Pooling Analysis of Genetic Data: The Association of Leptin Receptor (*LEPR*) Polymorphisms With Variables Related to Human Adiposity

M. Heo,* R. L. Leibel,† B. B. Boyer,‡ W. K. Chung,† M. Koulu,§ M. K. Karvonen,§ U. Pesonen,§ A. Rissanen, M. Laakso,†† M. I. J. Uusitupa,†† Y. Chagnon,‡‡ C. Bouchard,§§ P. A. Donohoue,*** T. L. Burns,††† A. R. Shuldiner,‡‡‡ K. Silver,‡‡‡ R. E. Andersen,§§§ O. Pedersen,§§§ S. Echwald,**** T. I. A. Sørensen,†††† P. Behn,‡‡‡‡ M. A. Permutt,§§§§ K. B. Jacobs,***** R. C. Elston,***** D. J. Hoffman* and D. B. Allison†††††**

**New York Obesity Research Center, Columbia University College of Physicians and Surgeons, New York, New York 10025, †Department of Pediatrics, Columbia University College of Physicians and Surgeons, New York, New York 10032, ‡Department of Molecular Biology, University of Alaska, Fairbanks, Alaska 99775, §Department of Pharmacology and Clinical Pharmacology, University of Turku, 20520 Turku, Finland, **Eating Disorder Unit, University of Helsinki, 00014 Helsinki, Finland, ††Department of Medicine, University of Kuopio, 70210 Kuopio, Finland, ‡‡Kinesiologie, Physical Activity Sciences Lab, Laval University, Ste-Foy, Quebec G1K7P4, Canada, §§Pennington Biomedical Research Center, Louisiana State University, Baton Rouge, Louisiana 70808, ***Department of Pediatrics, University of Iowa College of Medicine, Iowa City, Iowa 52242, †††Department of Preventive Medicine, University of Iowa College of Public Health, Iowa City, Iowa 52242, ‡‡‡Division of Endocrinology, University of Maryland, Baltimore, Maryland 21224, §§§Department of Geriatrics, Johns Hopkins University, Baltimore, Maryland 21224, ****Steno Diabetes Center, Hagedorn Research Institute, DK-2820 Gentofte, Denmark, ††††Danish Epidemiology Science Center, Institute of Preventive Medicine, Copenhagen 1399, Denmark, ‡‡‡‡Internal Medicine Associates, Bloomington, Indiana 47403, §§§§Department of Internal Medicine, Washington University School of Medicine, St. Louis, Missouri 63110, *****Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, Ohio 44109 and †††††Department of Biostatistics, University of Alabama, Birmingham, Alabama 35294-0022*

Manuscript received October 30, 2000
Accepted for publication July 5, 2001

ABSTRACT

Analysis of raw pooled data from distinct studies of a single question generates a single statistical conclusion with greater power and precision than conventional metaanalysis based on within-study estimates. However, conducting analyses with pooled genetic data, in particular, is a daunting task that raises important statistical issues. In the process of analyzing data pooled from nine studies on the human leptin receptor (*LEPR*) gene for the association of three alleles (*K109R*, *Q223R*, and *K656N*) of *LEPR* with body mass index (BMI; kilograms divided by the square of the height in meters) and waist circumference (WC), we encountered the following methodological challenges: data on relatives, missing data, multivariate analysis, multiallele analysis at multiple loci, heterogeneity, and epistasis. We propose herein statistical methods and procedures to deal with such issues. With a total of 3263 related and unrelated subjects from diverse ethnic backgrounds such as African-American, Caucasian, Danish, Finnish, French-Canadian, and Nigerian, we tested effects of individual alleles; joint effects of alleles at multiple loci; epistatic effects among alleles at different loci; effect modification by age, sex, diabetes, and ethnicity; and pleiotropic genotype effects on BMI and WC. The statistical methodologies were applied, before and after multiple imputation of missing observations, to pooled data as well as to individual data sets for estimates from each study, the latter leading to a metaanalysis. The results from the metaanalysis and the pooling analysis showed that none of the effects were significant at the 0.05 level of significance. Heterogeneity tests showed that the variations of the nonsignificant effects are within the range of sampling variation. Although certain genotypic effects could be population specific, there was no statistically compelling evidence that any of the three *LEPR* alleles is associated with BMI or waist circumference in the general population.

WHEN many studies on the same topic differ in terms of statistical inferences and conclusions thereof, combining the information from these separate studies by either metaanalysis or raw data pooling provides a means by which data from the individual studies can be combined to enhance statistical power. The primary advantages of such analyses include (1) reduction of type I errors by consolidating many tests of the same hypothesis with many samples into a single test with one pooled sample; (2) increased statistical power; and (3) direct tests of heterogeneity among samples/populations.

In genetic studies, analysis of pooled data can be especially challenging even if all studies investigated relationships between genotypes at the same gene with the same phenotypes. Genetic data sets usually consist of individuals related to other individuals, resulting in correlated

Corresponding author: David B. Allison, Department of Biostatistics, School of Public Health, University of Alabama, RPHB 327M, 1530 Third Ave. S., Birmingham, AL 35294-0022.
E-mail: dallison@ms.soph.uab.edu

observations. Therefore, methodologies accounting for such correlated observations should be employed. Typically, familial correlations depend upon the degree of relationship between pedigree members. With respect to genotypes, multiple alleles across multiple loci, or markers, within the same gene may be of interest, leading to the need to evaluate multilocus analyses for direct and epistatic effects, *i.e.*, interactions among the multiple alleles. Interactions among genotypes and covariates (*i.e.*, gene-by-environment interaction or effect modification) should also be considered in modeling; the use of appropriate covariates can define more precisely the effects of individual alleles. Pleiotropic (either relational or mosaic) effects of genotypes are also of interest and can be investigated through multivariate analysis. Finally, missing observations on genotypes and other variables are not uncommon.

Codes for the same genotypes and discrete variables are often different from study to study. In addition, there may be increased numbers of missing observations in pooled data because of different lists of covariates and alleles. The number of members in each pedigree is usually different within and among studies. The configurations of degrees of relationship among pedigree members are not the same over the different pedigrees. These characteristics of the pooled data require statistical modeling, allowing flexible construction of the residual covariance matrix. Moreover, the number of independent variables in a model can be very large when all the main effects and interaction effects are included. Heterogeneity of samples (*e.g.*, in demographic characteristics) among the different studies is also a concern and necessitates analysis adjusted for study profiles or study effects. To our knowledge, neither the guidelines nor the statistical software for handling such methodological and practical issues are currently well developed in the genetics literature, though some guidelines do exist in other contexts (*e.g.*, Cooper and Hedges 1994).

In this article, we illustrate these issues and demonstrate some appropriate (and in some cases *ad hoc*) statistical methodologies and procedures. We used pooled raw data on body mass index (BMI; measured as the weight in kilograms divided by the square of the height in meters) and waist circumference (WC) and the three amino acid substitutions (the polymorphisms or the allelic variants) *K109R* at exon 2, *Q223R* at exon 4, and *K656N* at exon 12 in the leptin receptor (*LEPR*) gene on human chromosome 1p. In the rest of this article, to denote any variant at a particular exon, we use the word "allele" rather than "polymorphism" in accordance with Elston (2000). The exon-coding nomenclature and the marker names for the alleles used are as in Chung *et al.* (1997). The alleles at each exon within the *LEPR* gene are diallelic and the *LEPR* gene is a candidate gene influencing human obesity.

Interindividual variation in human BMI, direct measures of fat mass, and fat distribution have been clearly shown to have a strong genetic component (Comuzzie and Allison 1998). Mice homozygous for inactivating mutations of *Lepr* become massively obese (Chua *et al.* 1996). Rare instances of human obesity secondary to homozygosity for inactivating *LEPR* mutant alleles have been identified (Clément *et al.* 1998). Recently, several studies have assessed the role of the three common *LEPR* alleles in interindividual variations in BMI and WC. These analyses yielded inconsistent results (*e.g.*, Gotoda *et al.* 1997; Matsuoka *et al.* 1997; Silver *et al.* 1997; Chagnon *et al.* 1999; del Giudice *et al.* 2000). In an attempt to resolve the physiological significance of the alleles at the *LEPR* gene with obesity-related phenotypes, teams of investigators representing nine studies provided their raw data for a pooling analysis (see appendix a).

SUBJECTS

A total of 3263 individuals were included in this study (Table 1). Sixty-two percent of these individuals are related to one or more subjects in the data set. The largest number of generations among the family pedigrees in the pooled data was two. Descriptive statistics are presented in Table 2 along with percentages of missing observations. The subjects are ethnically diverse—*i.e.*, African-American, Caucasian, Danish, Finnish, French-Canadian, and Nigerian. Approximately half are female (Table 2).

STATISTICAL ANALYSIS

General model: In general, genetic models can be represented in the following form:

$$f(\text{phenotype(s)}) = \text{genetic effects} + \text{main covariate effects} + \text{interaction effects} + \text{error}.$$

The function f of the phenotype(s) depends upon the model implemented. For example, f(phenotype(s)) is the squared difference of phenotypes of sibling pairs in the Haseman-Elston regression. The phenotypes can be univariate or multivariate. Genetic effects can be random as in variance components analysis or fixed as in usual association studies. Covariates are discrete and/or continuous, and their coefficients are usually fixed, although modeling random coefficients is possible. The model becomes a mixed-effects model when fixed and random effects are simultaneously included. Interaction effects can be among genetic effects (*i.e.*, locus-by-locus interaction, epistasis), among main covariates, and between genetic and main covariates (*i.e.*, locus-by-environment interactions). The expectation of the vector error term is zero but its covariance may not necessarily be in the form of $\sigma\mathbf{I}$, where $\mathbf{I}$ is the n-by-n identity matrix and n is the number of subjects under study.

Preliminary analysis: The main objective of the preliminary analysis was to test the association between alleles at each exon and BMI and WC. We analyzed BMI and WC separately for each exon. The main covariates were sex and age and no interaction effects were modeled. In this preliminary analysis, no missing data imputation was employed. The purpose of these approaches was to overview a crude overall association.

For the association analysis, we utilized the ASSOC routine in the S.A.G.E. (1997) software (George and Elston 1987), which allows for familial (residual) correlations. The sibling-based permutation test, a valid joint test for association in the presence of linkage (Allison *et al.* 1999), was also utilized to control for potential population admixture. The permutation of observations was performed within each sibship, which therefore leads to inferences conditional on sibship and hence on family. This step eliminates the possibility of confounding by population admixture. In other words, no matter what population the families came from, inferences conditional upon sibship are equivalent to those conditional upon population stratification, because sibships are "subunits" of a population with a uniform degree of admixture. This has been more fully described and demonstrated elsewhere (*e.g.*, Allison *et al.* 1999).

We also analyzed the proportion of alleles shared identical-in-state (IIS) by sibling pairs because the alleles at the three exons are of interest. Specifically, we regressed the square of the phenotypic difference between the sibling pairs (*cf.* Haseman-Elston procedure; Haseman and Elston 1972) on the proportion of alleles shared IIS by sibling pairs. We also regressed the grand-mean-centered cross product of phenotypes between the sibling pairs (*cf.* New Haseman-Elston procedure; Elston *et al.* 2000) on the proportion of alleles shared IIS by sibling pairs. This type of "IIS analysis" is also valid for association and linkage analysis. We used IIS instead of identical-by-descent (IBD) because, when the marker alleles are causative of variations in the phenotype, IIS linkage analysis should be more powerful than IBD linkage analysis and we were not interested in whether the phenotypes of interest are linked to unknown alleles (in which case IBD analysis would have been more powerful). All of these analyses were adjusted for age and sex.

For this preliminary analysis we took two methodological approaches: metaanalysis and pooling analysis. For the former, we applied the association analyses described above to each data set from each study and then "metaanalyzed" the results followed by heterogeneity analysis (Hedges and Olkin 1985). To combine the *P* values, we used Fisher's approach (Fisher 1954). For the latter, we applied association and linkage analyses to a single pooled data set. Finally, we compared the results of the metaanalysis with those obtained from the pooling analysis.

Main analysis: The objectives of the main analysis were to test the simultaneous effects of alleles in multiple exons, to test the epistatic effects of alleles in these exons, and to test effect modification by main covariates such as age, sex, diabetes, and ethnicity. BMI and WC were modeled as separate univariate phenotypes and as a combined multivariate phenotype. All exons were simultaneously included in all models. The main covariates included continuous age polynomial variables up to the third degree (*i.e.*, age, age^2, age^3). In addition, discrete variables (ethnicity, diabetes, sex, and study effects) were dummy coded. The powers of the variable were included because the range of age was wide (3–94 years), and BMI and WC are nonlinearly associated with age over that range, especially in children (Rolland-Cachera *et al.* 1982).

Two- and three-way interaction effects among genotypes and two-way interaction effects between genotype and main covariates were included in the analysis. However, with regard to two-way interaction effects, age^2 and age^3 were excluded to limit the number of independent variables in the models. The study effects were included only as main covariates without any related interaction terms. We used the S.A.G.E. ASSOC routine, fitted ordinary least squares (OLS) regression models, and conducted general linear model (GLM) multivariate analyses for the main analyses. In this main analysis, only pooling analyses were conducted. In the following section, the theoretical and practical issues that arose in conducting the main analysis are described.

Multiple imputation for missing values: There were many missing observations for the phenotypes and covariates, *e.g.*, diabetic status (Table 2), and for genotypes at exon 2 in particular (Table 3). Because deletion of missing values (*e.g.*, list-wise deletion) from analyses can introduce biases and inefficient use of collected data (Schafer 1997), we employed the multiple imputation (MI) method proposed by Rubin (1978, 1987, 1996) and described in Schafer (1997), assuming that the missing observations occurred at random as defined in Little and Rubin (1987). Briefly, imputed values for one missing observation are randomly drawn multiple times independently from an underlying probability model, *e.g.*, from a normal distribution whose mean and variance were determined by regression analysis; *i.e.*, the predicted value is the mean, and the mean squared error (MSE) is the variance. For the nonmissing observations, imputation is not necessary.

The MI method is appealing because it accommodates variation due to random imputation in the inference procedure (note that single imputed values are not true observations). In principle, more imputations provide for better inference with increased accuracy. Practically, however, three to five imputations appear satisfactory in terms of efficiency of estimation even with 50% observations missing (Schafer and Olsen 1998). Specifically, the efficiency of estimation by means of MI is $1/(1 + \gamma/m)$, where γ is the fraction of missing

TABLE 1

Data sets for the pooling analysis

Study ID	Name[a] of study	No. of subjects	Ethnicity	Reference
1	Heritage family study (HFS)	857 *related* subjects from 215 families	Caucasian, African-American	Chagnon *et al.* (2000)
2	Quebec family study (QFS)	687 *related* subjects from 175 families	Caucasian	Chagnon *et al.* (1999)
3	African-American (AfAm) study	228 unrelated	African-American	Our unpublished data
4	Nigerian study	221 unrelated	African	Tanizawa *et al.* (1994)
5	Finnish study 2	112 unrelated	Caucasian	Oksanen *et al.* (1998)
6	Baltimore study	251 unrelated	Caucasian, African-American, Hispanic[b]	Silver *et al.* (1997)
7	Finnish study 1	65 unrelated	Caucasian	Oksanen *et al.* (1998)
8	Muscatine Iowa family study (MIFS)	482 *related* subjects from 112 families[c]	Caucasian	Donohoue *et al.* (2000)
9	Danish study	360 unrelated	Caucasian	Echwald *et al.* (1997)
	Total	3263		

[a] These are not necessarily official names: see appendix a for the data providers.
[b] There were only two Hispanic subjects, who were excluded from the present analysis.
[c] The number of families with an obese child from over 400 families in this study.

information and m is the number of imputations (Rubin 1987). This returns 86 and 91% efficiency for $m = 3$ and 5, respectively, with $\gamma = 0.5$, *i.e.*, 50% observations missing. Herein the missing observations were imputed five times.

Random values from the normal distribution can be imputed for missing values of continuous variables. For missing discrete categorical values, however, the categorical values nearest to the randomly generated continuous values can be imputed. This imputation method for the categorical variables is acceptable in general even under normality assumptions (Schafer and Olsen 1998). However, this technique does not guarantee producing imputations with coherent genotypes (*i.e.*, genotypes satisfying Mendel's laws) for members within a family of multiple generations. To avoid possible inconsistencies of imputed genotypes among family members, we employed "restricted" imputation. For instance, when all family members had missing genotypes, only offspring genotypes were imputed for computational simplicity and convenience, or the "aa" genotype was excluded from possible offspring genotypes when the parental mating type was AA-AA or AA-Aa. See appendix b for statistical inferences that include *P*-value calculations on the basis of MI.

Practical issues and ad hoc methods: The interaction terms and dummy codes for genotypes and discrete variables created models with many independent variables, making interpretation of results difficult. Therefore, we applied OLS backward elimination to select significant variables in the presence of genotype variables. In doing so, we first "stacked" the five imputed pooled data sets into one single data set and then assigned a weight of one-fifth in count to every data point in the stacked data set, so that original nonmissing values have a weight of 1 and imputed values have a weight of $\frac{1}{5}$. By "stack" we mean accumulation of (the five rectangular) missing-imputed data sets by case. This procedure still does not account for familial correlations or variations due to MIs and therefore does not produce unbiased standard errors. However, this *ad hoc* method can help identify significant contributors to specific models; the OLS method provides unbiased coefficient estimates and reasonably good compatibility for testing compared to its counterpart that adjusts for correlations (M. A. Province, T. Rice and D. C. Rao, unpublished results). Therefore, for subsequent statistical analyses we took residualized BMI and WC resulting from OLS regression on some "significant" variables retained from the backward elimination using the stacked data set.

Although the S.A.G.E. ASSOC routine was developed to allow familial correlations and to produce consistent standard error estimates, it has several practical limitations in terms of the number of interactions that can be included. However, one can include more than one locus by coding the other loci as covariates since ASSOC allows multiple covariates to be included in the model.

TABLE 2

Descriptive statistics from a total of 3263 observations

Continuous variables	Mean ± SD	(min, max)	% missing
Age (year)	38.2 ± 15.8	(3, 94)	2.4
BMI (kg/m^2)	28.2 ± 7.2	(14, 77)	1.2
WC (cm)	90.2 ± 18.4	(33, 178)	22.9
Discrete variables	**Category**	**%[a]**	**% missing**
Sex	Female	50.9	<0.05
Diabetes	Yes	24.4	64.0
Ethnicity	Caucasian	74.7	None
	African-American	18.4	
	African	6.8	
	Hispanic	0.1	

[a] Out of nonmissing observations.

ASSOC provides the estimated coefficients for each covariate and their standard errors. Through the use of the statistics, we can test the effect of the additional loci by a Wald-type test. The choice of which locus is utilized as a genetic locus ("marker" in the S.A.G.E. ASSOC terminology) and which loci are treated as covariates should not affect the inferences about genotypic effects at each locus because the Wald (for the covariates) and likelihood-ratio tests (for the genetic locus, the marker) are asymptotically equivalent. On the other hand, most programs for OLS regression are easy and flexible to run and produce unbiased point estimates, but they do produce biased estimates of standard errors. To adjust the bias of the standard errors of the OLS regression, we calculated a "correction factor": the ratio of average estimated variances of the point estimates from S.A.G.E. ASSOC to that of estimated variances from the OLS regression. This calculation resulted in "corrected" flexible use of OLS in various model-fitting procedures in the presence of correlations among observations across data points. We also tried the OLS approach with the stacked data, using weights of $\frac{1}{5}$ to compare the estimates and *P* values obtained from the MI inference described in APPENDIX B.

With respect to multivariate analysis for testing pleiotropic effects, many approaches can be employed (*e.g.*, AMOS *et al.* 1990; ALLISON *et al.* 1998; MANGIN *et al.* 1998). In this study, we employed GLM multivariate analysis without controlling for familial correlations, because *P* values could be adjusted by using the correction factors described above.

RESULTS

Descriptive statistics: Demographic descriptive statistics are presented in Tables 1 and 2 along with percentages of missing observations. The range of age is large, as are the ranges of BMI and WC. The estimated allele frequencies at exons 2, 4, and 12 were 0.23, 0.48, and 0.20, respectively (Table 3). Using the maximum-likelihood (ML) test for departure from Hardy-Weinberg equilibrium (HWE) described by LYNCH and WALSH (1998, pp. 60–61) there was no evidence for a departure from HWE for alleles at any exon (exon 2, $P = 0.285$; exon 4, $P = 0.597$; and exon 12, $P = 0.537$). The MI resulted in little change in statistical significance (exon 2, $P = 0.960$; exon 4, $P = 0.770$; and exon 12, $P = 0.374$). On the basis of the ML test in LYNCH and WALSH (1998, pp. 98–99), however, the three exons are (as expected) all significantly in pairwise linkage disequilibrium regardless of the imputation (exon 2 *vs.* exon 4, $P < 0.001$; exon 2 *vs.* exon 12, $P < 0.001$; and exon 4 *vs.* exon 12, $P < 0.001$).

Table 3 also presents genotype-specific mean BMI and WC by locus. Within each exon, the mean values are almost the same before or after imputation. However, before imputation, overall means of BMI and WC at exon 2 are smaller than those at the other exons because the genotypes at exon 2 in the Baltimore study (Table 1) are all missing and Baltimore study subjects have relatively larger BMI and WC compared to subjects in the other studies. After imputation, on the other hand, the overall mean values become almost the same across the exons. These results, unadjusted for covariates or familial correlations, provide a preliminary view of the association between the phenotypes and genotypes.

Preliminary analysis: Results of the ASSOC analyses are presented in terms of differences of the estimated effects of the two genotypes, "wild-type" homozygote and heterozygote, on BMI and WC, separately, from those of the "mutant" homozygous genotypes (Table 4) after adjusting for age and sex. For example, subjects heterozygous (*K109R*) for the exon 2 allele had a meta-estimate effect size of 0.03 on BMI when compared to subjects with *K109K* genotype. No single effect was significant from either the individual studies or the metaanalysis (Table 4). The number of subjects for each estimate is presented in Table 5, which also shows *P*

TABLE 3

Proportion, mean, and SD of BMI and WC by genotypes

Locus	Genotypes	Proportions (%)[a]		Mean BMI ± SD (kg/m^2)		Mean WC ± SD (cm)	
		Before imputation	After imputation	Before imputation	After imputation	Before imputation	After imputation
Exon 2	*K109K*	59.8	58.8	27.3 ± 6.1	28.0 ± 7.1	88.0 ± 16.2	90.8 ± 17.8
	K109R	34.6	35.8	27.5 ± 6.5	28.4 ± 7.6	89.3 ± 16.8	91.8 ± 18.8
	R109R	5.6	5.4	26.9 ± 5.8	27.4 ± 6.4	87.4 ± 16.8	89.3 ± 17.5
	Total	13.9[b]	1.6[b]	27.4 ± 6.3	28.1 ± 7.2	88.4 ± 16.5	91.1 ± 18.1
Exon 4	*Q223Q*	27.0	26.9	28.0 ± 7.2	27.9 ± 7.2	90.3 ± 18.9	90.8 ± 18.1
	Q223R	49.5	49.7	28.3 ± 7.3	28.2 ± 7.3	90.3 ± 18.6	91.1 ± 17.5
	R223R	23.6	23.4	28.4 ± 7.2	28.3 ± 7.2	91.1 ± 17.7	91.2 ± 17.5
	Total	4.8[b]	1.1[b]	28.2 ± 7.3	28.1 ± 7.2	90.5 ± 18.5	91.0 ± 18.1
Exon 12	*K656K*	63.8	63.7	28.3 ± 7.4	28.2 ± 7.4	90.7 ± 18.7	91.3 ± 18.5
	K656N	32.4	32.5	28.1 ± 6.9	28.0 ± 6.9	89.7 ± 17.7	90.5 ± 17.5
	N656N	3.8	3.8	28.3 ± 7.1	28.2 ± 7.1	91.5 ± 18.0	91.6 ± 16.5
	Total	4.8[b]	1.1[b]	28.2 ± 7.2	28.1 ± 7.2	90.4 ± 18.4	91.0 ± 18.1

[a] Proportions of genotypes calculated from nonmissing observations. The proportion of missing values was calculated from all subjects.

[b] Percentage of total observations missing.

values to assess the significance of the phenotypic variation due to genotypic variation, by exon. The calculation of these *P* values was based on the contribution of the genotypic variation to the likelihood. The effects were again not significant for either phenotype for any exon from individual studies or from the metaanalysis. Table 6 presents results of the sibling-based permutation association test from individual studies, from the metaanalysis, and from the pooling analysis. In this sibling-based permutation association test, only studies with related subjects were analyzed, because the test requires siblings within each family. Table 6 also shows absence of significant genotypic effects on the phenotypes.

Results of the regression on proportions of alleles shared IIS are presented in Tables 7 and 8. These results were obtained from the regression of squared phenotype differences and the regression of grand-mean centered phenotype cross products, respectively. There was no statistical evidence that the alleles examined are associated with (or linked to) significant variation in either phenotype. However, the result of such IIS analysis of one study (Quebec family study) showed a significant linkage of the *Q223R* allele to BMI ($P = 0.04$), which is in agreement with the result of the IBD linkage analysis reported in Chagnon *et al.* (1999; $P = 0.02$).

Heterogeneity analysis: Heterogeneity analysis (Tables 4, 7, and 8) showed that the nonsignificant results are consistent over the three loci and over all of the individual studies, which implies that the variation in effects over the studies is within the range of sampling variation.

Main analysis: *Missing data imputation:* The results of missing data imputation are presented in Table 3 with a comparison of descriptive statistics before and after imputation, as presented earlier. Due to the "restricted" imputation adopted herein, ~5% of subjects still have missing genotypes for at least one exon even after imputation. Mean BMIs and WCs over the genotypes at exon 2 were increased after imputation. However, frequencies of alleles at each exon changed little.

Backward elimination and effect modifications: Significance of various effect modifications was evaluated by testing for interaction effects among genotypes and main covariates. We tested such interaction effects by OLS regression with the weighted stacked data. Backward elimination—starting with a full model with main genotype effects, epistatic effects (*i.e.*, interaction effects among genotypes), main covariate effects, and interaction among genotypes and covariates—was applied to identify significant interaction effects. The OLS backward elimination results are listed in Table 9.

Only the interaction effect between *R109R* at exon 2 and sex was significant for BMI, implying that male subjects with *R109R* genotype at exon 2 have significantly higher BMI than the other subjects. However, the contribution of this interaction effect to the variations of BMI is minimal (increase in $R^2 < 0.01\%$). The nonsignificant allele-by-environment interaction effects suggest that the genotypic effects, if any, might not be modified by the main covariates such as diabetes, sex, age, and ethnicity. In terms of epistatic effects, however, subjects with *K109R* at exon 2 and *N656N* at exon 12 (see HT2HM12 in Table 9) appeared to have significantly higher BMIs. Age, sex, ethnicity, diabetes, and study effects are major contributors to the variation of BMI and WC. Therefore, we took the "residualized," or adjusted, BMI and WC for sex, diabetes, ethnicity, age polynomials, and study effects for the following analyses.

Joint effects of multiple alleles at different loci and epistasis: The results for BMI from the S.A.G.E. ASSOC routine

TABLE 4

Estimated effects (with reference to mutant homozygotes) of each genotype at each locus adjusted for age and sex (and their standard errors)

	Exon 2		Exon 4		Exon 12	
Study	*K109R* (SE)	*R109R* (SE)	*Q223R* (SE)	*R223R* (SE)	*K656N* (SE)	*N656N* (SE)
BMI						
Finnish 1	0.39 (0.529)	0.90 (0.574)	0.40 (0.612)	1.03 (0.637)	−0.24 (0.517)	−0.55 (1.308)
Finnish 2	0.90 (0.757)	0.44 (1.147)	0.12 (0.929)	0.96 (1.012)	−0.24 (0.821)	0.27 (3.775)
QFS	0.12 (0.499)	−0.15 (0.933)	−0.83 (0.541)	−0.30 (0.745)	0.45 (0.545)	1.13 (1.137)
HFS	0.07 (0.407)	−0.20 (0.818)	0.65 (0.422)	0.73 (0.548)	0.24 (0.399)	−0.86 (1.123)
MIFS	−0.51 (0.625)	−1.45 (1.426)	−0.46 (0.654)	−0.32 (0.860)	−0.14 (0.590)	1.06 (1.969)
Baltimore	—	—	0.77 (1.448)	0.76 (1.766)	−0.98 (1.300)	0.84 (3.585)
Danish	−0.90 (0.678)	1.17 (1.974)	0.35 (0.670)	1.58 (0.892)	−0.24 (0.639)	0.64 (1.239)
Nigerian	−0.52 (—[a])	0.00 (—[a])	−0.76 (0.996)	−0.53 (1.017)	0.55 (0.710)	3.12 (1.796)
AfAm	0.13 (1.045)	5.24 (5.043)	1.80 (1.187)	−0.06 (1.251)	0.07 (0.988)	−1.51 (1.858)
Meta est[b]	0.03 (0.22), $P = 0.892$	0.33 (0.37), $P = 0.372$	0.13 (0.23), $P = 0.571$	0.50 (0.28), $P = 0.074$	0.06 (0.21), $P = 0.775$	0.31 (0.51), $P = 0.543$
Q[c]	4.46, d.f. = 6, $P = 0.615$	4.37, d.f. = 6, $P = 0.626$	8.76, d.f. = 8, $P = 0.363$	5.85, d.f. = 8, $P = 0.664$	2.64, d.f. = 8, $P = 0.955$	5.68, d.f. = 8, $P = 0.683$
WC						
Finnish 1	—	—	—	—	—	—
Finnish 2	3.92 (2.023)	1.63 (3.045)	0.46 (2.469)	3.34 (2.700)	0.50 (2.223)	2.21 (10.079)
QFS	0.65 (1.224)	−0.73 (2.248)	−0.84 (1.317)	−0.28 (1.808)	1.05 (1.336)	2.51 (2.764)
HFS	0.67 (1.057)	0.03 (2.111)	1.07 (1.102)	1.70 (1.416)	0.72 (1.040)	−0.76 (2.987)
MIFS	−1.14 (1.593)	−4.98 (3.623)	−1.71 (1.666)	−3.88 (2.180)	−0.84 (1.513)	4.58 (5.007)
Baltimore	—	—	−1.52 (3.216)	−3.58 (3.913)	−0.91 (2.908)	−2.36 (7.412)
Danish	—	—	—	—	—	—
Nigerian	−0.67 (—[a])	0.00 (—[a])	−3.94 (2.630)	−3.46 (2.676)	0.23 (1.885)	10.02 (4.528)
AfAm	—	—	—	—	—	—
Meta est[b]	0.70 (0.67), $P = 0.296$	−0.56 (1.28), $P = 0.662$	−0.40 (0.68), $P = 0.556$	−0.23 (0.86), $P = 0.789$	0.37 (0.63), $P = 0.557$	2.48 (1.67), $P = 0.137$
Q[c]	3.87, d.f. = 3, $P = 0.276$	2.09, d.f. = 3, $P = 0.554$	4.56, d.f. = 5, $P = 0.472$	8.60, d.f. = 5, $P = 0.126$	1.21, d.f. = 5, $P = 0.944$	4.55, d.f. = 5, $P = 0.473$

[a] Failed to converge and excluded from metaanalysis.
[b] Metapooled estimate weighted by standard errors (Hedges and Olkin 1985).
[c] χ^2 statistics for testing heterogeneity of effect sizes (Hedges and Olkin 1985).

are displayed in Table 10 with the five imputed pooled data sets. No single genotype at any exon has significant effect on BMI. The joint effects of all the genotypes at all exons are also not significant. These results are consistent with those from OLS regression without controlling for the familial correlations (Table 10). Similar results were observed (Table 11) with respect to WC. Interestingly, estimated coefficients, their standard errors, and *P* values for individual genotypes obtained from the OLS regressions after employing the estimation process described in appendix b were almost exactly the same as those from the OLS regressions with the stacked data (with weights of ⅕; Tables 10 and 11).

The correction factors from both BMI and WC analyses were, however, <1, which is counterintuitive, implying that the standard errors estimated without controlling for familial correlations are bigger than those that include such controlling. In regard to testing epistasis, *i.e.*, two- and three-way interaction effects among genotypes over all exons, we applied the OLS regression to accommodate all of the interactions in a particular model. The results are displayed in terms of *P* values for simultaneous effect of such interactions in Table 12. This analysis shows that the epistatic effects among genotypes of the three exons are not jointly significant from the results of MI nor from the results of weighted OLS, regardless of the presence of the main genotype effects in models after adjusting for sex, diabetes, ethnicity, age polynomials, and study effects.

Multivariate analysis for pleiotropic effects: The results for testing the pleiotropic effects are shown in Table 13. Although *P* values vary with imputation, no main effects or interaction effects are significant. These results indicate that the polymorphisms do not have any statistically significant simultaneous effects on BMI and WC.

DISCUSSION

We presented practical and *ad hoc* statistical methodologies that can accommodate many of the challenges

TABLE 5

***P* values of joint effects of the three genotypes at each locus on the phenotypes from association tests adjusted for age and sex**

	BMI			WC		
Study	Exon 2 (*n*)	Exon 4 (*n*)	Exon 12 (*n*)	Exon 2 (*n*)	Exon 4 (*n*)	Exon 12 (*n*)
Finnish 1	0.297 (63)	0.229 (65)	0.839 (65)	—	—	—
Finnish 2	0.443 (112)	0.514 (112)	0.956 (112)	0.147 (110)	0.333 (110)	0.956 (110)
QFS	0.932 (649)	0.257 (652)	0.527 (653)	0.730 (630)	0.791 (633)	0.574 (634)
HFS	0.937 (776)	0.273 (775)	0.541 (776)	0.807 (775)	0.465 (774)	0.719 (775)
MIFS	0.499 (403)	0.783 (434)	0.819 (436)	0.364 (402)	0.209 (434)	0.499 (436)
Baltimore	—	0.861 (245)	0.715 (245)	—	0.657 (222)	0.914 (222)
Danish	0.320 (358)	0.198 (358)	0.775 (358)	—	—	—
Nigerian	0.791 (197)	0.745 (203)	0.180 (204)	0.946 (191)	0.317 (197)	0.102 (196)
AfAm	0.557 (207)	0.131 (214)	0.708 (213)	—	—	—
Overall[a]	0.885	0.437	0.967	0.723	0.575	0.786

From χ^2 test with 2 d.f. from the likelihood-ratio test.
[a] From Fisher combination of *P* values (Fisher 1954).

encountered in pooled genetic data analysis. Although data pooling may be ideal for enhancement of power of statistical inferences, it is a daunting task to manage and analyze pooled data, as we have seen so far. Even a task as seemingly simple as coordinating the pooling of different data sets by creating a coherent coding system and uniform variable names can prove to be time consuming. Developing a coherent coding system for the pedigree members is important and creating appropriate dummy family members is often required for application of software. Therefore, investigators planning a pooling study should be aware of the large amount of time required for data management before the pooled data are analyzed.

In addition, the availability of statistical software that can handle the analytic problems raised here is limited. This deficiency creates a particular type of problem because, even after all data management issues are resolved, analysts are sometimes forced to change software and/or write additional programs with specific computer languages to conduct appropriate and necessary procedures and analyses. For example, one statistical software package may be able to provide a multiple-imputed data set but not be able to generate the estimates from the imputed data using advanced statistical analytical approaches. To obtain such estimates, data analysts generally need to apply a different software package to the multiple-imputed data; it is important to note that some types of software are not compatible with all forms of electronic data. This emphasizes another consideration. While there are many forms of genetic data analysis software available, few are flexible

TABLE 6

Results from sibling-based permutation test for association adjusted for age and sex

Phenotype	Study	Exon 2	Exon 4	Exon 12
		Pooling analysis: χ^{2a}, *P* (n^b)		
BMI		3.06, 0.080 (1067)	1.73, 0.189 (1092)	0.25, 0.619 (1095)
WC		2.01, 0.155 (1065)	1.37, 0.242 (1092)	1.85, 0.174 (1095)
		Analysis by study and metaanalysis: χ^2, *P* (*n*)		
BMI	QFS	0.53, 0.465 (379)	0.53, 0.469 (381)	0.02, 0.878 (383)
	HFS	3.00, 0.083 (476)	1.09, 0.296 (476)	0.28, 0.601 (476)
	MIFS	0.84, 0.359 (212)	0.35, 0.555 (235)	0.05, 0.827 (236)
	Overall[c]	$P = 0.200$	$P = 0.528$	$P = 0.948$
WC	QFS	0.274, 0.601 (379)	0.09, 0.761 (381)	0.79, 0.375 (383)
	HFS	1.59, 0.208 (475)	1.49, 0.223 (475)	1.28, 0.259 (475)
	MIFS	0.83, 0.364 (211)	0.38, 0.538 (236)	0.03, 0.866 (237)
	Overall[c]	$P = 0.403$	$P = 0.571$	$P = 0.550$

[a] Degrees of freedom are 1 for the chi-square distributions.
[b] *n* represents the available number of siblings.
[c] From Fisher combination of *P* values (Fisher 1954).

TABLE 7

Results from sibling-pair IIS analysis on the squared phenotype difference

Phenotype	Study	Exon 2	Exon 4	Exon 12
		Pooling analysis: β (SE) *P*, *n*[a]		
BMI		−12.3 (10.9) 0.260, 1084	−16.1 (9.6) 0.093, 1117	−3.4 (11.4) 0.769, 1112
WC		−33.4 (63.4) 0.598, 1080	−86.1(55.7) 0.123, 1114	−56.9 (65.6) 0.386,1109
		Analysis by study and metaanalysis: β (SE) *P*, *n*		
BMI	QFS	−14.6 (21.1), 0.490, 431	−39.1 (19.0) 0.040, 439	−12.0 (23.3) 0.607, 440
	HFS	−6.8 (11.0) 0.538, 480	0.4 (10.0) 0.965, 480	3.14 (10.8) 0.771, 480
	MIFS	13.4 (25.1) 0.594, 172	14.6 (19.9) 0.464, 198	−15.6 (24.9) 0.533, 192
	Meta est[b]	−5.6 (9.1) 0.538	−4.4 (8.1) 0.586	−1.69 (9.1) 0.852
	Q[c]	0.77, d.f. = 2, P = 0.682	4.48, d.f. = 2, P = 0.107	0.71, d.f. = 2, P = 0.702
WC	QFS	39.3 (116) 0.735, 432	−156 (106) 0.140, 439	−75.2 (128), 0.558, 440
	HFS	−59.6 (72.7) 0.413, 478	−9.2 (66.1) 0.889, 478	−36.9 (71.5) 0.606, 478
	MIFS	−6.3 (156) 0.968, 170	−19.6 (119) 0.870, 197	−131 (149) 0.379, 191
	Meta est[b]	−28.3 (57.3) P = 0.622	−44.7 (50.7) P = 0.378	−58.6 (57.6) P = 0.308
	Q[c]	0.54, d.f. = 2, P = 0.762	1.44, d.f. = 2, P = 0.488	0.35, d.f. = 2, P = 0.841

[a] *n* represents the available number of sibling pairs.
[b] Metapooled estimate (Hedges and Olkin 1985).
[c] χ^2 statistics for testing heterogeneity of effect sizes (Hedges and Olkin 1985).

enough to meet the rapidly increasing genetic statistical needs. For example, statistical methodologies and theories exist, such as GLM and generalized estimating equations, that can flexibly account for varying familial correlation matrices due to different pedigree structures. At the same time, to our knowledge, there are no statistical packages that can easily handle such variations.

As such, the procedures and analyses proposed herein should be an example for conducting a pooled analysis in the absence of "bona fide" methodologies, although some methodological issues are still in question. For example, the empirical correction factors obtained in this study were all <1 (Tables 10 and 11), implying that the estimated standard errors from OLS methods are bigger than those from methods accounting for familial correlations. This result may initially seem to be counterintuitive, because treating correlated observations as if they were uncorrelated often yields smaller standard errors due to inflated information. However, appropriate control for correlation may yield smaller standard errors. For example, suppose that random variables X and Y are perfectly correlated [*i.e.*, corr(X, Y) = 1] and $X = Y + \delta$ for some nonrandom δ. Now, we have n pairs of observed X and Y. If we apply a paired t-test to testing the null hypothesis of $\delta = 0$, the result will be significant no matter how small a nonzero δ, because

TABLE 8

Results from sibling-pair IIS analysis: the product of deviations of the phenotype from grand mean

Phenotype	Study	Exon 2	Exon 4	Exon 12
		Pooling analysis: β (SE) *P*, *n*[a]		
BMI		4.4 (4.4) 0.312, 1084	−4.4 (4.1) 0.283, 1117	−3.5 (4.7) 0.458, 1112
WC		6.9 (30.9) 0.822, 1080	−24.1(28.5) 0.399, 1114	0.3 (32.6) 0.992, 1109
		Analysis by study and metaanalysis: β (SE) *P*, *n*		
BMI	QFS	−1.1 (6.9) 0.876, 432	−8.6 (7.0) 0.216, 439	−5.4 (8.0) 0.496, 440
	HFS	11.8 (6.9) 0.090, 480	0.8 (6.3) 0.900, 480	0.8 (6.8) 0.901, 480
	MIFS	2.1 (10.0) 0.833, 172	−7.2 (8.1) 0.373, 198	−4.5 (10.1) 0.658, 192
	Meta est[b]	4.7 (4.4) 0.281	−4.4 (4.1) 0.283	−2.4 (4.6) 0.608
	Q[c]	1.83, d.f. = 2, P = 0.400	1.16, d.f. = 2, P = 0.560	−2.4, d.f. = 2, P = 0.817
WC	QFS	−35.8 (44.5) 0.422, 432	−74.3 (44.5) 0.095, 438	−5.3 (51.6) 0.918, 440
	HFS	60.0 (49.1) 0.222, 478	−28.2 (44.7) 0.527, 478	−7.8 (48.4) 0.873, 478
	MIFS	−56.1 (90.2) 0.535, 170	94.0 (67.5) 0.166, 197	35.5 (85.8) 0.679, 191
	Meta est[b]	−0.1 (31.0) P = 0.998	−25.3 (28.6) P = 0.376	−0.6 (32.6) P = 0.986
	Q[c]	2.53, d.f. = 2, P = 0.283	4.34, d.f. = 2, P = 0.11	0.21, d.f. = 2, P = 0.902

[a] *n* represents the available number of sibling pairs.
[b] Metapooled estimate (Hedges and Olkin 1985).
[c] χ^2 statistics for testing heterogeneity of effect sizes (Hedges and Olkin 1985).

TABLE 9

Backward elimination results

Phenotype	Effects	Variables[a]	Coefficients	SE	*P* value
BMI		(Constant)	20.775	0.565	<0.001
	Main genotypes	HM_EX2	−1.315	0.641	0.040
	Epistasis	HT2HM12	5.736	2.325	0.014
	Covariates	Study 1	−2.419	0.347	<0.001
		Study 2	−1.916	0.344	<0.001
		Study 3	−1.805	0.677	0.008
		Study 4	−3.804	0.499	<0.001
		Study 5	5.822	0.644	<0.001
		Study 6	8.529	0.474	<0.001
		Study 7	−3.857	0.812	<0.001
		AA	2.338	0.408	<0.001
		Age	0.347	0.029	<0.001
		Age^2	−0.003	<0.001	<0.001
	Interactions	HM2_Sex	2.219	0.960	0.027
WC		(Constant)	63.700	1.617	<0.001
	Covariates	Study 2	−4.612	0.793	<0.001
		Study 4	−6.813	1.180	<0.001
		Study 5	−14.679	1.574	<0.001
		Study 6	19.593	1.080	<0.001
		Study 7	−9.175	1.948	<0.001
		Study 8	−2.629	0.945	0.005
		Sex	8.471	0.570	<0.001
		Diabetes	−1.841	0.693	0.008
		AA	2.569	0.825	0.002
		Age	0.901	0.074	<0.001
		Age^2	−0.006	0.001	<0.001

[a] Variable labels: Genotypes: HM_EX2, 1 for subjects with *R109R* at exon 2, 0 for the other subjects. Epistasis: HT2HM12, 1 for subjects with *K109R* at exon 2 and *N656N* at exon 12, 0 for the other subjects. Main Covariates: Study no.: 1 for subjects in Study no., 0 otherwise (see Table 1). The subjects in the ninth study are the referent. AA: 1 for African-American subjects, 0 for the other subjects. Sex: 1 for male subjects, 0 for female subjects. Diabetes: 1 for diabetic subjects, 0 for the other subjects. Interactions: HM2_Sex, interaction between HM2 and sex, where HM2 is 1 for subjects with *R109R* at exon 2 and 0 for the other subjects. Therefore HM2_Sex is 1 for male subjects with *R109R* at exon 2 and 0 for the other subjects.

the estimated standard error of δ, the denominator of the test statistic, will be 0. But if we apply a two-sample *t*-test, which ignores the correlation, then the results will depend on the magnitude of δ and the number of subjects, because the estimated standard error will not be zero. However, the question of whether this reasoning also applies to the situation described in this article is still unanswered. If this were the case in general, *P* values based on OLS methods would provide only an upper limit of "true" *P* values. Thus, the OLS-based *P* values may not provide a reasonable conclusion about hypotheses tests unless the OLS *P* values are large. Therefore, when the *P* values from OLS methods are borderline (between 0.05 and 0.10), application of correlation-adjusting methods may be needed for more accurate *P* values.

The MI method adds (or more precisely "allows for") uncertainty around the unknown missing genotypes by generating *multiple* randomly imputed genotypes where the imputed values are the predicted values *plus some random error* for which the expected squared value is equal to the estimated variance of the prediction. This is done multiple times and the variance in the results that occur from imputation to imputation enter into the calculation of standard errors and *P* values, thereby "penalizing" one for uncertainty rather than artifactually augmenting one's certainty. In some cases, the variance around the imputation is zero because the missing genotypes are known without error by Mendel's laws. In those cases, the imputation adds the correct amount of uncertainty; it just happens that that amount is zero. More broadly, the justification of the regression imputation is that genotypes can be predicted on the basis of observed phenotypes and covariates just as phenotypes can be predicted on the basis of observed genotypes. Although imputing such genotypes does not in and of itself create new information, the regression imputation of missing genotypes in this way allows one to use the full information that is available in a data set by, for example, not requiring one to drop subjects who are missing genotypic information at one locus but who have information at other loci when conducting a multilocus analysis; *i.e.*, it avoids list-wise deletions.

Under the null hypothesis of no association, there is

TABLE 10

Joint effects of genotypes on BMI adjusted for sex, ethnicity, diabetes, and age polynomials

	Exon 2				Exon 4				Exon 12					
	K109R		*R109R*		*Q223R*		*R223R*		*K656N*		*N656N*			
Imputation	Coef	SE	Coef	SE	Coef	SE	Coef	SE	Coef	SE	Coef	SE	$\chi^2(6)$[b]	*P* value
						S.A.G.E. ASSOC analysis								
1	−0.010	0.194	−0.175	0.400	0.147	0.221	0.187	0.240	0.061	0.174	−0.099	0.219	0.906	0.989
2	0.157	0.202	0.054	0.432	0.001	0.219	0.092	0.279	0.107	0.202	−0.024	0.488	1.130	0.980
3	0.207	0.203	0.119	0.420	−0.003	0.221	0.285	0.279	0.084	0.199	0.023	0.464	1.288	0.972
4	−0.055	0.204	−0.437	0.427	0.144	0.221	0.312	0.279	0.082	0.201	−0.063	0.481	2.184	0.902
5	0.035	0.204	−0.142	0.423	0.054	0.221	0.254	0.282	0.143	0.201	−0.037	0.475	1.591	0.953
Overall[a]	0.067	0.201	−0.116	0.421	0.068	0.221	0.226	0.273	0.095	0.196	−0.040	0.447		
P value[a]	0.740		0.782		0.756		0.407		0.626		0.929			0.988[d]
Imputation	Coef	SE	Coef	SE	Coef	SE	Coef	SE	Coef	SE	Coef	SE	*F* change	*P* value
						Ordinary regression analysis								
1	−0.064	0.249	−0.455	0.529	0.245	0.278	0.159	0.336	0.063	0.245	0.183	0.606	0.293	0.941
2	0.078	0.248	−0.263	0.533	0.102	0.277	0.107	0.335	0.137	0.245	0.266	0.607	0.184	0.981
3	0.194	0.249	−0.187	0.526	0.110	0.278	0.076	0.336	0.102	0.245	0.313	0.600	0.263	0.954
4	−0.106	0.251	−0.752	0.529	0.244	0.279	0.337	0.337	0.094	0.246	0.234	0.607	0.527	0.788
5	0.005	0.250	−0.323	0.518	0.144	0.278	0.263	0.336	0.196	0.245	0.272	0.605	0.281	0.946
Overall[c]	0.021	0.249	−0.396	0.527	0.169	0.278	0.188	0.336	0.118	0.245	0.254	0.605		
P value[c]	0.932		0.452		0.543		0.575		0.629		0.675			0.985[d]
Overall[c]	0.022	0.249	−0.395	0.527	0.169	0.278	0.189	0.336	0.119	0.245	0.254	0.605	0.246	0.961[c]
P value[c]	0.930		0.454		0.544		0.573		0.627		0.675			
						Correction factor: within Var(S.A.G.E.)/within Var(OLS)								
		0.652		0.636		0.630		0.657		0.636		0.524		Average: 0.623

[a] See APPENDIX B for $\dim(Q_i) = 1$.
[b] χ^2 with 6 d.f. for testing joint effects of the genotypes calculated as twice minus the log-likelihood difference between with and without the genotypes.
[c] Obtained from the weighted OLS regression.
[d] See APPENDIX B for $\dim(Q_i) > 1$.

TABLE 11

Joint effects of genotypes on WC adjusted for sex, ethnicity, diabetes, and age polynomials

	Exon 2				Exon 4				Exon 12					
	K109R		*R109R*		*Q223R*		*R223R*		*K656N*		*N656N*			
Imputation	Coef	SE	Coef	SE	Coef	SE	Coef	SE	Coef	SE	Coef	SE	$\chi^2(6)$[b]	*P* value
							S.A.G.E. ASSOC analysis							
1	0.492	0.581	0.158	1.271	−0.297	0.636	−0.657	0.809	−0.136	0.573	−0.799	1.392	1.838	0.934
2	1.072	0.597	0.727	1.290	−0.807	0.648	−1.242	0.823	−0.138	0.586	−1.318	1.463	5.660	0.462
3	0.907	0.589	0.986	1.253	−0.547	0.640	−1.360	0.812	−0.050	0.578	−0.588	1.420	5.082	0.533
4	0.674	0.600	0.114	1.275	−0.265	0.654	−0.677	0.828	−0.070	0.590	−0.241	1.475	1.913	0.928
5	0.770	0.589	0.367	1.241	−0.336	0.642	−0.412	0.817	0.038	0.579	−0.560	1.425	1.919	0.927
Overall[a]	0.783	0.591	0.470	1.267	−0.450	0.644	−0.870	0.818	−0.071	0.581	−0.701	1.439		
P value[a]	0.186		0.711		0.484		0.288		0.902		0.626			0.999[d]
Imputation	Coef	SE	Coef	SE	Coef	SE	Coef	SE	Coef	SE	Coef	SE	*F* change	*P* value
							Ordinary regression analysis							
1	0.358	0.609	−0.062	1.295	−0.187	0.682	−0.645	0.822	−0.091	0.599	−0.101	1.483	0.168	0.985
2	0.839	0.608	0.416	1.309	−0.589	0.682	−1.117	0.822	0.040	0.601	−0.801	1.492	0.571	0.753
3	0.633	0.609	0.579	1.287	−0.244	0.680	−1.159	0.821	0.012	0.599	−0.145	1.469	0.494	0.813
4	0.452	0.617	−0.164	1.302	−0.059	0.687	−0.615	0.829	−0.012	0.605	0.402	1.495	0.233	0.966
5	0.601	0.612	0.111	1.266	−0.073	0.681	−0.246	0.822	0.136	0.600	0.194	1.478	0.180	0.982
Overall[a]	0.578	0.611	0.176	1.293	−0.230	0.682	−0.756	0.823	0.017	0.601	−0.090	1.484		
P value[a]	0.345		0.892		0.736		0.358		0.977		0.952			0.999[d]
Overall[c]	0.576	0.611	0.173	1.292	−0.232	0.682	−0.753	0.823	0.019	0.601	−0.088	1.484	0.265	0.953[c]
P value[c]	0.346		0.893		0.734		0.360		0.975		0.953			
							Correction factor: within Var(S.A.G.E.)/within Var(OLS)							
		0.937		0.961		0.891		0.987		0.935		0.936		Average: 0.941

[a] See APPENDIX B for $\dim(Q_j) = 1$.
[b] χ^2 with 6 d.f. for testing joint effects of the genotypes calculated as twice minus the log-likelihood difference between with and without the genotypes.
[c] Obtained from the weighted OLS regression.
[d] See APPENDIX B for $\dim(Q_j) > 1$.

TABLE 12

Epistasis effects adjusted for sex, ethnicity, diabetes, and age polynomials

	P values of joint effect of the two- and three-way interactions among genotypes			
	In the presence of main genotype effects		In the absence of main genotype effects	
Imputation	BMI	WC	BMI	WC
1	0.122	0.659	0.330	0.899
2	0.051	0.389	0.204	0.564
3	0.079	0.209	0.241	0.383
4	0.046	0.312	0.127	0.616
5	0.013	0.257	0.067	0.572
	Overall *P* values			
MI[a]	0.195	0.687	0.470	0.942
Weighted OLS[b]	0.212	0.706	0.471	0.893

[a] *P* values based on multiple imputation; see APPENDIX B for $\dim(Q_i) > 1$.

[b] *P* values obtained from the weighted OLS.

no reason to suspect that such imputation, if properly accounted for, would bias the expected results. However, under an alternative hypothesis, MI methods can give different results. This results from (a) reducing possible biases if the data are missing at random (MAR) but not missing completely at random (MCAR) in the terminology of LITTLE and RUBIN (1987); (b) increasing the precision of estimates by using all of the information available in the data set in contrast to analysis of only complete cases; and (c) thereby increasing power. In our example, the nonsignificant results were not changed after missing values were imputed. That is, analysis of complete cases and MI yielded equivalent results (data not shown). This further consolidates the null association and serves as a form of sensitivity analysis. MI is now generally accepted and widely advocated among statisticians. However, it has not made significant headway into many applied areas, including genetic research. We hope that this illustration of its use may encourage further adoption of it or something analogous by genetic researchers.

In terms of combining *P* values, when the dimension of the parameter vector is one, the procedure described in APPENDIX B [when $\dim(Q_i) = 1$] is well justified theoretically and by simulation studies (*e.g.*, RUBIN 1996). Therefore, it is surprising that the overall combined *P* values based on weighted OLS are all very close to the *P* values based on that procedure (Tables 10 and 11), despite the fact that the weighting does not account for incompleteness of random missing imputation. This could support the use of weighting because of its simplicity, effectiveness, and reliability, which in turn may allow its use to be extended to the case of multidimensional parameters. In the cases of multidimensions, on the other hand, the particular procedure adopted herein [APPENDIX B; when $\dim(Q_i) > 1$] can be used as only a rough guide—providing an estimate of the range of *P* values between one-half and twice the calculated value (LI *et al.* 1991), although the use of this procedure has been advocated because of its computational simplicity in comparison with other procedures. However, the *P* values obtained from this procedure and from the weighted OLS are again very similar (Table 12). Calculation of the overall *P* values for pleiotropic effects (Table 13) based on weighting was not possible because of software limitation.

With respect to allele-by-allele interaction analysis, it should be pointed out that when several loci, or markers, are in close physical proximity to each other, as in the current case, and interactions among loci are tested for, such interactions, if observed, may be due to linkage disequilibrium and not true epistasis. To understand why, consider a hypothetical situation of two diallelic loci, *A* and *B*, with alleles A_1 and A_2 and B_1 and B_2, respectively, in close proximity to each other. Assume that they are in equilibrium and that neither has any

TABLE 13

Pleiotropic effects of the alleles on BMI and WC adjusted for sex, ethnicity, diabetes, and age polynomials

	P values from Wilk's λ test statistic						
	Main			Two-way interaction			Three-way interaction
Imputation	Exon 2	Exon 4	Exon 12	Exons 2 and 4	Exons 2 and 12	Exons 4 and 12	Exons 2, 4, and 12
1	0.332	0.893	0.298	0.483	0.353	0.498	0.635
2	0.086	0.340	0.025	0.865	0.023	0.278	0.491
3	0.009	0.423	0.057	0.641	0.027	0.607	0.808
4	0.039	0.630	0.028	0.262	0.007	0.366	0.673
5	0.0005	0.137	0.001	0.110	0.0004	0.043	0.152
Overall *P* values[a]	0.586	0.921	0.433	0.984	0.786	0.874	0.958

[a] *P* values based on multiple imputation; see APPENDIX B for $\dim(Q_i) > 1$.

individual or interactive effect on the phenotype. Then at some point in history, an allelic variation might have occurred at locus *C*, which is also very close to *A* and *B*, and that *C* now has alleles C_1 and C_2, with the C_2 allele conferring a predisposition to increased phenotypic values. Suppose further that C_2 arose on a chromosome with alleles A_2 and B_2 and, due to tight linkage, to this day C_2 occurs primarily (though not necessarily exclusively) on chromosomes with the A_2, B_2 haplotype. Finally, assume that loci *A* and *B* are genotyped in a study but *C* is not (*i.e.*, *C* is unobserved). Then, given a sufficiently large sample, one will detect an interaction "effect" of the *A* and *B* loci. However, this is solely due to the fact that when A_2 and B_2 occur together, they represent a haplotype with a higher likelihood of having a C_2 allele at the *C* locus. For the present study, such "phase" information is not available. However, if a polymorphism under study does not cause variation but is both linked to and in disequilibrium with a polymorphism that causes variation in the phenotype, power to detect epistasis *may* be enhanced through the use of estimated haplotypes rather than single nucleotide polymorphisms (Fallin and Schork 2000; Fallin *et al.* 2001). The power will depend critically on the degree of disequilibrium between the polymorphism and the causative allele. Nevertheless, the haplotype analysis should be very close (though not identical) to our analyses testing for epistasis.

With respect to comparison between metaanalysis and pooling analysis, both methods should yield similar results in terms of estimated effects and their significance, as was the case in this study. Metaanalysis can be as powerful as pooling analysis in certain particular situations (*e.g.*, Olkin and Sampson 1998). However, pooling analysis has advantages over metaanalysis, including the ability to run analyses in a consistent fashion across studies, to test certain assumptions (*e.g.*, normality), and to run analyses beyond those the original investigators ran. In addition, pooling analysis enables us to consistently provide the same analysis to each data set and impose a degree of quality control on the data analyses that is uniform across all data sets. Finally, raw data pooling allows examination for outliers, use of transformations, and full use of the statistician's usual armamentarium. In this sense, raw data pooling may be preferable to metaanalysis of published summary statistics. Nevertheless, if all analyses are conducted identically and correctly both within each data set separately and in the pooled data set, then there may be no theoretical difference in power.

Interaction of study effects with genotypic effects was assessed by means of testing heterogeneity of effects among studies. The heterogeneity of the effects was not significant, as presented earlier. Because of this nonsignificance and concerns about possible overfitting, we did not include study-by-allele interaction effects in the main pooling analysis; *i.e.*, we estimated equal genetic effects across the studies. Although the study effects are large (Table 9), this does not mean that the genotypic effects are different over the studies, but rather that average levels of the phenotypes are different in the different studies. However, we acknowledge that modeling study effects alone may not capture all possible heterogeneity across the study populations due to differences in sampling schemes, degrees of demographic homogeneity within samples, and so on. For example, the Danish samples (Echwald *et al.* 1997) were collected under a unique sampling scheme. Moreover, the Finnish samples (Oksanen *et al.* 1998) may be relatively homogenous compared to the other study samples, although the significance of this remains open to questions (Abbott 2000).

An alternative approach would be a mixed-effects model with cohort as a random effect. Such a model could include empirical Bayes estimation and testing (Carlin and Louis 2000), which we are currently exploring. Such mixed-effects (random effect for studies) model approaches to testing heterogeneity are certainly a reasonable alternative that can be pursued in future work.

On the basis of the results from the backward elimination procedures (Table 9), subjects who are heterozygous at exon 2 and homozygous at exon 4 are significantly greater in imputed missing BMI compared to the other subjects (Table 9). Although these subjects (there were five such) could deserve more investigation, only one subject with this combination is extremely obese, with BMI 51. It is therefore unlikely that this particular allele interaction can cause obesity and there is a limitation in epistasis analysis because we used only three exons; *i.e.*, we do not know the pathways of how the three exons interact with unknown alleles in other exons and introns. Furthermore, if we had adjusted for the number of tests, each adjusted *P* value would have been much higher than that reported in this article. For example, if all tests were independent, a multiple-test adjusted *P* value p_b may be written as $p_b = 1 - (1 - p)^t$, where t is the number of tests and p is an unadjusted *P* value. It follows that p_b is 0.06 even with $t = 6$ and $P = 0.01$. This further confirms that it is unlikely that all the nonsignificant results from this study are due to type 2 error, although, when the tests are dependent, the adjusted *P* value will be <0.06 but >0.01. Moreover, even 1% of variation of phenotype due to their allelic variants would have been detected with >99% power with 3000 subjects. As far as confidence intervals of point estimates are concerned, they can be immediately computed from the standard errors provided in the tables.

The lack of association between the amino acid substitutions and obesity indices, despite a large sample, suggests that the substitutions do not affect the phenotypes. While amino acid substitutions may result in either nonfunctioning or poorly functioning proteins, or even

functional proteins (if the effect is silent), it is important to note that among complex traits with multiple pathways, such as obesity, the absence of association might not necessarily indicate a lack of effect. It may simply be that persons with the amino acid substitution compensated by other means or that additional genotypic factors may be involved and need to be taken into account before the phenotype becomes manifest. However, the lack of association does not rule out the possibility that the three alleles may influence intermediate traits, or phenotypes, not examined as part of the analyses conducted in this article.

In conclusion, conducting appropriate statistical procedure and analysis of pooled genetic data requires careful data management and flexible adaptation of methods and software to effectively model biological effects of the genes under study. In the absence of well-developed guidelines, we hope that the procedures and methods illustrated herein can be useful as an example for future pooling of genetic studies of quantitative trait loci.

We thank Drs. Jose Fernandez and Gary Gadbury for their valuable input. This study was supported in part by the National Institutes of Health grants R01DK51716, R01DK52431, R01ES09912, F33DK09919, P30DK26687, R01HD29569, R01GM28356, and P41RR03655.

LITERATURE CITED

Abbott, A., 2000 Manhattan versus Reykjavik. Nature **406:** 340–342.

Allison, D. B., B. Thiel, P. St. Jean, R. C. Elston, M. C. Infante *et al.*, 1998 Multiple phenotype modeling in gene-mapping studies of quantitative traits: power advantages. Am. J. Hum. Genet. **63:** 1190–1201.

Allison, D. B., M. Heo, N. Kaplan and E. R. Martin, 1999 Sibling-based tests of linkage and association for quantitative traits. Am. J. Hum. Genet. **64:** 1754–1764.

Amos, C. I., R. C. Elston, G. E. Bonney, B. J. B. Keats and G. S. Berenson, 1990 A multivariate approach for detecting linkage, with application to a pedigree with adverse lipoprotein phenotype. Am. J. Hum. Genet. **47:** 247–254.

Carlin, B. P., and T. A. Louis, 2000 Empirical Bayes: past, present and future. J. Am. Stat. Assoc. **95** (452): 1286–1289.

Chagnon, Y. C., W. K. Chung, L. Pérusse, M. Chagnon, R. L. Leibel *et al.*, 1999 Linkage and associations between the leptin receptor (*LEPR*) gene and human body composition in the Quebec Family Study. Int. J. Obes. **23:** 278–286.

Chagnon, Y. C., J. H. Wilmore, I. B. Borecki, J. Gagnon, L. Pérusse *et al.*, 2000 Association between the leptin receptor gene and adiposity in middle-aged Caucasian males from the HERITAGE family study. J. Clin. Endocrinol. Metab. **85:** 29–34.

Chua, Jr., S. C., W. K. Chung, X. S. Wu-Peng, Y. Zhang, S. M. Liu *et al.*, 1996 Phenotypes of mouse diabetes and rat fatty due to mutations in the *OB* (leptin) receptor. Science **271:** 994–996.

Chung, W. K., L. Power-Kehoe, M. Chua, F. Chu, M. Devoto *et al.*, 1997 Exonic and intronic variation in the leptin receptor (*OBR*) of obese humans. Diabetes **46:** 1509–1511.

Clément, K., C. Vaisse, N. Lahlou, S. Cabrol, V. Pelloux *et al.*, 1998 A mutation in the human leptin receptor gene causes obesity and pituitary dysfunction. Nature **392:** 398–401.

Comuzzie, A. G., and D. B. Allison, 1998 The search for human obesity genes. Science **280:** 1374–1377.

Cooper, H., and L. V. Hedges (Editors), 1994 *The Handbook of Research Synthesis.* Russell SAGE Foundation, New York.

del Giudice, E. M., L. Perrone, P. Forabosco, M. Devoto, M. T. Carbone *et al.*, 2000 Linkage study of early-onset obesity to leptin receptor gene in Italian children. Nutr. Res. **20:** 1059–1063.

Donohoue, P. A., T. L. Burns, M. C. B. Mendoza, W. K. Chung and R. L. Leibel, 2000 Lys656Asn variant of the leptin receptor gene (LEPR) and the β-3 adrenergic receptor (β3AR) gene linked to body mass index in humans: the Muscatine study. Pediatr. Res. **47:** 127A.

Echwald, S. M., T. D. Sørensen, T. I. A. Sørensen, A. Tybjærg-Hansen, T. Andersen *et al.*, 1997 Amino acid variants in the human leptin receptor: lack of association to juvenile onset obesity. Biochem. Biophys. Res. Commun. **233:** 248–252.

Elston, R. C., 2000 Introduction and overview. Stat. Methods Med. Res. **9:** 527–541.

Elston, R. C., S. Buxbaum, K. B. Jacobs and J. M. Olson, 2000 Haseman and Elston revisited. Genet. Epidemiol. **19:** 1–17.

Fallin, D., and N. J. Schork, 2000 Accuracy of haplotype frequency estimation for biallelic loci, via the expectation-maximization algorithm for unphased diploid genotype data. Am. J. Hum. Genet. **67:** 947–959.

Fallin, D., A. Cohen, L. Essioux, I. Chumakov, M. Blumenfeld *et al.*, 2001 Genetic analysis of case/control data using estimated haplotype frequencies: application to APOE louc variation and Alzheimer's disease. Genome Res. **11:** 143–151.

Fisher, R. A., 1954 *Statistical Methods for Research Workers*, Ed. 12. Hafner Publishing, New York.

George, V. T., and R. C. Elston, 1987 Testing the association between polymorphic markers and quantitative traits in pedigrees. Genet. Epidemiol. **4:** 193–201.

Gotoda, T., B. S. Manning, A. P. Goldstone, H. Imrie, A. L. Evans *et al.*, 1997 Leptin receptor gene variation and obesity: lack of association in a white British male population. Hum. Mol. Genet. **6:** 869–876.

Haseman, J. K., and R. C. Elston, 1972 The investigation of linkage between a quantitative trait and a marker locus. Behav. Genet. **2:** 3–19.

Hedges, L. V., and I. Olkin, 1985 *Statistical Methods for Meta-Analysis.* Academic Press, New York.

Li, K. H., T. E. Raghunathan and D. B. Rubin, 1991 Large-sample significance levels from multiply-imputed data using moment-based statistics and F reference distribution. J. Am. Stat. Assoc. **86:** 1065–1073.

Little, R. J. A., and D. B. Rubin, 1987 *Statistical Analysis With Missing Data.* John Wiley & Sons, New York.

Lynch, M., and B. Walsh, 1998 *Genetics and Analysis of Quantitative Traits.* Sinauer Associates, Sunderland, MA.

Mangin, B., P. Thoquet and N. Grimsley, 1998 Pleiotropic QTL analysis. Biometrics **54:** 88–99.

Matsuoka, N., Y. Ogawa, K. Hosoda, J. Matsuda, H. Masuzaki *et al.*, 1997 Human leptin receptor gene in obese Japanese subjects: evidence against either obesity-causing mutations or association of sequence variants with obesity. Diabetologia **40:** 1204–1210.

Oksanen, L., J. Kapiro, P. Mustajoki and K. Kontula, 1998 A common pentanucleotide polymorphism of the 3′-untranslated part of the leptin receptor gene generates a putative stem-loop motif in the mRNA and is associated with serum insulin levels in obese individuals. Int. J. Obes. **22:** 634–640.

Olkin, I., and A. Sampson, 1998 Comparison of meta-analysis versus analysis of variance of individual patient data. Biometrics **54:** 317–322.

Rolland-Cachera, M. F., M. Sempe, M. Guilloud-Bataille, E. Patois, F. Pequignot-Guggenbuhl *et al.*, 1982 Adiposity indices in children. Am. J. Clin. Nutr. **36:** 178–184.

Rubin, D. B., 1978 Multiple imputations in sample surveys—a phenomenological Bayesian approach to nonresponses, pp. 20–34 in *Proceedings of the Survey Research Methods Section.* American Statistical Association.

Rubin, D. B., 1987 *Multiple Imputations for Nonresponse in Surveys.* John Wiley & Sons, New York.

Rubin, D. B., 1996 Multiple imputation after 18 years. J. Am. Stat. Assoc. **91:** 473–489.

S.A.G.E., 1997 *Statistical Analysis for Genetic Epidemiology*, Release 3.1 (Computer program package available from the Department of Epidemiology and Biostatistics, Rammelkamp Center for Education and Research, MetroHealth Campus, Case Western University, Cleveland).

Schafer, J. L., 1997 *Analysis of Incomplete Multivariate Analysis* (Monographs on Statistics and Applied Probability Series 72). Chapman & Hall, New York.

Schafer, J. L., and M. K. Olsen, 1998 Multiple imputation for multivariate missing-data problems: a data analyst's perspective. Multivar. Behav. Res. **33:** 545–571.

Silver, K., J. Walston, W. K. Chung, F. Yao, V. V. Parikh *et al.*, 1997 The Gln223Arg and Lys656Asn polymorphisms in the human leptin receptor do not associate with traits related to obesity. Diabetes **46:** 1898–1900.

Tanizawa, Y., A. C. Riggs, S. Dagog-Jack, M. Vaxillaire, P. Froguel *et al.*, 1994 Isolation of the human LIM/homeodomain gene islet-1 and identification of a sample sequence repeat polymorphism. Diabetes **43:** 935–941 (erratum, Diabetes **43:** 1171).

Communicating editor: C. Haley

APPENDIX A: DATA PROVIDERS

See Table 1 for the study names.

Finnish Studies 1 and 2: Markku Koulu, M. Karvonen, U. Pesonen, A. Rissanen, M. Laakso, and M. Uusitupa.

QFS and HFS: Claude Bouchard and Yvonne Chagnon.

MIFS: Trudy L. Burns and Patricia A. Donohoue.

Baltimore Study: Ross E. Andersen, Alan R. Shuldiner, and Kristi Silver.

Danish Study: Soren Echwald, Olaf Pedersen, and T. I. A. Sørensen.

Nigerian Study and AfAm Study: Philip Behn and M. Alan Permutt.

APPENDIX B: INFERENCE BASED ON MULTIPLE IMPUTATION (Shafer 1997)

We conducted appropriate analysis with each imputed complete data set to obtain five estimates, Q_i, $i = 1, \ldots, m(= 5)$, and their estimated variance $U_i = \mathrm{Var}(Q_i)$, $i = 1, \ldots, m(= 5)$. In the following, $\dim(Q_i)$ denotes the dimension of the parameter vector Q_i.

When $\dim(Q_i) = 1$, the final estimate, its variance T, and sampling distribution are

$$\bar{Q} = \sum Q_i/m,\ T = \mathrm{Var}(\bar{Q}) = \bar{U} + (1 + 1/m)B,$$

the total variance, where

$$\bar{U} = \sum U_i/m,$$

the within-imputation variance, and

$$B = 1/(m - 1)\sum (Q_i - \bar{Q})^2,$$

the between-imputation variance, and

$$(\bar{Q} - Q_0)/\sqrt{T} \sim t_v,$$

where Q_0 is the true parameter value and the number of degrees of freedom is

$$v = (m - 1)\left[1 + \frac{\bar{U}}{(1 + 1/m)B}\right]^2.$$

When $\dim(Q_i) = k > 1$, calculation of an "overall" P value for testing H_0: $Q = Q_0$ can be performed on the basis of the Wald test statistic; that is,

$$d_i^W = (Q_i - Q_0)^T\, U_i^{-1}(Q_i - Q_0),$$

or equivalently,

$$d_i^W = \chi_k^2(1 - p_i),$$

where p_i is the P value obtained from the ith imputed complete data set. Then we obtained a test statistic

$$D_2 = (\bar{d}^w/k - (m + 1)(m - 1)^{-1}r_2)/(1 + r_2) \sim F_{k,v},$$

where

$$\bar{d}^w = \sum d_i^w/m,\ r_2 = (1 + 1/m)\left[1/(m - 1)\sum\left(\sqrt{d_i^w} - \sqrt{\bar{d}^w}\right)^2\right],$$

and

$$v = k^{-3/m}(m - 1)(1 + 1/r_2)^2.$$

We then obtained the overall P value,

$$p = \Pr[D_2 > F_{k,v}].$$