Genetic Association Studies

Kathryn L. Lunetta, PhD

With the completion of the HapMap project[1] and the development of technology that allows the examination of ≥1 million genetic polymorphisms at once, genetic association studies are becoming more comprehensive. This article first provides a brief overview of the rationale for genetic association studies; it then discusses the primary features differentiating genetic from standard association studies and emphasizes these differences with an example. Finally, this article reviews methods for addressing 2 of the main pitfalls of genetic association studies: population stratification and multiple testing. The principal focus of this primer is population-based association studies using unrelated individuals. A future article will address family-based linkage and association studies.

Rationale

Traditional epidemiological studies focus on assessing the impact of specific risk factors on disease risk in populations. The goal of a genetic association study is to establish statistical associations between ≥1 genetic polymorphisms and phenotypes or disease states and thus to identify genetic risk factors that can later be studied in a more comprehensive manner using traditional epidemiological methods. Ideally, the statistical analyses brings us to the point where 1 or several genetic variants are identified as the potential functional variants within a gene, so that laboratory scientists can then use experimental methods to determine what functional purpose the variants have and how it might relate to the phenotype. Historically, the term polymorphism has been used to refer to genetic mutations that occur with a frequency ≥1% in the population. This article refers to genomic locations with multiple alleles interchangeably as genetic variants or polymorphisms. Pollex and Hegele[2] describe many types of genetic variants found in the human genome and review the current state of knowledge concerning copy number variants and cardiovascular disease. This article focuses on single-nucleotide polymorphisms (SNPs), although much of what is presented is relevant to all types of variants.

We expect to see an association between a genetic variant and phenotype when the variant has a functional effect on the trait or when it is in linkage disequilibrium (LD) with a functional variant. LD is the nonindependence of alleles at 2 (or more) loci in a population resulting from their close proximity on a chromosome. The LD between 2 loci is a function of the crossover rate and the number of generations since the mutation occurred or was introduced into the population. LD makes genome-wide association studies possible. Although there are millions of polymorphisms in the human genome, many are in LD with each other and thus carry redundant information. Testing 1 variant gives information about others, so it is not necessary to test all polymorphisms. Several recent articles provide excellent analyses and comparisons of the extent of LD in various human populations.[3–6]

Genetic Association Studies

Genetic association studies should not be pursued unless the trait being studied has established evidence for heritability; ie, evidence for familial correlation or disease clustering should be unequivocal. The primary features differentiating how we test for association with a genetic polymorphism versus typical epidemiological covariates are that we test genetic variants for Hardy-Weinberg equilibrium (HWE) before testing for association, and we must specify a genetic model for the association test. This article briefly describes the options for study focus and design and presents the methods and rationale for HWE testing and genetic model selection.

Study Focus: Candidate Gene Versus Genome-Wide Design

Until recently, most genetic association studies examined a single polymorphism or a set of polymorphisms near a single gene or focused on a candidate region defined by a linkage peak determined by a family study. With the ever-improving genotyping technology, genome-wide association studies, with hundreds of thousands or even millions of polymorphisms genotyped, have become feasible. The genome-wide approach is unbiased in that it does not require prior hypotheses about what types of genes or polymorphisms are most likely to be associated with the phenotypes of interest. However, testing a large number of polymorphisms required for a genome-wide study comes at a price; the power to definitively identify associations is low when such a large number of tests is performed. This issue is discussed below. Candidate gene studies limit the number of tests to a small subset of the genome and focus hypotheses on sets of genes that we have prior reason to believe might be associated with

From the Department of Biostatistics, Boston University School of Public Health, Boston, Mass.
Correspondence to Kathryn L. Lunetta, Department of Biostatistics, Boston University School of Public Health, 715 Albany St, Crosstown Center, 3rd Floor, Boston, MA 02118. E-mail klunetta@bu.edu
(*Circulation*. 2008;118:96-101.)

Circulation is available at http://circ.ahajournals.org
DOI: 10.1161/CIRCULATIONAHA.107.700401

the phenotype of interest. However, even in the context of a candidate gene study, we may end up testing tens of thousands of SNPs. Selecting SNPs that best represent the common variation in the genome (eg, tag SNPs)[7] helps to minimize the number of SNPs tested, but using tag SNPs lowers power compared with testing the functional SNPs.[8,9]

Study Designs and Outcome Measures

Study design and choice of outcome for genetic association studies should follow the same concepts and principles used for any epidemiological study. Typical study designs include case-control, in which individuals with and without the outcome of interest are ascertained, and random ascertainment, in which a random sample of individuals from a population are studied. The phenotypic outcome of interest affects the choice of study ascertainment; eg, rare outcomes are best studied with a case-control design because random samples will select few individuals with the outcome of interest. Recent articles in this primer series have presented overviews of regression for analysis of quantitative outcomes[10] and survival methods for time-to-onset outcomes.[11] Detailed review of the variety of study designs and outcome measures is beyond the scope of this tutorial. More comprehensive information is available in epidemiology and statistics texts, eg, the text by Jewel.[12]

Hardy-Weinberg Equilibrium

In association studies, we generally test each SNP for HWE before testing for association with phenotypes. The Hardy-Weinberg law states that the genotype frequencies and allele frequencies of a large, randomly mating population remain constant from generation to generation provided migration, mutation, and selection do not take place. Therefore, HWE is the stable distribution of frequencies of the genotypes AA, Aa, and aa in the proportions p^2, 2pq, and q^2, respectively (where p and q are the frequencies of the alleles A and a). HWE is a consequence of random mating within a population in the absence of mutation, migration, natural selection, or random drift. Practically speaking, it is the state in which the maternally and paternally inherited alleles of an individual at a particular locus are statistically independent. A significant departure from HWE for a SNP in a sample may indicate nonrandom mating and possibly population stratification, nonrandom genotyping error, or missing genotype data in which 1 allele or genotype is more often misclassified or missing than the other. Genotyping error or missing genotype data may lead to spurious associations if the probability that genotypes are missing or misclassified differs for different phenotypes.[13,14] Nonrandom mating may be due to structured or stratified samples and can result in spurious association as a result of confounding, as described below.

We test a SNP for HWE by comparing the observed genotype counts in a sample with those expected under HWE. The simplest test is a goodness-of-fit χ^2 test. We estimate the SNP allele frequencies p and $q=1-p$ by determining their proportions in the sample and then determining the expected genotype counts using the HWE expected frequencies Np^2, 2Npq, and Nq^2, where N is the number of individuals genotyped. Then, a goodness-of-fit χ^2 test compares the observed and expected counts. Because the goodness-of-fit test gives inflated type I error rates under some conditions, including rates for polymorphisms with small minor allele frequencies, alternative exact tests of HWE have been developed[15,16] and are becoming more widely used.

In a sample that is not ascertained on the basis of any specific phenotype, the HWE test should be performed on the full sample. For ascertained samples such as a case-control samples, if the population prevalence of the trait is low, Hardy-Weinberg testing should be conducted in the controls because we expect departure from HWE among cases for any polymorphism that is associated with case status. For common traits, we expect both cases and controls to depart from HWE for polymorphisms associated with case status. SNPs with genotypes that depart significantly from Hardy-Weinberg–expected proportions usually are excluded from association analyses. The criterion used to decide whether or not to omit a SNP from association analyses depends on a number of factors, including the number of SNPs tested and the call rate (proportion of observations successfully genotyped); often, SNPs with HWE test values of $P<0.01$ or 0.001 are omitted from association analyses.

Genetic Model

There are 2 approaches for testing association between polymorphisms and an outcome: allelic and genotypic tests. To determine the appropriate test for association, we must first specify a genetic model. We can assume dominance of one of the alleles by treating the heterozygote and one of the homozygote genotypes as a single category. This dichotomization of the SNP genotypes forces heterozygotes to have the same risk or mean phenotype as one of the homozygotes. Additive models impose a structure in which each additional copy of the variant allele increases the response, whether log odds ratio, log hazard ratio, or mean phenotype, by the same amount. A general genetic model retains the 3 distinct genotype classes and makes no assumptions about how the risk or mean for heterozygotes compares with the 2 homozygotes. The general model requires 2 *df* for testing association for a SNP, whereas the other models require only 1 *df*.

For categorical outcomes, the simplest association test is a χ^2 test of independence computed on a cross-classification table of outcome versus alleles or genotypes. The test has degrees of freedom $(m-1)(n-1)$, where n is the number of phenotypic classes and m is the number of genotypic or allelic classes. For example, $m=3$ for a SNP genotype test if all 3 genotypes are observed and $m=2$ for a SNP allele test. Allelic association tests assume that the 2 alleles within each individual are independent (ie, that they are in HWE). Armitage's trend test and other tests that assume additivity of allele effects are alternatives that do not impose this assumption[17] and are therefore preferred. Under HWE, the allele-based test and the trend test are asymptotically equivalent. The general model, or genotype-based test, which treats the 3 genotypes as separate categories, is the most flexible choice, but the additional degree of freedom required results in a test that is less powerful than the correct genetic model when the correct model is known. The options for the genetic model are the same for any regression-based association analysis in

Table 1. Recognized Acute MI at Exam 6 by ESR1 *Pvu*II genotype in the Framingham Offspring Study

	Observed Genotype Counts, n (%)				Observed Allele Counts, n (%)		Expected Counts Under HWE, n				
	TT	CT	CC	Total	T	C	TT	CT	CC	χ^2 Statistic	*P*
MI	16 (27)	21 (36)	22 (37)	59	53 (45)	65 (55)	11.90	29.19	17.90	4.65	0.03
No MI	509 (30)	841 (50)	330 (20)	1680	1859 (55)	1501 (45)	514.27	830.46	335.27	0.27	0.60
Total	525 (30)	862 (50)	352 (20)	1739	1912 (55)	1566 (45)	525.55	860.89	352.55	0.003	0.96

which one can use a factor with 3 levels to allow a general genetic model or code the alleles as dominant, recessive, or additive. For example, for quantitative phenotypes, ANOVA, a type of linear regression meant for quantitative outcomes and categorical predictors, can be used to test for association between a genotype and a phenotype. Instead of comparing counts of cases and controls for each genotype, we look for differences in mean phenotype among the genotype classes.

Adjustment for Covariates

Unlike classic epidemiology studies, SNP association studies are unlikely to be confounded by behavioral and environmental factors because these factors usually do not alter genotype. If behavioral or environmental factors affect phenotype independently of the genes of interest, however, adjustment may increase precision. Adjustment for traits or comorbidities that also may be associated with SNPs will remove confounding resulting from these factors.

Example

Estrogen receptor α (ESR1) polymorphisms were tested for association with cardiovascular disease outcomes[18] in a subset of independent individuals in the Framingham offspring cohort. Details concerning the Framingham offspring cohort selection criteria have been described.[19] Here, we use the previously published genotype and phenotype data for the ESR1 polymorphism c.454 to 397T>C, also known as −397T/C, as *Pvu*II, and by its RefSNP accession ID rs2234693, to illustrate a simple SNP association analysis. Table 1 shows the number of individuals by acute recognized myocardial infarction (MI) status and genotype. The proportion of individuals with the CC genotype is greater among individuals who have had a recognized acute MI than among those who have not. The second set of columns in Table 1 display the allele counts and percentages. We determine the allele counts by summing the total number of T and C alleles in each category; there are twice as many alleles as genotypes. The third set of columns in Table 1 display the expected genotype counts under the assumption of HWE. This sample of individuals was selected randomly from the population of Framingham, so testing for HWE in the full sample is appropriate. There is no evidence to reject the assumption of HWE in the sample ($P=0.96$). For the small subset of 59 individuals with recognized acute MI, there is evidence for a lack of HWE ($P=0.03$). We expect departure from HWE among cases for polymorphisms associated with case status. Table 2 presents the odds ratios (ORs) and test statistics for tests of association between acute MI status and ESR1 genotype under several models. Every model provides significant evidence for association except the model that combines CT and CC genotypes (the dominant C allele model). The reason is that the crude ORs indicate that individuals carrying the CC genotype are at increased risk compared with those with the TT genotype (OR, 2.12), whereas individuals carrying the CT genotype have somewhat decreased risk compared with those with the TT genotype (OR, 0.79). For this example, the trend test and the allele test produce nearly identical statistics. When we impose an additive model on the data, we force the odds for CT individuals to be between that of CC and TT individuals. It is evident from the reduced level of significance of these 2 additive model tests compared with the general genotype model that additivity is not a good fit to the data. The model treating the T allele as dominant, which combines the TT and CT genotypes into 1 category, provides the smallest *P* value of all the association tests. The difference between the test statistics for the general model and the specific models provides information about the fit of each model. The general model always has the largest χ^2 statistic; specific models with similar χ^2 statistics provide the best fit to the data.

Three quantitative phenotypes measured at entry to the study also were tested for association with the polymorphism. Table 3 presents the mean and SD for body mass index, total cholesterol, and high-density lipoprotein cholesterol, along with the *P* value from an ANOVA F test comparing the 3 means. None of the phenotypes differ significantly in mean

Table 2. Tests of Association Between ESR1 *Pvu*II Genotype and Recognized Acute MI

	Genotype					
	TT	CT	CC	χ2 Statistic	*df*	*P*
Crude OR (vs TT)	1	0.79	2.12	11.36	2	0.0034
Additive model OR (vs TT)	1	1.52	2.31	5.00	1	0.025
Dominant C allele OR (vs TT)	1	1.17	1.17	0.27	1	0.60
Dominant T allele OR (vs TT+CT)	1	1	2.43	10.99	1	0.0009
Allele table, C vs T allele	1 (T)	1.52 (C)		4.99	1	0.025

Table 3. Baseline Characteristics of Participants by ESR1 *Pvu*II Genotype

	Genotype			
	TT	CT	CC	*P*
BMI, kg/m^2	25 (4.0)	25 (4.2)	25 (4.6)	0.32
Total cholesterol, mg/dL	194.4 (36.0)	196.5 (39.2)	193.6 (36.1)	0.39
HDL cholesterol, mg/dL	50.7 (50.8)	50.9 (50.9)	50.0 (50.0)	0.61

BMI indicates body mass index; HDL, high-density lipoprotein. Values are mean (SD).

by genotype ($0.32 \leq P \leq 0.61$). Given how similar the means are across genotypes, it is clear that neither an additive model nor a dominant model would increase the evidence for association.

Pitfalls

Two of the common problems that must be addressed in any genetic association study are population stratification and the large numbers of hypotheses tested.

Population Stratification

Population stratification refers to the situation in which individuals in a study differ by ethnic background or another potentially confounding factor for different phenotypes. For example, a study might ascertain cases and controls so that cases have a greater proportion of subjects of Hispanic descent than controls. Spurious association resulting from population stratification can occur if both the phenotype distribution and the genotype distribution differ among the subpopulations (eg, ethnicity). When we know the subpopulation membership of individuals, we can perform stratified analyses and remove all confounding. For example, an analysis of a sample consisting of black and Asian individuals can be stratified by ethnicity. Alternatively, we can use family-based study designs and family-based association tests, which stratify analyses by family. However, in many situations, we do not have reliable information about the structure, nor do we have a family-based study design. Under these conditions, a number of options exist. First, we should adjust for any covariates that may be related to population structure. These may include self-reported ethnicity and geographic location (eg, study site for a multisite study or place of birth).[20] After removing the effects of these potential confounders, one can adjust for the residual, average level of stratification using the method known as genomic control,[21] which removes the average bias resulting from population structure. Some SNP allele frequencies vary across populations and are therefore susceptible to stratification bias. Genomic control may undercorrect for stratification for SNPs with extreme differences across subpopulations that can occur in a population that otherwise appears to have low levels of structure. For example, a set of 178 SNPs typed on a sample of Europeans yielded no evidence for population stratification using genomic control, yet a specific SNP in the *LCT* gene demonstrated significant association with height that was later attributed to stratification bias.[22] There are several alternatives to genomic control. For genome-wide association data, individuals can be clustered into genetically homogeneous subsets using pairwise identity-by-state information across all loci. Association analyses can then be performed stratified by cluster.[23] Alternatively, principal-components analysis can be used to adjust for genetic ancestry.[24,25] For data sets with fewer SNPs, the model-based structured association method of Pritchard et al[26,27] assigns individuals to latent subpopulations and then performs stratified association tests. In practice, when we have a new, significant association to report, it is useful to gather data on allele frequencies across many populations from public databases. If the SNP tends to have similar allele frequencies across populations, it is unlikely to be subject to spurious association resulting from stratification.

Multiple Testing

Contemporary association studies consider multiple polymorphisms. Additionally, some studies report the results of multiple, often correlated phenotypes or the results of multiple genetic models or covariate adjustments. We define the power of an association test to be the probability that the association test rejects the null hypothesis under the condition that the polymorphism is truly associated with the phenotype. A type I error occurs when we reject the null hypothesis when, in fact, there is no true association between the polymorphism and the phenotype. The nominal significance level is the type I error rate, α, selected for the individual association tests. Traditionally, when only 1 or a few tests are performed, we set $\alpha = 0.05$. For studies in which we test many hypotheses, the nominal significance level chosen for a study dictates the proportion of all of the reported tests that are found to be significant, even when none of the hypotheses are true. Usually, when a large number of hypotheses (eg, SNPs) are tested, we adjust the nominal significance level downward so that we do not falsely reject too many hypotheses. In the context of candidate gene and genome-wide association studies, many methods have been proposed to account for the large number of tests performed while attempting to retain high power. Two complementary approaches exist for minimizing the effects of multiple testing: We can incorporate some strategy for limiting the number of association tests performed, and we can adjust for the number of tests that we do perform.

For a study of a single phenotype and multiple SNPs, the most efficient way to limit the number of tests is to perform a single test per SNP. For a SNP that is truly associated with the phenotype, the most powerful test is the one that most closely reflects the true, underlying genetic model. However, because the true genetic model is not known, the general or additive genetic model is usually the best choice. The genetic model and test to be used should be determined before analysis. A second option to limit the number of tests includes performing a test of association between the phenotype and haplotypes or multilocus genotypes rather than each SNP individually. For a study with multiple phenotypes and multiple SNPs, 1 option for limiting the number of tests is to use a multivariate test to test for association between the set of phenotypes and each SNP. For any SNP associated

with the set of phenotypes, individual tests of association between each phenotype and the SNP will help to determine which phenotype or subset of phenotypes is associated with the SNP.

The simplest correction for multiple testing is the Bonferroni adjustment, in which we multiply nominal *P* values by the total number of tests performed. The underlying assumption is that the set of tests are independent; therefore, Bonferroni correction is conservative in the context of correlated tests. The adjustment controls the family-wise error rate, which is the probability of at least 1 type I error. For example, if one sets the experiment-wide error rate at 0.05, then the Bonferroni-adjusted *P* values must be <0.05 to be considered significant, and the probability of observing at least 1 such result in the entire experiment is ≤ 0.05. Nyholt[28] introduced a refinement of the Bonferroni procedure for the case of SNPs in linkage disequilibrium. One estimates the effective number of independent SNPs among a set of genotyped SNPs; this number is then substituted for the total number of SNPs tested in the Bonferroni adjustment. Because the effective number of independent SNPs is always less than or equal to the total number of SNPs tested, this method is less conservative than the Bonferroni procedure.

Permutation testing is an alternative way to adjust for large numbers of tests in typical association studies. This method incorporates the correlation between phenotypes and/or between genotypes and is therefore less conservative than Bonferroni adjustment. The basic idea is to permute phenotype(s) with respect to the genotype(s) among observations, thus removing any association between phenotypes and genotypes but retaining the correlation among phenotypes and among genotypes resulting from LD within an individual. The process is done thousands of times; all of the association test statistics and corresponding *P* values that were computed on the original data set are recomputed on each permuted data set. Finally, the minimum *P* value from among the original data set association tests is compared with the distribution of the minimum *P* values obtained from the set of permuted data sets.

The false discovery rate, first proposed by Benjamini and Hochberg,[29] is a less stringent form of adjustment. In contrast to the Bonferroni adjustment, the false discovery rate controls the expected proportion of false discoveries among all rejected hypotheses. In general, controlling the false discovery rate allows us to reject more hypotheses than controlling the family-wise error rate. However, for genome-wide or large-scale candidate gene association studies, a simple false discovery rate adjustment may still be very stringent and result in low power to detect associations with modest effect size. Weighted false discovery rates[30,31] and stratified false discovery rates[32] set different criteria for significance (or follow-up) for different categories of tests. These methods will result in a greater power to detect true associations if different subsets of SNPs have a higher proportion of truly associated SNPs (true positives) than the full set of SNPs.

Guidelines for Publishing

Guidelines for publishing association studies are evolving, along with the study designs and numbers of polymorphisms tested. Two principles that have emerged are that data should be presented in a way that facilitates replication and/or meta-analysis by other researchers and that the results for all polymorphisms tested (not just positive associations) should be made freely available. The first principle suggests that unique, universally recognized names for each polymorphism should be used in articles such as the reference SNP identifier (rs number) and that effect estimates (eg, regression parameters, or odds ratios) and their SEs should be presented, along with *P* values and genotype counts, for association tests. The criteria for publishing a genetic association study vary widely among journals and are beyond the scope of this review.

Conclusions

This primer has discussed some of the analytic features of genetic studies that differ from other epidemiological studies. Some features of large-scale candidate gene and genome-wide studies with emerging importance that have not been addressed include replication and validation studies and meta-analyses. Other articles in this series will provide reviews of family-based study designs and methods, including family-based (transmission) association tests.

Disclosures

None.

References

1. A haplotype map of the human genome. *Nature*. 2005;437:1299–1320.
2. Pollex RL, Hegele RA. Copy number variation in the human genome and its implications for cardiovascular disease. *Circulation*. 2007;115:3130–3138.
3. Barrett JC, Cardon LR. Evaluating coverage of genome-wide association studies. *Nat Genet*. 2006;38:659–662.
4. Evans DM, Cardon LR. A comparison of linkage disequilibrium patterns and estimated population recombination rates across multiple populations. *Am J Hum Genet*. 2005;76:681–687.
5. Shifman S, Kuypers J, Kokoris M, Yakir B, Darvasi A. Linkage disequilibrium patterns of the human genome across populations. *Hum Mol Genet*. 2003;12:771–776.
6. Wall JD, Pritchard JK. Haplotype blocks and linkage disequilibrium in the human genome. *Nat Rev Genet*. 2003;4:587–597.
7. Carlson CS, Eberle MA, Rieder MJ, Yi Q, Kruglyak L, Nickerson DA. Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. *Am J Hum Genet*. 2004;74:106–120.
8. Pritchard JK, Przeworski M. Linkage disequilibrium in humans: models and data. *Am J Hum Genet*. 2001;69:1–14.
9. Zhang K, Calabrese P, Nordborg M, Sun F. Haplotype block structure and its applications to association studies: power and study designs. *Am J Hum Genet*. 2002;71:1386–1394.
10. Crawford SL. Correlation and regression. *Circulation*. 2006;114:2083–2088.
11. Rao SR, Schoenfeld DA. Survival methods. *Circulation*. 2007;115:109–113.
12. Jewell NP. *Statistics for Epidemiology*. Boca Raton, Fla: Chapman & Hall/CRC; 2004.
13. Moskvina V, Craddock N, Holmans P, Owen MJ, O'Donovan MC. Effects of differential genotyping error rate on the type I error probability of case-control studies. *Hum Hered*. 2006;61:55–64.
14. Moskvina V, Schmidt KM. Susceptibility of biallelic haplotype and genotype frequencies to genotyping error. *Biometrics*. 2006;62:1116–1123.
15. Emigh TH. A comparison of tests for Hardy-Weinberg equilibrium. *Biometrics*. 1980;36:627–642.
16. Wigginton JE, Cutler DJ, Abecasis GR. A note on exact tests of Hardy-Weinberg equilibrium. *Am J Hum Genet*. 2005;76:887–893.

17. Sasieni PD. From genotypes to genes: doubling the sample size. *Biometrics*. 1997;53:1253–1261.
18. Shearman AM, Cupples LA, Demissie S, Peter I, Schmid CH, Karas RH, Mendelsohn ME, Housman DE, Levy D. Association between estrogen receptor alpha gene variation and cardiovascular disease. *JAMA*. 2003; 290:2263–2270.
19. Kannel WB, Feinleib M, McNamara PM, Garrison RJ, Castelli WP. An investigation of coronary heart disease in families: the Framingham Offspring Study. *Am J Epidemiol*. 1979;110:281–290.
20. Ardlie KG, Lunetta KL, Seielstad M. Testing for population subdivision and association in four case-control studies. *Am J Hum Genet*. 2002;71: 304–311.
21. Devlin B, Roeder K. Genomic control for association studies. *Biometrics*. 1999;55:997–1004.
22. Campbell CD, Ogburn EL, Lunetta KL, Lyon HN, Freedman ML, Groop LC, Altshuler D, Ardlie KG, Hirschhorn JN. Demonstrating stratification in a European American population. *Nat Genet*. 2005;37:868–872.
23. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81:559–575.
24. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. *PLoS Genet*. 2006;2:e190.
25. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*. 2006;38:904–909.
26. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics*. 2000;155:945–959.
27. Pritchard JK, Stephens M, Rosenberg NA, Donnelly P. Association mapping in structured populations. *Am J Hum Genet*. 2000;67:170–181.
28. Nyholt DR. A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. *Am J Hum Genet*. 2004;74:765–769.
29. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Royal Stat Soc Series B*. 1995;57:289–300.
30. Genovese R, Roeder K, Wasserman L. False discovery control with p-value weighting. *Biometrika*. 2006;93:509–524.
31. Roeder K, Bacanu SA, Wasserman L, Devlin B. Using linkage genome scans to improve power of association in genome scans. *Am J Hum Genet*. 2006;78:243–252.
32. Greenwood CM, Rangrej J, Sun L. Optimal selection of markers for validation or replication from genome-wide association studies. *Genet Epidemiol*. 2007;31:396–407.

KEY WORDS: genetics ■ statistics ■ epidemiology