. Genotypes for 1,000 sibpairs were simulated for a variety of effect sizes and allele frequencies. Under the null hypothesis of no associations, all tests apart from test III (using all data but ignoring the phenotypic correlation structure) followed a χ^2_1 distribution. Test statistic β_3 is inflated because the estimated variance of the regression coefficient is smaller than the true variance. When a QTL was simulated, the empirical average test statistics were extremely close to those predicted (results not shown). The test statistic from the weighted linear regression was close to a single analysis in which the correlation structure of the errors was taken into account (results not shown). The mean test statistic of these analyses was the largest, as predicted by the theory.

For a small subset of 10 of the replicates, we analysed the data with a model that took account of the uncertainty of the inferred genotypes, using PAP [Hasstedt and Moll, 1989]. The results showed that no power was gained, relative to either the weighted least squares analysis (analysis IV) or the general linear model (analysis VI) by performing this more complex analysis.

DISCUSSION

We have quantified the value, in terms of increased power and contribution to the expected test statistic, of including ungenotyped relatives in association studies. The increase in power for MZ twin pairs and sibpairs is in the order of 10 to 30%, a substantial increase in effective sample size. Our main result is given in equation (3), which shows that the expected contribution to the test statistic of association of an ungenotyped relative with a phenotype is proportional to the squared difference between the additive genetic relationship and the phenotypic correlation of the genotyped and ungenotyped individuals. This contribution is largest when either the genetic relationship is high but the phenotypic correlation low (e.g., for MZ twins and a trait with a low heritability) or when the genetic relationship is low but the phenotypic correlation high. The latter scenario may be found in societies where a large extended family group lives in the same household environment.

Assuming that the inferred genotypes of the ungenotyped individuals are known without error (when they are not) does not appear to have a large effect on power, analogous to QTL linkage mapping studies [Gessler and Xu, 1996]. Hence, association analysis can be accomplished in two steps: firstly, inferred missing genotypes from the genotype of the relative(s), and, secondly, by fitting standard regression or variance component analyses. In this respect, the proposed association analysis is very similar to QTL linkage analysis using least squares regression or maximum likelihood variance components approaches, which are also performed using a two-step procedure [Almasy and Blangero, 1998; George et al., 2000; Haseman and Elston, 1972].

For selected samples, in which one member of a pair is genotyped and both individuals have an extreme phenotype (e.g., resulting from a design using extremely discordant and concordant relative

pairs), power may be increased if the phenotypes are also used to infer the missing genotype. One suggestion for these designs is to analyse x as a function of y , taking into account that the x values in the ungenotyped individuals have smaller variance than those in the genotyped individuals. However, for the effect sizes considered in this study, the relationship between phenotype and QTL genotype is weak, so that most information to predict the genotype of an unrelated relative comes from the observed genotype of the genotyped individual, and not from the phenotypes.

For more complex pedigrees than the ones considered in this study, the phenotypic correlation structure of the entire sample of individuals with phenotypes should be taken into account. Repeatedly using equations (3) and (5) will over-predict the extra information from all ungenotyped individuals within the pedigree. However, in a deep pedigree the contribution from distant relatives is small, because the genetic relationship is small and the phenotypic correlation between a genotyped individual and a distant relative is likely to be small. To assess the expected contribution from ungenotyped relatives with phenotypes in an association study in a large deep pedigree, gene-drop simulation studies could be performed.

If there is admixture or stratification in the population, then our suggested analysis approach would amplify spurious association. In the absence of observed genotype data on relatives, so that a transmission equilibrium test cannot be performed [Allison, 1997], there are methods that control for population sub-structure [Pritchard and Donnelly, 2001]. Imputing missing genotypes does not in itself create spurious association.

ACKNOWLEDGMENTS

We thank Nick Martin and Brian Kinghorn for discussions and helpful comments, and the referees for many constructive suggestions.

REFERENCES

- Allison DB. 1997. Transmission-disequilibrium tests for quantitative traits. *Am J Hum Genet* 60:676–690.
- Almasy L, Blangero J. 1998. Multipoint quantitative-trait linkage analysis in general pedigrees. *Am J Hum Genet* 62:1198–1211.
- Boerwinkle E, Chakraborty R, Sing CF. 1986. The use of measured genotype information in the analysis of quantitative phenotypes in man. I. Models and analytical methods. *Ann Hum Genet* 50:181–194.
- Falconer DS, Mackay TFC. 1996. *Introduction to quantitative genetics*. Harlow: Longman.
- George AW, Visscher PM, Haley CS. 2000. Mapping quantitative trait loci in complex pedigrees: a two-step variance component approach. *Genetics* 156:2081–2092.
- Gessler DD, Xu S. 1996. Using the expectation or the distribution of the identity by descent for mapping quantitative trait loci under the random model. *Am J Hum Genet* 59:1382–1390.
- Haseman JK, Elston RC. 1972. The investigation of linkage between a quantitative trait and a marker locus. *Behav Genet* 2:3–19.
- Hasstedt SJ. 2002. Pedigree analysis package, Version 5.0. Salt Lake City: University of Utah.
- Hasstedt SJ, Moll PP. 1989. Estimation of genetic model parameters: variables correlated with a quantitative phenotype exhibiting major locus inheritance. *Genet Epidemiol* 6:319–332.
- Hernandez-Sanchez J, Visscher P, Plastow G, Haley C. 2003. Candidate gene analysis for quantitative traits using the transmission disequilibrium test: The example of the melanocortin 4-receptor in pigs. *Genetics* 164:637–644.
- Kerr RJ, Kinghorn BP. 1996. An efficient algorithm for segregation analysis in large populations. *J Anim Breed Genet Z Tierzucht Zuchtungsbiol* 113:457–469.
- Kinghorn BP, Kennedy BW, Smith C. 1993. A method of screening for genes of major effect. *Genetics* 134:351–360.
- Lange K, Sinsheimer JS, Sobel E. 2005. Association testing with Mendel. *Genet Epidemiol* 29:36–50.
- Lange K, Westlake J, Spence MA. 1976. Extensions to pedigree analysis. III. Variance components by the scoring method. *Ann Hum Genet* 39:485–491.
- Li H, Thompson E. 1997. Semiparametric estimation of major gene and family-specific random effects for age of onset. *Biometrics* 53:282–293.
- Lynch M, Walsh B. 1998. *Genetics and analysis of quantitative traits*. Sunderland, MA: Sinauer Associates.
- Pritchard JK, Donnelly P. 2001. Case-control studies of association in structured or admixed populations. *Theor Popul Biol* 60:227–237.
- Visscher PM, Hopper JL. 2001. Power of regression and maximum likelihood methods to map QTL from sib-pair and DZ twin data. *Ann Hum Genet* 65:583–601.
- Vitezica ZG, Elsen JM, Rupp R, Diaz C. 2005. Using genotype probabilities in survival analysis: a scrapie case. *Genet Sel Evol* 37:403–415.
- Wacholder S, Hartge P, Struewing JP, Pee D, McAdams M, Brody L, Tucker M. 1998. The kin-cohort study for estimating penetrance. *Am J Epidemiol* 148:623–630.