# Letters to the Editor

*Am. J. Hum. Genet. 69:1146–1148, 2001*

## The Importance of Genealogy in Determining Genetic Associations with Complex Traits

*To the Editor:*

Most common diseases, such as asthma, type 2 diabetes, bipolar disorder, and cardiovascular disease, are known to have genetic components, but the susceptibility genes have been notoriously difficult to localize and to identify. These complex diseases likely have a large number of genetic and nongenetic risk factors that together have varying effects on phenotype. Many investigators have recommended founder populations for complex-trait mapping, with the expectation that fewer susceptibility alleles will be segregating in these restricted gene pools (Lander and Schork 1994; Wright et al. 1999; Shifman and Darvasi 2001). Some or all individuals in these populations are inbred, but often the exact relationships between all members are either unknown or not taken into account. It is tempting to use such populations for their presumed homogeneity, even in the absence of accurate pedigree information. The failure to take full pedigree information into account can either reduce the power to detect linkage (Dyer et al., in press) or inflate LOD scores (Miano et al. 2000). The failure to account for relatedness among individuals will also affect association studies. In particular, many statistical tests of association are not strictly valid, owing to the lack of true independence between individuals. Nonetheless, such populations have been used widely in association studies (e.g., de Silva et al. 1999; Laprise et al. 2000; Ospina-Duque et al. 2000; Summerhill et al. 2000; Bitti et al. 2001; Hegele et al. 2001), and some authors have even recommended the inclusion of founder populations in case-control studies, owing to their decreased heterogeneity (Shifman and Darvasi 2001). However, the impact that ignoring pedigree relationships has on tests of association has not been evaluated.

The Hutterites are an extreme example of a large, complex pedigree with multiple inbreeding loops. We are in the unique position of having complete genealogical information on this 12,903-person, 13-generation pedigree (Abney et al. 2000). Additionally, we have extensive phenotype characterization and a dense microsatellite map (of 568 short-tandem-repeat–polymorphism markers) of ~750 members of this population, who are descendants of just 64 Hutterite founders (Ober et al. 2000). Thus, we were able to assess the effect that ignoring pedigree information has on statistical tests of association.

We performed two separate genomewide scans of association on each of three quantitative phenotypes: serum immunoglobulin E (IgE), serum LDL, and body-mass index (BMI). These phenotypes were chosen to represent quantitative traits associated with diverse complex diseases (asthma, cardiovascular disease, and diabetes, respectively). All phenotypes were adjusted for age and sex and were transformed so that the residuals were approximately normally distributed (Abney et al. 2001). The heritabilities of IgE and BMI were completely accounted for by additive genetic variance, with heritabilities of .63 and .54, respectively; the heritability of LDL had a strong dominance component in addition to additive genetic variance, with a broad heritability of .96 (discussed in detail by Abney et al. [2001]). To estimate the effect that each allele at each locus has on the trait values, we used a statistical test of association, developed specifically for use in large, inbred pedigrees (Ober et al. 2001 [in this issue]). Pedigree structure is taken into account by the use of variance components to model the polygenic background (Abney et al. 2000, 2001). When pedigree structure is ignored, the method is equivalent to a linear regression of the trait on age, sex, and genotype, with a Bonferroni correction applied to the $P$ value for the $t$ test for significance of genotype. In the first scan, we included the variance components and therefore took into account the relatedness between individuals. In the second scan, we did not include any pedigree information (additive and dominance variances were 0). In both scans, the observed $P$ values were adjusted for multiple comparisons, by use of a Bonferroni correction.

The two methods yielded dramatically different results. In general, the significance of association with a given marker was considerably inflated when pedigree structure was not included, although in some cases the reverse was true (see fig. 1). Only two loci were among the five most significant results, by both methods, for the three traits (the same marker at 7p21 was associated with IgE, by both methods, and the same marker at 8q12

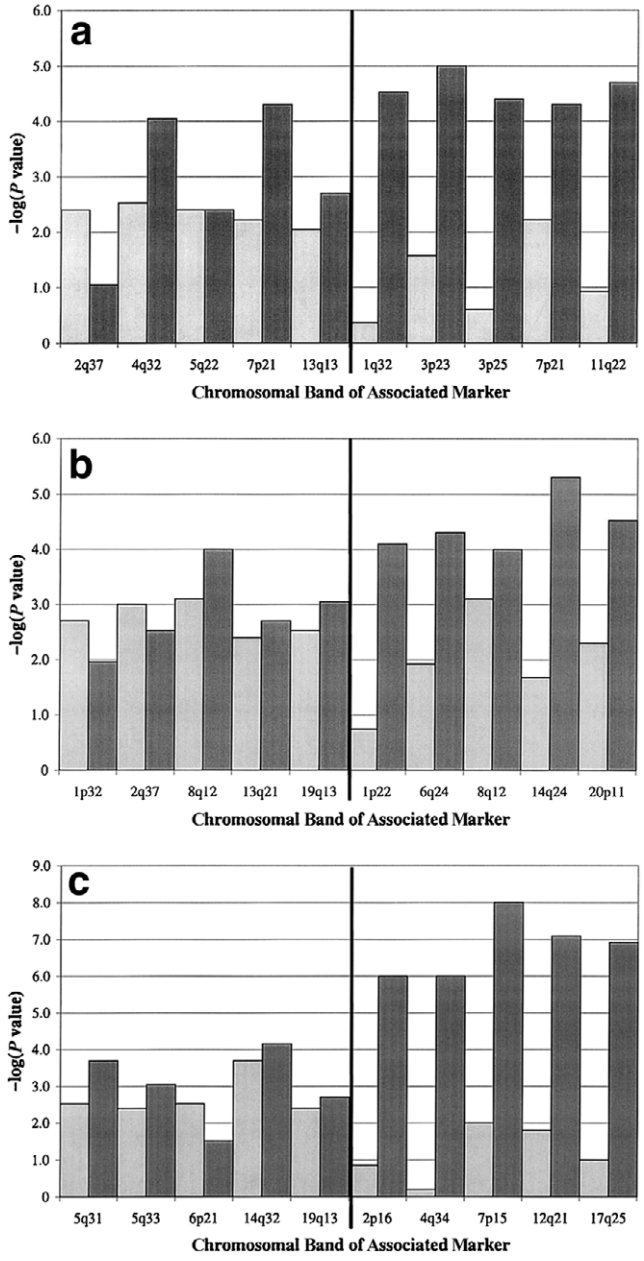


**Figure 1** Results of two genome scans—one including pedigree structure (*lighter bars*) and one not including pedigree structure (*darker bars*)—for IgE (*a*), LDL (*b*), and BMI (*c*). The five most significant loci when structure is included (*left sides*) and when structure is not included (*right sides*) are shown.

was associated with LDL, by both methods). In addition, many more loci showed evidence of association when the pedigree structure was not included (see fig. 2). In fact, 10%–22% of all markers appeared to have a strong association ($P < .01$) with the phenotype, when pedigree structure was not included. Results for 58 single-nucleotide polymorphisms showed similar trends (data not shown).

Although the Hutterites are an extreme example of a complex pedigree, there are several other populations known to have similar structures (Badner et al. 1990; Slutsky et al. 1997; Hsueh et al. 2000). Moreover, individuals from a variety of smaller, island populations who either have been or are currently being studied may be more related to each other than can be discerned from the recently collected pedigree data (de Silva et al. 1999; Mathias et al. 2000; Bitti et al. 2001). In fact, even presumably outbred populations may contain hidden consanguinity (Broman and Weber 1999), and cryptic relatedness may be a problem in association studies of rare disorders (Bacanu et al. 2000). This problem may be avoided by the use of statistical tools designed to detect misspecified or cryptic relationships (McPeek and Sun 2000; Sun et al., in press).

We cannot prove that the inclusion of the pedigree structure in the method results in true associations, until the alleles contributing to these quantitative traits are found; however, we believe that the number of associations found when structure is ignored is unrealistic. Presumably, a profound failure of the assumption of independence between individuals, in method 2, results in a dramatically increased number of type 1 errors. Overall, our data suggest that failing to take into account extended-familial relationships can result in a large number of false-positive results, and some "true" associations may be missed. In addition, the level of significance could be overestimated by several orders of magnitude. In an association study in which it is not possible to take into account all familial relationships, as we have done with the Hutterites, another option is to use genomic controls (Devlin and Roeder 1999). Otherwise, naïve approaches to genetic-association analysis could result

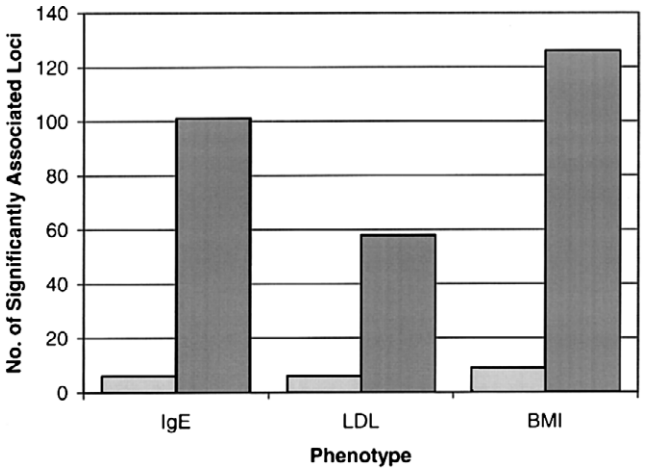


**Figure 2** Number of significantly associated ($P < .01$) loci when pedigree structure is included (*lighter bars*) and when pedigree structure is not included (*darker bars*).

in an enormous amount of time and of money spent in following up artifactual associations.

## Acknowledgments

We thank Harvey Dytch for assistance with computer programming and data management. This work was supported by National Institutes of Health grants DK55889, HD56399, and HG01645.

DINA L. NEWMAN,[1] MARK ABNEY,[1,2] MARY SARA MCPEEK,[1,2] CAROLE OBER,[1] AND NANCY J. COX[1]

*Departments of [1]Human Genetics and [2]Statistics*
*University of Chicago*
*Chicago*

## References

Abney MA, McPeek MS, Ober C (2000) Estimation of variance components of quantitative traits in inbred populations. Am J Hum Genet 66:629–650

——— (2001) Narrow and broad heritabilities of quantitative traits in a founder population. Am J Hum Genet 68:1302–1307

Bacanu SA, Devlin B, Roeder K (2000) The power of genomic control. Am J Hum Genet 66:1933–1944

Badner JA, Sieber WK, Garver KL, Chakravarti A (1990) A genetic study of Hirschsprung disease. Am J Hum Genet 46: 568–580

Bitti PP, Murgia BS, Ticca A, Ferrai R, Musu L, Piras ML, Puledda E, Campo S, Durando S, Montomoli C, Clayton DG, Mander AP, Bernardinelli L (2001) Association between the ancestral haplotype HLA A30B18DR3 and multiple sclerosis in central Sardinia. Genet Epidemiol 20:271–283

Broman KW, Weber JL (1999) Long homozygous chromosomal segments in reference families from the Centre d'Etude du Polymorphisme Humain. Am J Hum Genet 65:1493–1500

de Silva AM, Walder KR, Aitman TJ, Gotoda T, Goldstone AP, Hodge AM, de Courten MP, Zimmet PZ, Collier GR (1999) Combination of polymorphisms in OB-R and the OB gene associated with insulin resistance in Nauruan males. Intl J Obes Relat Metab Disord 23:816–822

Devlin B, Roeder K (1999) Genomic control for association studies. Biometrics 55:997–1004

Dyer TD, Williams JT, Goring, HHH Blangero J. The effect of pedigree complexity on quantitative trait linkage analysis. Genet Epidemiol Suppl (in press)

Hegele RA, Wang J, Harris SB, Brunt JH, Young TK, Hanley AJ, Zinman B, Connelly PW, Anderson CM (2001) Variable association between genetic variation in the *CYP7* gene promoter and plasma lipoproteins in three Canadian populations. Atherosclerosis 154:579–587

Hsueh WC, Mitchell BD, Aburomia R, Pollin T, Sakul H, Gelder Ehm M, Michelsen BK, Wagner MJ, St Jean PL, Knowler WC, Burns DK, Bell CJ, Shuldiner AR (2000) Diabetes in the Old Order Amish: characterization and heritability analysis of the Amish Family Diabetes Study. Diabetes Care 23:595–601

Lander ES, Schork NJ (1994) Genetic dissection of complex traits. Science 265:2037–2048

Laprise C, Boulet LP, Morissette J, Winstall E Raymond V (2000) Evidence for association and linkage between atopy, airway hyper-responsiveness, and the $\beta$ subunit Glu237Gly variant of the high-affinity receptor for immunoglobulin E in the French-Canadian population. Immunogenetics 51:695–702

Mathias RA, Bickel CA, Beaty TH, Petersen GM, Hetmanski JB, Liang KY, Barnes KC (2000) A study of contemporary levels and temporal trends in inbreeding in the Tangier Island, Virginia, population using pedigree data and isonymy. Am J Phys Anthropol 112:29–38

McPeek MS, Sun L (2000) Statistical tests for detection of misspecified relationships by use of genome-screen data. Am J Hum Genet 66:1076–1094

Miano MG, Jacobson SG, Carothers A, Hanson I, Teague P, Lovell J, Cideciyan AV, Haider N, Stone EM, Sheffield VC, Wright AF (2000) Pitfalls in homozygosity mapping. Am J Hum Genet 67:1348–1351

Ober C, Abney M, McPeek MS (2001) The genetic dissection of traits in a founder population. Am J Hum Genet 69: 1068–1079 (in this issue)

Ober C, Tsalenko A, Parry R, Cox NJ (2000) A second-generation genomewide screen for asthma-susceptibility alleles in a founder population. Am J Hum Genet 67:1154–1162

Ospina-Duque J, Duque C, Carvajal-Carmona L, Ortiz-Barrientos D, Soto I, Pineda N, Cuartas M, et al (2000) An association study of bipolar mood disorder (type I) with the 5-HTTLPR serotonin transporter polymorphism in a human population isolate from Colombia. Neurosci Lett 292:199–202

Shifman S, Darvasi A (2001) The value of isolated populations. Nat Genet 28:309–310

Slutsky AS, Zamel N, University of Toronto Genetics of Asthma Research Group (1997) Genetics of asthma: the University of Toronto Program. Am J Respir Crit Care Med Suppl 156: S130–S132

Summerhill E, Leavitt SA, Gidley H, Parry R, Solway J Ober C (2000) $\beta_2$-Adrenergic receptor Arg16/Arg16 genotype is associated with reduced lung function, but not with asthma, in the Hutterites. Am J Respir Crit Care Med 162:599–602

Sun L, Abney M, McPeek MS. Detection of misspecified relationships in inbred and outbred pedigrees. Genet Epidemiol Suppl (in press)

Wright AF, Carothers AD, Pirastu M (1999) Population choice in mapping genes for complex diseases. Nat Genet 23:397–404

Address for correspondence and reprints: Dr. Dina L. Newman, University of Chicago, 920 East 58th Street, CLSC 501, Chicago, IL 60637. E-mail: dnewman@genetics.uchicago.edu


0002-9297/2001/6905-0023$02.00