Am. J. Hum. Genet. 49:350–365, 1991

Estimation of Hardy-Weinberg and Pairwise Disequilibrium in the Apolipoprotein AI-CIII-AIV Gene Cluster

M. B. Haviland,* A. M. Kessling,† J. Davignon,‡ and C. F. Sing*

*Department of Human Genetics, University of Michigan, Ann Arbor; †Department of Biochemistry and Molecular Genetics, Saint Mary's Hospital Medical School, London; and ‡Institut de Recherches Cliniques de Montreal, Quebec

Summary

Departures from Hardy-Weinberg (HW) equilibria and pairwise disequilibria were estimated in a sample of unrelated healthy individuals typed for six RFLPs in the apo AI-CIII-AIV gene region. The sample was composed of males and females, selected for health, from two populations, those of exclusively French-Canadian (FC) and those of some non-French-Canadian (NFC) ancestry. An approach suggested by Weir and Cockerham, which includes estimates of nonrandom association (disequilibria) between three and four alleles at two loci as well as the traditional associations between two alleles, at two loci was used. The pattern of departures from HW equilibria suggested that the genetic structures of the FC and NFC are different. Departure from HW equilibrium at an RFLP locus could not be predicted from information about other loci in the same gene region. Nonrandom associations were also evident from the pairwise analyses. Two pairs of loci had significant diallelic disequilibria, while two other pairs had significant triallelic disequilibria. All of the RFLP pairs had at least one measure of disequilibrium at its maximum value determined by allele frequencies. Inferences about pairwise disequilibria depended on the statistical approach used. Sizes of the pairwise disequilibria were not correlated with the physical distance between loci. The impact of these disequilibria on RFLP-phenotype association studies is discussed.

Introduction

Coronary artery disease (CAD) is the leading cause of mortality in North America and most of central and western Europe (Higgins and Luepker 1989). Development of CAD depends on the interactions of an individual's genetic predisposition with exposures to environmental effects, determined by such factors as diet and smoking (Davignon et al. 1983). The etiology of CAD is complex (Sing and Moll 1990). It has been well established that CAD aggregates in families but does not segregate as a Mendelian trait (Berg 1983). Epidemiologic studies (e.g., Kannel et al. 1976) have shown that continuously distributed biological traits are associated with risk of developing CAD. It is hypothesized that mutations in genes that alter the levels of these traits may be predictors of risk of CAD. Hence, efforts have focused on the study of quantitative levels of risk-factor traits—and on the genes that affect them—to understand the genetic basis of CAD. Plasma lipid levels and levels of the lipoprotein particles that transport the lipids are among the major risk-factor traits that have been considered (e.g., see Gordon et al. 1977; Maciejko et al. 1983; Pownall and Gotto 1983; Hamsten et al. 1986; Kottke et al. 1986).

The genes coding for the apolipoproteins AI (apo AI), CIII (apo CIII), and AIV (apo AIV) are candidates for determining genetic variation in plasma lipids because their products are known to be involved in the structure, transport, and metabolism of cholesterol-rich particles (reviewed by Herbert et al. 1983; Breslow 1988; Lusis 1988). These three genes are closely linked on the long arm of chromosome 11 in humans (Karathanasis et al. 1983*a*; Law et al. 1984; Schamaun et al. 1984). Apo AI is a major component of high-density lipoprotein (HDL) (Schaefer et al. 1978). HDL is involved in removing cholesterol from cells

Received January 15, 1991; revision received March 27, 1991.

Address for correspondence and reprints: Charles F. Sing, Department of Human Genetics, University of Michigan Medical School, Ann Arbor, MI 48109.

0002-9297/91/4902-0011$02.00

and as such may be involved in protection against CAD (Oram et al. 1981; Miller 1983; Kottke 1986). Apo AI also acts as a cofactor of lecithin:cholesterol acyltransferase, an enzyme involved in reverse cholesterol transport (Fielding et al. 1972). Apo CIII is a component of very-low-density lipoprotein (VLDL) and HDL (Gotto et al. 1986). Apo AIV is found in chylomicrons, HDL, and VLDL (Karathanasis and Yunis 1986).

Many investigators have attempted to associate variation at RFLP loci in the apo AI-CIII-AIV region to variation in discrete disease classifications and in continuous lipid levels. Associations between variation in a RFLP marker and phenotypic variation are dependent on the presence of linkage disequilibrium between the marker and the mutation responsible for the observed phenotypic effects. The results of these association studies have been inconclusive. For example, Kessling et al. (1986) found no evidence for an association between the allele frequencies determined by the *Sst*I RFLP, located 3′ to the apo CIII gene, and the presence or absence of hypertriglyceridemia, while Rees et al. (1985*a*) and Shoulders et al. (1986) did find an association. The *Sst*I polymorphism has been associated with differences between CAD cases and controls by some investigators (Ferns et al. 1985; Rees et al. 1985*b*) but not by others (Aburatani et al. 1988; Paulweber et al. 1988). There have also been inconsistencies among studies comparing lipid levels among RFLP genotypes. In a Caucasian sample the *Sst*I RFLP genotypes did not have significantly different HDL levels (Paulweber et al. 1988), while in a Japanese sample they did (Aburatani et al. 1988). In Caucasian CAD patients, genotypes defined by the *Pst*I RFLP, located 3′ of the apo AI coding sequence, had significantly different HDL levels in some samples (Paulweber et al. 1988; Wile et al. 1989) but not in others (Ordovas et al. 1986; Monsalve et al. 1989). In the same studies, apo AI levels were significantly different among the *Pst*I RFLP genotypes in one sample (Paulweber et al. 1988) but not in others (Monsalve et al. 1989; Wile et al. 1989).

There are many reasons why marker-phenotype association studies may be inconclusive (Cooper and Clayton 1988). In different populations there may be either different mutations affecting the phenotype or different nonrandom associations between RFLP alleles and the mutation affecting the phenotype, i.e., different genetic structures (Bodmer and Thomson 1977; Schull and Hanis 1990). Also, the statistical approach taken may lead to incorrect inferences. Typically, differences in allele frequencies among disease classes are compared with a contingency χ^2 or log likelihood test (e.g., see Kessling et al. 1985; Rees et al. 1985*b*; Monsalve et al. 1989). The average phenotypic levels among genotypes defined by a single RFLP locus are generally compared using the one-way analysis of variance (e.g., see Aburatani et al. 1988; Kessling et al. 1988*a*; Wile et al. 1989). Then, in both the discrete and continuous cases, the results from using different RFLP loci in the same region have been compared in some studies to infer which RFLP(s) is closest to a functional mutation that determines the phenotypic effect. This approach assumes that the RFLP allele frequencies at different loci are independent. Comparisons among single RFLP analyses in a particular study are unwarranted when significant nonrandom associations between frequencies of marker alleles are present. Finally, the results from these analyses (both discrete and continuous) cannot be used to localize the mutation(s) affecting a phenotype, because the magnitude of nonrandom associations between each RFLP locus and the mutation affecting the phenotype may not be linearly related to the physical distance separating them (e.g., see Litt and Jorde 1986; Borresen et al. 1988; Hegele et al. 1990). To evaluate these issues, one needs to understand the genetic structure of the candidate gene region of interest.

The purpose of the study reported in this paper was to estimate (1) departures from Hardy-Weinberg (HW) equilibria for six RFLP loci in the apo AI-CIII-AIV region and (2) associations of allele frequencies between pairs of these loci (pairwise disequilibria). A sample of unrelated healthy individuals consisting of those of exclusively French-Canadian (FC) and those of some non-FC (NFC) ancestry was studied. The statistical approach suggested by Weir and Cockerham (1989) was used. In addition to the traditional diallelic associations, it considers nonrandom associations among three (triallelic) and four (quadriallelic) alleles at two loci in the characterization of pairwise disequilibria.

The FC and NFC samples had different patterns of departures from HW equilibria. The presence of significant departures from HW equilibria in the FC sample was probably due to small sample size—i.e., chance, in conjunction with small expected genotype frequencies, genotypic selection which occurred when this healthy sample was collected, and/or data exclusion due to methodological constraints. There were statistically significant diallelic and triallelic disequilibria, although these inferences depended on the sta-

tistical approach used. Every pair of RFLPs had at least one measure of disequilibrium at its bounds set by allele frequency. The common allele for all six RFLPs was at a frequency of about .9, and the sample size was relatively small, so that, in each pairwise analysis, six of the nine genotype cells had expected values less than five. These small numbers, coupled with chance, probably account for the large number of disequilibria found to be at their bounds. The Weir and Cockerham (1989) approach is contrasted to the "traditional" approach, which considers only the pairwise nonrandom associations between two alleles on the same chromosome. The inferences were the same for these gametic diallelic disequilibria, but the traditional X^2 value was consistently smaller than the Weir and Cockerham value. Finally, this study of the apo AI-CIII-AIV region demonstrates the relevance that genetic-structure knowledge obtained from disequilibria statistics has for designing and interpreting RFLP-phenotype association studies.

Subjects and Methods

Sample

The sample used in the present study included 160 males and 92 females, all drawn from a larger sample of healthy unrelated white-collar workers at Hydro-Quebec, a utility company in Montreal. This larger sample has been described elsewhere (Xhignesse et al., in press, Kessling et al., submitted; Lussier-Cacan et al., submitted). The a priori exclusion criteria used to define the healthy sample consisted of age below 20 or above 59 years, known hypertension, diabetes, glucose intolerance, hyperlipidemia, gross obesity, nonfasting state, or the taking of prescription or nonprescription medication (except for women taking contraceptive or replacement hormones). Blood samples were drawn from 626 subjects by peripheral venepuncture after an overnight fast, and DNA was extracted from 385 blood samples. An a posteriori exclusion procedure was also used, consisting of the following criteria: hyperuricemia, hyperglycemia, abnormal liver function tests, urea or electrolytes, non-Caucasian, missing results, known kinship with others in the sample, or lipid levels greater than 4 SD from the sample mean. Of the 347 individuals who remained, 252 were successfully typed for all six RFLPs. Data exclusion occurred because RFLP typings were rejected when three investigators independently reading the autoradiographs did not agree on the typing of an individual's genotype. Typings were repeated when DNA was present in sufficient amounts. An individual was not included in the study if the amount of DNA was insufficient for a repeat of the assay. For each RFLP, typings for approximately 30 individuals were not available.

The individuals were also asked whether they knew of any ancestor who was not FC and to list the surnames and maiden names of their grandparents. FCs are descendants of a colony of approximately 6,000–8,000 immigrants from northeastern and western France who settled in eastern Canada (Laberge 1966). Geographic and social constraints have kept the population relatively isolated. A subset of the 252 individuals used in this study, consisting of 111 males and 61 females, were identified as FC by the questionnaire; another 65 individuals were classified as NFC by these criteria; and 15 had unknown ancestry.

Laboratory Procedures

DNA extraction, digestion, blotting, hybridization, and autoradiography were carried out as described elsewhere (Kessling et al., submitted). Two probes were used. Taken 5′ to 3′, they were a 2.2-kb *Pst*I genomic fragment (Kessling et al. 1985), which includes the apo AI gene and which was used to detect the *Xmn*I, *Taq*I, *Pst*I, and *Sst*I RFLPs, and 1.05 (Kessling et al. 1988*b*), a genomic probe which is from the region between the apo CIII and apo AIV genes and which was used to detect the *Pvu*II RFLP.

The restriction map is shown in figure 1. In this study, the most common allele is labeled A. The *Xmn*I(1) RFLP (A = insert; a = no insert) marks a 300-bp insert of Alu-type sequence 4 kb upstream of the apo AI gene (Coleman et al. 1986; Kessling et al. 1988*a*). The second *Xmn*I RFLP (*Xmn*I [2]; A = uncut; a = cut fragment) is caused by a point mutation 3.7 kb 5′ of the apo AI gene (Kessling et al. 1985). The

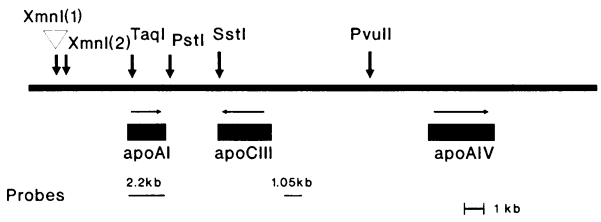

Figure 1 Diagram of apo AI-CIII-AIV region on chromosome 11, depicting locations of apolipoprotein genes, probes used, and six RFLP loci. The arrows above the genes denote the direction of transcription.

*Taq*I RFLP (A = cut; a = uncut fragment) is within the apo AI gene (Cohen et al. 1986), and the *Pst*I RFLP (A = cut; a = uncut fragment) is 3′ of the apo AI coding sequence (Kessling et al. 1985). The *Sst*I RFLP (A = uncut; a = cut fragment) is in the 3′ untranslated region of the apo CIII gene (Rees et al. 1983), and the *Pvu*II RFLP (A = uncut; a = cut fragment) is between the apo CIII and apo AIV genes (Oettgen et al. 1986).

Statistical Analysis

Allele frequencies were estimated by the gene-counting method. Disequilibria were calculated as functions of the differences between the observed and expected genotype frequencies. Estimation of departures from HW equilibria, pairwise disequilibria, and tests of statistical significance followed an approach proposed by Weir and Cockerham (1989) and will be reviewed briefly here. Traditionally, departure from HW equilibrium and gametic diallelic pairwise disequilibrium have been the only nonrandom associations considered (e.g., see Thompson et al. 1988; Hegele et al. 1990). The Weir and Cockerham approach differs because it includes nonrandom associations among three and four alleles at two loci, as well as associations between two alleles at two loci (see fig. 2). All the nonrandom associations among alleles A and a at locus 1 and B and b at locus 2 can be estimated as follows:

$$
\begin{aligned}
D_A &= P_A^A - p_A^2 \ ; \\
D_B &= P_B^B - p_B^2 \ ; \\
D_{AB} &= P_{AB} - p_A p_B \ ; \\
D_{A/B} &= P_{A/B} - p_A p_B \ ; \\
\Delta_{AB} &= P_{AB} + P_{A/B} - 2p_A p_B = D_{AB} + D_{A/B} \ ; \\
D_{AAB} &= P_A^{AB} - p_A \Delta_{AB} - p_B D_A - p_A^2 p_B \ ; \\
D_{ABB} &= P_B^{AB} - p_B \Delta_{AB} - p_A D_B - p_A p_B^2 \ ; \\
\Delta_{AABB} &= P_{AB}^{AB} - 2p_A D_{ABB} - 2p_B D_{AAB} - 2p_A p_B \Delta_{AB} \\
&\quad - p_A^2 D_B - p_B^2 D_A - \Delta_{AB}^2 - D_A D_B - p_A^2 p_B^2 \ .
\end{aligned}
$$

D_A measures departure from HW equilibrium at locus 1. D_B measures departure from HW equilibrium at locus 2. D_{AB} measures gametic diallelic disequilibrium between an allele at locus 1 and an allele at locus 2 on the same chromosome. $D_{A/B}$ measures nongametic diallelic disequilibrium between an allele at locus 1 and an allele at locus 2 on opposite chromosomes. Δ_{AB} measures composite diallelic disequilibrium between an allele at locus 1 and an allele at locus 2 (Δ_{AB} is used when coupling and repulsion heterozygotes cannot be distinguished). D_{ABB} measures triallelic disequilibrium between an allele at locus 1 and the two alleles at locus 2, adjusted for the departures from HW equilibria and diallelic disequilibria (there is also a symmetric triallelic measure D_{AAB}). And Δ_{AABB} measures quadriallelic disequilibrium between the two alleles at locus 1 and the two alleles at locus 2, adjusted for the departures from HW equilibria and diallelic and triallelic

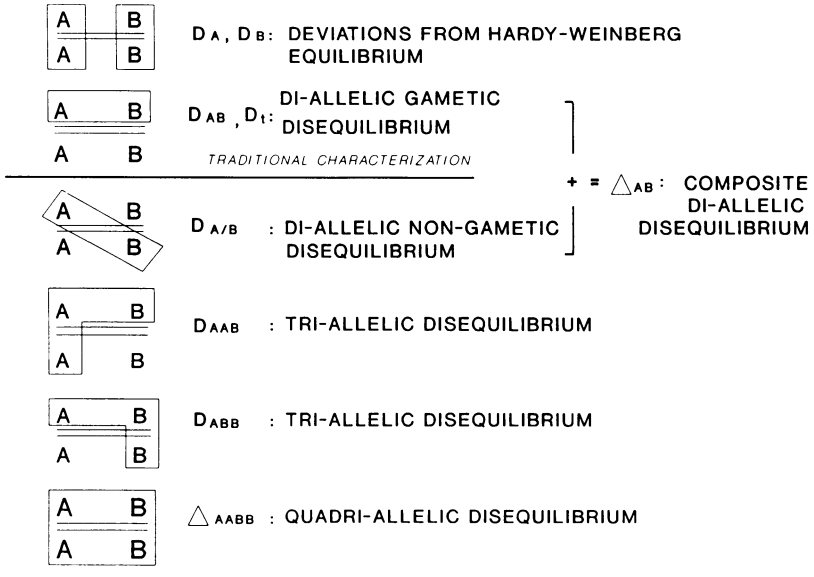

Figure 2 Weir and Cockerham (1989) characterization of two-locus disequilibria. A = alleles at locus 1; B = alleles at locus 2. Alleles on the same side of the double lines are on the same gamete. The boxes denote the alleles for which the disequilibrium is being defined.

disequilibria (the quadriallelic estimate of disequilibrium is also a composite measure). P denotes relative genotype frequencies, and p denotes relative allele frequencies.

In the sample considered here, gametic phase was not known for every individual. So, composite diallelic and quadriallelic disequilibria as well as the departures from HW equilibria and triallelic disequilibria were estimated. Then the diallelic composite disequilibrium was decomposed into the gametic and nongametic associations. To accomplish this, two-locus haplotype frequencies were estimated using an algorithm developed by Hill (1974; but see Weir and Cockerham 1979). One assumption made in the application of this algorithm is HW equilibrium, so those RFLP loci not in HW equilibrium were not included in this particular analysis. These estimated haplotype frequencies were used to estimate the gametic diallelic disequilibria. Then the nongametic diallelic disequilibrium was estimated by subtracting the gametic from the composite disequilibrium.

Weir and Cockerham also derived the estimates of the variances for the disequilibria by using multinomial theory. The expression for the departure from HW equilibrium is

$$\text{var}(D_A) = [p_A^2 p_a^2 + (1 - 2p_A)^2 D_A - D_A^2]/n \ ,$$

where n is the sample size. The variances for the other disequilibrium measures include information about the higher-order associations. For example, the estimator of the variance for the gametic diallelic disequilibrium includes measures of departures from HW equilibria and nongametic diallelic and quadriallelic disequilibria, while the estimator of the variance for the composite diallelic disequilibrium includes measures of departures from HW equilibria and triallelic and quadriallelic disequilibria, as can be seen below.

$$\begin{aligned}\text{var}(D_{AB}) = {} & [p_A p_a p_B p_b + (1 - 2p_A)(1 - 2p_B)D_{AB} \\ & + D_A D_B - D_{AB}^2 + D_{A/B}^2 \\ & + 2D_{AB}D_{A/B} + \Delta_{AB}^{AB}]/2n \ ; \\ \text{var}(\Delta_{AB}) = {} & [(p_A p_a + D_A)(p_B p_b + D_B) \\ & + (1 - 2p_A)(1 - 2p_B)\Delta_{AB}/2 \\ & + (1 - 2p_A)D_{ABB} + (1 - 2p_B)D_{AAB} \\ & + \Delta_{AABB}]/n \ .\end{aligned}$$

The estimators of the variances for the triallelic and quadriallelic disequilibria are more complex. They are given by Weir and Cockerham (1989).

To test the null hypothesis that a particular measure of disequilibrium (δ) is significantly different from zero, Weir and Cockerham suggest the statistic $X^2 = \delta^2/\text{var}(\delta)$, where the variance is calculated by assuming that the null hypothesis ($\delta = 0$) is true. This statistic is distributed approximately as a X^2 with 1 df.

When the above expressions are used for the disequilibria and variances (with the disequilibrium being tested set equal to zero), the X^2 statistics for testing the hypotheses that the departure from HW equilibrium and gametic and composite diallelic disequilibria are each zero are as follows:

$$\begin{aligned}X^2(D_A) = {} & nD_A^2/p_A^2 p_a^2 \ ; \\ X^2(D_{AB}) = {} & nD_{AB}^2/[(p_A p_a p_B p_b + D_A D_B \\ & + D_{A/B}^2 + \Delta_{AB}^{AB})/2] \ ; \\ X^2(\Delta_{AB}) = {} & n\Delta_{AB}^2/[(p_A p_a + D_A)(p_B p_b + D_B) \\ & + (1 - 2p_A)D_{ABB} + (1 - 2p_B)D_{AAB} \\ & + \Delta_{AABB}] \ .\end{aligned}$$

Note that $X^2(D_A)$ is the traditional test but that $X^2(D_{AB})$ and $X^2(\Delta_{AB})$ (and, although not shown, the X^2 for the higher-order disequilibria) include estimates of the other disequilibria in their variance.

The disequilibrium coefficients and their significance tests were calculated using programs written in the SAS/IML language (SAS Institute Inc. 1985). The sample was divided into groups based on gender and ancestry. First, the allele frequencies were compared among the groups by using the G-test (Sokal and Rohlf 1981). If the allele frequencies were not significantly different, then the departures from HW equilibria were compared to see whether the arrangement of alleles into genotypes was the same among the groups. Finally, groups with similar allele frequencies and patterns of departures from HW equilibria were combined, and diallelic, triallelic, and quadriallelic disequilibria were calculated.

There are two possible testing procedures which can be used for the pairwise disequilibria. In both cases, the disequilibrium being tested was set to zero under the null hypothesis. The first approach (reduced) started with tests of the higher-order disequilibria. If they were not significant, these terms were set to zero in the lower-order variances. If there was evidence that the higher-order disequilibria were not zero, then their estimated values were used in computing the lower-order variances (Weir and Cockerham 1989). The second approach (complete) used all the estimates of disequilibria in computing the variances (except the one being tested), even if they were not significantly different from zero. Neither approach can be considered to be more conservative than the other, because the resultant increase or decrease in the level of variance is a function of the sign of the disequilibria.

All the tests were taken as significant at a single test type I error rate of $\Pr \leqslant .05$, to minimize the

probability of a type II error. Zerba et al. (in press) demonstrate that correcting for multiple tests of significant disequilibria leads to the probability of a type II error being several times greater than the probability of a type I error. Also, Rothman (1990) suggests that, in general, correcting for multiple tests of observational data may not be warranted. Experimentwise type I errors are usually computed assuming that the null hypotheses are true for each comparison. Thus, when previous knowledge suggests that one or more of the null hypotheses are false, the experimentwise type I error rate will be conservative, leading to type II errors. For these reasons, a single test type I error rate was used.

For comparison, the traditional gametic diallelic disequilibrium was also considered. The estimator of gametic diallelic disequilibrium is the same for the traditional and Weir and Cockerham approaches (i.e., $D_{AB} = D_t$). But the variance of the traditional gametic diallelic disequilibria (e.g., see Thompson et al. 1988; Hegele et al. 1990) does not include measures of departures from HW equilibria and nongametic diallelic and quadriallelic disequilibria, as does the Weir and Cockerham variance; i.e., $\text{var}_t(D_t) = p_A p_a p_B p_b / n$. The significance of the traditional D_t is calculated using $X_t^2 = D_t^2 / \text{var}_t(D_t)$, which is also distributed approximately as a χ^2 with 1 df.

Last, the range that each of the disequilibria can take is set by the allele frequencies. The bounds of each of the disequilibria, except the quadriallelic, were calculated (Weir and Cockerham 1989). For example, for the composite diallelic disequilibria,

$$\begin{aligned}
\Delta_{max} &= 2p_A p_b && \text{if } \Delta_{AB} > 0,\ P_A^A < P_B^B,\ P_a^a > P_b^b \\
&= 2(X - Y) && \text{if } \Delta_{AB} > 0,\ P_A^A < P_B^B,\ P_a^a < P_b^b \\
&= 2p_a p_B && \text{if } \Delta_{AB} > 0,\ P_A^A > P_B^B,\ P_a^a < P_b^b \\
&= 2(X + Y) && \text{if } \Delta_{AB} > 0,\ P_A^A > P_B^B,\ P_a^a > P_b^b \\
\Delta_{min} &= -2p_A p_B && \text{if } \Delta_{AB} < 0,\ P_A^A > P_B^B,\ P_a^a < P_b^b \\
&= 2(-Z - Y) && \text{if } \Delta_{AB} < 0,\ P_A^A > P_B^B,\ P_a^a > P_b^b \\
&= -2p_a p_b && \text{if } \Delta_{AB} < 0,\ P_A^A < P_B^B,\ P_a^a > P_b^b \\
&= 2(-Z + Y) && \text{if } \Delta_{AB} < 0,\ P_A^A < P_B^B,\ P_a^a < P_b^b
\end{aligned}$$

where $X = (p_A p_b + p_a p_B)/2$; $Y = (P_a^A + P_b^B)/4$; and $Z = (p_A p_B + p_a p_b)/2$. In this way, estimated disequilibria can also be expressed as the percent of their maximum possible values.

Results

The sample used in this paper was composed of male and female FC and NFC individuals. Allele frequencies and departures from HW equilibria for the overall total, total FCs, total NFCs, FC males, and FC females are given in table 1. First, the allele frequencies were tested for homogeneity. Comparisons of departures from equilibria are affected by their range of possible values which are determined by allele frequencies. Since the allele frequencies were not significantly different either between FCs and NFCs or between FC males and FC females ($Pr > .05$; analyses not shown), a comparison of patterns of departures from HW equilibria was appropriate to determine whether the arrangement of alleles into genotypes was the same across groups. If the allele frequencies are not homogeneous, then group differences in departures from HW equilibria could simply be due to differences in allele frequencies.

The results from the HW analysis are shown in table 1. The variances given were calculated using the estimated departures from HW equilibria, although the departures were set to zero under the null hypothesis when that hypothesis was tested. The patterns of departures from HW equilibria varied among the groups. In the total sample, the genotype frequencies at the *Xmn*I(2) ($D_A = .0060$) and *Sst*I ($D_A = .0191$) RFLPs varied significantly from HW proportions. The *Pst*I ($D_A = -.0040$) and *Pvu*II ($D_A = -.0029$) RFLPs had departures from equilibria at their maximum possible values, given the allele frequencies, but they were not significantly different from zero.

The patterns of departure from HW equilibrium in the FCs and NFCs were compared because the two populations may have different genetic structures. As can be seen in table 1, half of the departures from equilibria at the RFLPs sampled (*Xmn*I[1], *Xmn*I[2], and *Taq*I) have different signs in the NFCs than in the FCs. Also, the significant departure from equilibrium at the *Sst*I RFLP observed for the total sample was present only in the NFCs, while the significant departure from equilibrium at the *Xmn*I(2) RFLP was present only in the FCs. Finally, in the FCs, even though they were not significant, the *Xmn*I(1) ($D_A = -.0081$), *Pst*I ($D_A = -.0030$), and *Pvu*II ($D_A = -.0022$) RFLPs all had negative departures from equilibria which were at their bounds, and each was missing the aa homozygous class.

A priori one would not expect genotype frequencies between males and females from the same population to differ, but, since this sample was selected for health, it is possible that different RFLP genotypes were excluded by the selection procedure in males and females. The differences in departures from HW equilibria between the FC males and FC females were smaller than the difference between the FCs and NFCs. The departure from equilibrium at the *Xmn*I(2) RFLP (D_A

Table 1

Analysis of Departure from HW Equilibrium for Six RFLPs in apo AI-CIII-AIV Region

		FC			
	TOTAL (n = 252)	Total (n = 172)	Males (n = 111)	Females (n = 61)	NFC TOTAL (n = 65)
A = *Xmn*I (1):					
Frequency (A).....	.9048	.9099	.9234	.8852	.8692
D_A	−.0011	−.0081	−.0059	−.0132	.0137
Variance	.000026	.000007	.000007	.000038	.000310
% Maximum	12.50	100.00	100.00	100.00	12.02
X^2.....	.044	1.687	.763	1.025	.940
A = *Xmn*I (2):					
Frequency (A).....	.9563	.9506	.9504	.9508	.9692
D_A	.0060	.0092	.0066	.0140	−.0009
Variance	.000027	.000056	.000068	.000219	.000001
% Maximum	14.45	19.55	13.92	29.89	100.00
X^2.....	5.260*	6.576*	2.150	5.448*	.066
A = *Taq*I:					
Frequency (A).....	.9464	.9390	.9459	.9262	.9538
D_A	.0051	.0079	.0061	.0110	−.0021
Variance	.000026	.000054	.000067	.000205	.000003
% Maximum	9.99	13.78	11.90	16.03	100.00
X^2.....	2.516	3.268	1.573	1.567	.152
A = *Pst*I:					
Frequency (A).....	.9365	.9448	.9369	.9590	.9077
D_A	−.0040	−.0030	−.0040	−.0017	−.0085
Variance	.000002	.000002	.000004	.000002	.000020
% Maximum	100.00	100.00	100.00	100.00	100.00
X^2.....	1.158	.588	.503	.111	.672
A = *Sst*I:					
Frequency (A).....	.9067	.9099	.8964	.9344	.9000
D_A	.0191	.0093	.0073	.0121	.0515
Variance	.000077	.000075	.000118	.000209	.000591
% Maximum	22.57	11.37	7.84	19.74	57.26
X^2.....	12.833*	2.223	.683	2.376	21.315*
A = *Pvu*II:					
Frequency (A).....	.9464	.9535	.9685	.9262	.9308
D_A	−.0029	−.0022	−.0010	−.0054	−.0048
Variance	.000001	.000001	.000001	.000011	.000009
% Maximum	100.00	100.00	100.00	100.00	100.00
X^2.....	.807	.409	.118	.387	.360

* Pr ≤ .05.

= .014) was significant only in the females. Since a small number of individuals were expected in the Aa and aa cells, chance may have been responsible for this difference. The other five RFLPs in the males and females were similar in the percent of the maximum value possible, given the allele frequencies. Thus, all subsequent analyses were done using the FCs with males and females pooled.

The results from the pairwise analysis are shown in table 2. Two composite diallelic disequilibria were statistically significant, that for an association between the *Xmn*I(1) and *Taq*I RFLPs (Δ_{AB} = .009) and that for an association between the *Xmn*I(1) and *Pvu*II RFLPs (Δ_{AB} = −.008). The *Xmn*I(1)-*Taq*I RFLP association was significant when the complete model was used (i.e., all estimated higher-order disequilibria were kept in the model), while the *Xmn*I(1)-*Pvu*II association was significant when the reduced model was used (i.e., when those higher-order disequilibria that were not significantly different from zero were removed from the model). The composite diallelic disequilibrium between the *Pst*I and *Pvu*II RFLPs was at its maximum value (Δ_{AB} = −.005) but was not statistically significant.

The higher-order disequilibria tended to be an order of magnitude smaller than the diallelic disequilibria. Twenty-seven of the 30 measures of triallelic disequilibria were at their maximum value, although only two of them were significant. There was a significant association between the *Xmn*I(2) and *Pvu*II RFLPs ($D_{AAB} = -.002$) when either the reduced or complete model was used. When the complete model was used, there was a significant association between the *Taq*I and *Pst*I RFLP alleles ($D_{AAB} = -.002$). None of the quadriallelic disequilibria were significant. The significant pairwise disequilibria are summarized in figure 3.

The composite diallelic disequilibria gave information about nonrandom associations between loci, but the frequency associations could be between alleles on the same or opposite chromosomes. To decompose the composite disequilibria, the haplotype frequencies were first estimated using all the RFLPs except *Xmn*I (2), which departed from HW equilibrium. Table 3 presents estimates of the gametic and nongametic diallelic disequilibria calculated using the complete model, in which all estimated higher-order disequilibria were included. For comparison, the traditional gametic disequilibria are also given. The variances for the Weir and Cockerham approach were calculated using both the complete model and the reduced model, the latter of which did not include nonsignificant higher-order disequilibria. Inferences from tests of significance were the same for both models (data not shown).

The significant composite diallelic disequilibrium was due to frequency associations between alleles on the same chromosome in one case. The gametic disequilibrium between the *Xmn*I(1)-*Taq*I RFLPs was significant ($D_{AB} = D_t = .011$), while the nongametic disequilibrium was not. The other significant composite measure was due to frequency associations both between alleles on the same chromosome and between alleles on opposite chromosomes. Both the gametic and nongametic disequilibrium between the *Xmn*I(1)-*Pvu*II RFLPs were at their bounds ($D_{AB} = D_t = D_{A/B} = -.004$), although neither was statistically significant. The gametic disequilibria for the *Xmn*I(1)-*Sst*I ($D_{AB} = D_t = -.008$) and *Taq*I-*Pvu*II ($D_{AB} = D_t = -.003$) RFLPs, as well as both the gametic and nongametic disequilibria for the *Pst*I-*Pvu*II ($D_{AB} = D_t = D_{A/B} = -.002$) RFLPs, were at the bounds set by their allele frequencies. However, their estimates were not statistically significant. The χ^2 values for tests of the traditional D_t were consistently smaller than those for the corresponding test suggested by Weir and Cockerham, but the inferences about the gametic diallelic disequilibria were the same.

Discussion

Mutations in the apo AI-CIII-AIV gene region are hypothesized to be involved in the development of CAD. Supporting evidence comes from the study of patients with premature atherosclerosis. An inversion involving the apo AI and apo CIII genes (Karathanasis et al. 1987) and an insertion in the apo AI gene (Karathanasis et al. 1983*b*) have been reported in patients with premature atherosclerosis. Recently, many population-based studies of the association between single RFLPs and either continuous variation in lipid levels or presence or absence of some disease trait have been undertaken to identify mutations acting in the population at large in disequilibrium with RFLP markers (for reviews, see Cooper and Clayton 1988; Lusis 1988; Fisher et al. 1989). The results of these studies have been inconclusive because the significant results have not been consistent among populations sampled. The reasons for these inconsistencies may be grouped into two categories. If all the studies are correct in their inferences and if RFLPs are associated with functionally important mutations in some populations and not in others, then the populations are genetically different. The functionally important mutation(s) may be present in some and not in other populations, and/or the nonrandom frequency associations between the marker alleles and the functionally important mutation alleles may differ among the populations. The second category of reasons for the inconsistencies applies if the results of some studies are spurious. The inferences made may be misleading because of the same selection procedure or small sample size. Also, RFLP-phenotype association studies that compare results of multiple single-RFLP analyses may be inappropriate for addressing the question of whether there are phenotypically important mutations in linkage disequilibrium with a specific RFLP. This is because these studies have tended to ignore allele frequency associations between RFLP markers. The analysis of the disequilibria present in the apo AI-CIII-AIV region reported in this paper was undertaken to thoroughly investigate that region's genetic structure and suggest how these results may impact RFLP-phenotype association studies.

In this study, the analysis of the departures from HW equilibria suggests two conclusions. First, the Caucasian population sampled is composed of two

Table 2

Estimation of Composite Diallelic and Triallelic and Quadriallelic Disequilibrium (FCs only)

				X^2 [a]	
	D	Variance	% Maximum	Complete Variance	Reduced Variance
A = *Xmn*I (1); B = *Xmn*I (2):					
Δ_{AB}	−.006000	.00001107	67.36	1.503	1.492
D_{AAB}	−.000541	.00000011	100.00	. . .	. . .
D_{ABB}	.000751	.00000042	100.00	.298	.368
Δ_{AABB}	.000099	.00000001		. . .	. . .
A = *Xmn*I (1); B = *Taq*I:					
Δ_{AB}	.009346	.00003674	9.40	5.082*	3.118
D_{AAB}	.000842	.00000035	100.00	.405	.358
D_{ABB}	.001618	.00000073	100.00	.750	.793
Δ_{AABB}	.000204	.00000003		. . .	. . .
A = *Xmn*I (1); B = *Pst*I:					
Δ_{AB}	−.001234	.00001931	12.39	.069	.072
D_{AAB}	−.000111	.00000016	100.00	. . .	. . .
D_{ABB}	−.000068	.00000006	100.00	. . .	. . .
Δ_{AABB}	−.000014	.00000000		.004	.004
A = *Xmn*I (1); B = *Sst*I:					
Δ_{AB}	−.007521	.00002411	46.31	1.458	1.442
D_{AAB}	−.000678	.00000023	100.00	. . .	. . .
D_{ABB}	.000894	.00000067	100.00	.376	.480
Δ_{AABB}	.000104	.00000003		. . .	. . .
A = *Xmn*I (1); B = *Pvu*II:					
Δ_{AB}	−.008383	.00000502	100.00	3.036	3.878*
D_{AAB}	−.000755	.00000008	100.00	. . .	. . .
D_{ABB}	−.000390	.00000003	100.00	. . .	. . .
Δ_{AABB}	−.000140	.00000001		.039	.039
A = *Xmn*I (2); B = *Taq*I:					
Δ_{AB}	.005594	.00003795	6.03	1.248	1.470
D_{AAB}	.000986	.00000040	100.00	.416	.390
D_{ABB}	−.001991	.00000523	40.65	2.821	1.706
Δ_{AABB}	−.000277	.00000012		.075	.075
A = *Xmn*I (2); B = *Pst*I:					
Δ_{AB}	.003262	.00001900	1.70	.934	.663
D_{AAB}	.000803	.00000026	100.00	.343	.369
D_{ABB}	.000180	.00000006	100.00	.075	.069
Δ_{AABB}	.000078	.00000000		. . .	. . .
A = *Xmn*I (2); B = *Sst*I:					
Δ_{AB}	−.006000	.00001156	67.36	1.472	1.207
D_{AAB}	.000751	.00000043	100.00	.298	.280
D_{ABB}	.000321	.00000022	100.00	.132	.093
Δ_{AABB}	−.000018	.00000001		.005	.005
A = *Xmn*I (2); B = *Pvu*II:					
Δ_{AB}	.001217	.00002727	.69	.061	.057
D_{AAB}	−.002306	.00000557	100.00	6.897*	4.046*
D_{ABB}	.000057	.00000006	100.00	.018	.004
Δ_{AABB}	−.000216	.00000006		.054	.054
A = *Taq*I; B = *Pst*I:					
Δ_{AB}	.001977	.00003225	1.91	.141	.210
D_{AAB}	−.002144	.00000515	59.60	4.558*	2.881
D_{ABB}	.000109	.00000010	100.00	.037	.040
Δ_{AABB}	−.000241	.00000008		.065	.065

(*continued*)

Table 2 (continued)

	D	Variance	% Maximum	X^2 [a] Complete Variance	X^2 [a] Reduced Variance
A = *Taq*I; B = *Sst*I:					
Δ_{AB}	.000625	.00003852	.56	.010	.011
D_{AAB}	.001086	.00000061	100.00	.462	.463
D_{ABB}	−.001786	.00000464	38.06	1.910	1.249
Δ_{AABB}	−.000232	.00000010		.071	.071
A = *Taq*I; B = *Pvu*II					
Δ_{AB}	−.002772	.00000844	48.81	.517	.480
D_{AAB}	.000372	.00000014	100.00	.156	.175
D_{ABB}	−.000129	.00000002	100.00	. . .	. . .
Δ_{AABB}	.000027	.00000000		. . .	. . .
A = *Pst*I; B = *Sst*I:					
Δ_{AB}	.001673	.00002413	1.66	.136	.107
D_{AAB}	.000092	.00000007	100.00	.034	.032
D_{ABB}	.001114	.00000046	100.00	.534	.580
Δ_{AABB}	.000120	.00000001		. . .	. . .
A = *Pst*I; B = *Pvu*II:					
Δ_{AB}	−.005138	.00000243	100.00	1.824	2.191
D_{AAB}	−.000284	.00000002	100.00	. . .	. . .
D_{ABB}	−.000239	.00000001	100.00	. . .	. . .
Δ_{AABB}	−.000053	.00000000		.013	.013
A = *Sst*I; B = *Pvu*II:					
Δ_{AB}	−.002569	.00001463	30.65	.327	.295
D_{AAB}	.000580	.00000026	100.00	.264	.295
D_{ABB}	−.000119	.00000003	100.00	. . .	. . .
Δ_{AABB}	.000047	.00000000		. . .	. . .

[a] Ellipses indicate that significance of disequilibrium could not be estimated because variance calculated under null hypothesis was approximately zero.

* Pr ≤ .05.

groups differing in genetic structure: those of FC ancestry and those of NFC ancestry. Also, the magnitude, sign, and significance of a departure from HW equilibrium at an RFLP could not be predicted from information about departures from equilibria at other RFLPs in the same region. In the FCs, one of the departures from HW equilibria was significant for an excess of heterozygotes, while three others were at their bounds for an excess of homozygotes, although the latter three were not significant.

Is the apo AI-CIII-AIV region in HW equilibrium? The answer to this question is complex because the inferences change depending on the RFLP and sample considered. The differences may be due to the sample selection procedure, i.e., sampling for health, data exclusion due to methodological constraints, small sample sizes, and/or evolutionary forces acting on the populations differently. The allele frequencies are not significantly different between the FCs and the NFCs, suggesting that those long-term evolutionary forces that change allele frequencies—i.e., forces such as mutation, migration, selection, and small population size—have not acted differently in the two groups. Rather, differences in sample selection procedure are probably the reasons for the differences in genotype frequencies between the FC population and the NFC population.

The presence of significant departures from HW equilibria may be due to the sample selection procedure. The observed and expected numbers of individu-

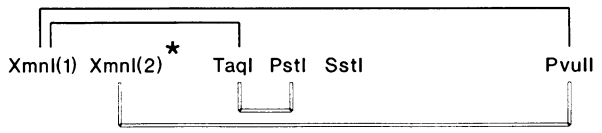

Figure 3 Genetic structure of apo AI-CIII-AIV region. The asterisk (*) signifies departure from HW equilibrium; the single line (——) signifies composite diallelic disequilibrium; and the double line (═══) signifies triallelic disequilibrium.

Table 3

Estimation of Gametic and Nongametic Diallelic Disequilibrium (FC only)

	D	Variance	% Maximum	X^2 (complete variance)
A = *Xmn*I (1); B = *Taq*I:[a]				
D_t	.010981	.00002732	22.08	4.413*
D_{AB}	.010981	.00003670	22.08	8.566*
$D_{A/B}$	−.001635	.00001099	29.72	.185
A = *Xmn*I (1); B = *Pst*I:				
D_t	−.001489	.00002488	29.92	.089
D_{AB}	−.001489	.00000931	29.92	.177
$D_{A/B}$	.000255	.00001302	.51	.005
A = *Xmn*I (1); B = *Sst*I:				
D_t	−.008121	.00003909	100.00	1.687
D_{AB}	−.008121	.00000357	100.00	3.360
$D_{A/B}$	.000600	.00002099	.82	.018
A = *Xmn*I (1); B = *Pvu*II:[a]				
D_t	−.004192	.00002114	100.00	0.831
D_{AB}	−.004192	.00000115	100.00	1.712
$D_{A/B}$	−.004192	.00000115	100.00	1.712
A = *Taq*I; B = *Pst*I:				
D_t	.001250	.00001739	2.41	.090
D_{AB}	.001250	.00001076	2.41	.197
$D_{A/B}$	.000727	.00000958	1.40	.067
A = *Taq*I; B = *Sst*I:				
D_t	.000429	.00002732	.77	.007
D_{AB}	.000429	.00001410	.77	.014
$D_{A/B}$	.000196	.00001361	.35	.003
A = *Taq*I; B = *Pvu*II:				
D_t	−.002839	.00001478	100.00	.545
D_{AB}	−.002839	.00000082	100.00	1.087
$D_{A/B}$	.000068	.00000760	.15	.001
A = *Pst*I; B = *Sst*I:				
D_t	.001999	.00002488	3.98	.161
D_{AB}	.001999	.00001693	3.98	.315
$D_{A/B}$	−.000326	.00001202	6.55	.008
A = *Pst*I; B = *Pvu*II:				
D_t	−.002569	.00001345	100.00	.490
D_{AB}	−.002569	.00000057	100.00	.998
$D_{A/B}$	−.002569	00000057	100.00	.998
A = *Sst*I; B = *Pvu*II:				
D_t	−.003029	.00002114	72.26	.434
D_{AB}	−.003029	.00000408	72.26	.861
$D_{A/B}$	.000460	00001167	1.09	.020

* $Pr \leq .05$.

[a] Significant composite diallelic disequilibrium.

als in any genotype class are not very different, because the more common allele frequencies are all about .9 and because the sample size is relatively small. Thus, a selection procedure—i.e., selection for health, data exclusion due to methodological constraints, or small sample size—that is slightly biased toward excluding one genotype over another may have created the disequilibria seen in this study. Selection for health may be genotype specific if there is in the apo AI-CIII-AIV region a functional mutation that has a large effect on the health of the individuals excluded from this study and that is associated with one of the RFLPs studied. It is unclear how selection at one locus would have affected the HW equilibrium at the other loci. Since health, as it is defined in this study, is influenced by many genes most of which have small effects, it seems unlikely that selection for health would have created the departures from HW equilibria seen. It is impossi-

ble to be certain that one of the RFLP genotypes was not preferentially present in individuals for whom typings were unobtainable. A study of the effects of data exclusion due to methodological constraints on genotype frequencies is needed to determine whether it is the reason for the departures from HW equilibria. Small sample size, i.e., chance, probably has the largest impact on the departures from HW equilibria seen in this study, because the expected number in the Aa and aa cells for each RFLP is small. At this point, whether the population structure or the selection procedure created the departures from equilibria seen in the apo AI-CIII-AIV region in this sample is unclear, but it seems likely that chance has played a major role.

Nonrandom frequency associations were also present between alleles at different RFLP loci in the FCs. Both composite diallelic and triallelic disequilibria were significant. It should be noted that the presence of significant departures from HW equilibria or diallelic disequilibria did not predict the presence of disequilibria at the higher-order levels (triallelic or quadriallelic), or vice versa. The composite diallelic disequilibria were composed primarily of associations either between alleles on the same chromosome or between alleles on both the same and opposite chromosomes. Also, all of the pairs of RFLP loci had at least one measure of disequilibrium at its maximum possible value determined by the allele frequencies. The presence of triallelic disequilibria, like the departures from HW equilibria, suggests that the allelic frequency associations may be due to the arrangement of haplotypes into genotypes. As mentioned above, the selection for health, small sample size, and data exclusion due to methodological constraints may be affecting these pairwise results as well. For each pairwise analysis, six of the nine genotype cells had expected values less than 5 because the common allele at each RFLP had a frequency of about .9 and because the sample size was relatively small. Thus, chance again is probably responsible for many of the triallelic disequilibria being at their bounds. Also, the pairwise analyses were done without correcting for multiple tests. So, the significant disequilibria may be due to chance—although this seems unlikely, because the power to detect disequilibria is generally low (Thompson et al. 1988).

Inferences made about the composite diallelic and triallelic disequilibria depended on the variance term used, either complete or reduced. The Weir and Cockerham X^2 value was consistently larger than the traditional value, although the inferences made were the same. Thompson et al. (1988) found that it takes very large sample sizes to detect diallelic gametic disequilibrium, especially if negative, when the more common allele has a frequency of approximately .9 at both loci. This suggests that some of the other disequilibria in this region may in fact be significantly different from zero.

The pairwise analysis also demonstrated that the distance between RFLP markers does not predict the magnitude or significance of diallelic, triallelic, or quadriallelic disequilibrium. This phenomenon at the diallelic level has also been reported by others who have studied this region (Ferns and Galton 1986; Antonarakis et al. 1988; Thompson et al. 1988). An extreme example was reported by Sinnock and Sing (1972), who demonstrated, with a very large sample, that significant associations can occur between markers on different chromosomes. Thompson et al. (1988) suggest that the lack of correlation between disequilibrium estimates and distance may be due to low power in the ability to detect disequilibria. Another explanation is that in small regions mutation and the evolutionary forces that affect allele frequencies—forces such as selection, migration, and drift—may have more of an impact on the underlying genetic structure than does recombination (Borresen et al. 1988). These disequilibrium results are relevant in the interpretation of associations between RFLPs and phenotypic variation.

The inconsistencies among RFLP-phenotype association studies may be due to interpopulation differences in the genetic structure at the apo AI-CIII-AIV region. The RFLP allele frequencies in the apo AI-CIII-AIV region in this Canadian sample did not differ from frequencies reported in other Caucasian populations (Paul et al. 1987; Thompson et al. 1988; Wile et al. 1989). There is evidence in the literature that other populations, such as the Japanese (Paul et al. 1987; Thompson et al. 1988), Chinese (Rees et al. 1985*a*), and American blacks (Antonarakis et al. 1988; Thompson et al. 1988), have different allele frequencies at some of the RFLPs in the apo AI-CIII-AIV region. As demonstrated in the HW analysis, even when allele frequencies are similar, how they arrange into genotypes may differ—and thus the genetic structure of the populations may differ. Although not addressed in this paper, the associations between alleles among loci may also vary among populations. Antonarakis et al. (1988) reported that, in the apo AI-CIII-AIV region, 13/45 deviations from gametic diallelic equilibrium were significant in a Mediterranean sample but that only 4/45 were significant in American

blacks. This suggests that the evolutionary history and effects of recombination in this region have varied among populations. Thus, comparisons of RFLP-phenotype association studies may not be appropriate, because the allele frequencies and associations between the RFLPs and the functional mutation may be different in different populations.

The presence of significant deviations from equilibrium should be taken into account in the design of RFLP-phenotype association studies. Zerba et al. (in press) demonstrated that departures from HW equilibria in a sample (and not in the larger population) that are due to nonevolutionary factors will cause an underestimation or overestimation of the genetic variance. This error in estimated variance biases the estimated sum-of-squares regression and can lead to inferences about the association between some continuous-trait variation and variation at a marker locus, inferences which do not hold for the population if HW equilibria in the population represents the "truth." Therefore, care must be taken to ensure that the sample used is representative of the population about which inferences are being made.

Since significant pairwise disequilibria may be present between alleles in a sample, the RFLP loci cannot be considered to be statistically independent. Thus, comparisons of inferences made from studies using different RFLPs may not be warranted (Chakraborty et al. 1986; Hanis et al. 1991). Rather than comparing results from multiple single-RFLP-phenotype association studies to determine the one RFLP most associated with the trait of interest, one can consider the patterns of associations of multiple RFLPs to determine whether variation in a region is associated with variation in the trait (Kessling et al., submitted). Some investigators have suggested using only one of a pair of loci in disequilibrium, making the assumption that, when there is association, no extra information would be gained by using both loci (e.g., see Schwartz et al. 1990). As can be seen from the data presented here, where the *Xmn*I(1) RFLP locus is in disequilibrium with both the *Taq*I and *Pvu*II RFLPs but where the *Taq*I and *Pvu*II RFLPs are in equilibrium, this assumption does not hold. Thus, using significant pairwise disequilibria between markers as a way to choose those markers for RFLP-phenotype association studies is not a valid strategy.

Last, our study emphasizes the lack of correlation between the level of disequilibrium and the distance between two RFLP loci. This finding suggests that a significant association between an RFLP and some trait should not be interpreted as meaning that the functionally important mutation in linkage disequilibrium with the RFLP is closer in distance to that specific RFLP than it is to other RFLPs in the region that may not have been associated with the continuous trait. Rather than the one-way analysis of variance or χ^2 test, cladistics (Templeton et al. 1987, 1988), an approach using evolutionary information to develop a testing scheme to detect the haplotypes carrying functionally important mutations, may prove to be a more effective strategy in the search for those individuals who carry functional mutations affecting a phenotype (Sing et al., in press).

In conclusion, information about departures from HW equilibria and pairwise disequilibria is necessary for interpreting and selecting appropriate analytical strategies for RFLP-phenotype association studies. When different populations have different underlying genetic structures, it is inappropriate to compare RFLP-phenotype association studies among them. Also, when RFLP-phenotype association studies are designed, the sample should be selected to ensure that only one Mendelian population is represented. Finally, because pairwise disequilibria are not directly related to physical distance in small regions of DNA, inferences about which RFLP is closest to a functional mutation will be misleading if they are made by comparing results of multiple single-RFLP analyses.

Acknowledgments

We would like to thank both the administration at Hydro-Quebec and Suzanne Lussier-Cacan, Marianne Xhignesse, Odette Bouffard, Ann Chamberland, and Suzanne Ouellette at the Institut de Recherches Cliniques de Montreal. Also, thanks goes to Drs. Jim Crow, Kim Zerba, Sharon Reilly, Tim Rebbeck, Pat Moll, and Bruce Weir for their helpful discussion and suggestions. This work was supported in part by grants from Health and Welfare, Canada (6605-2087-52), the Medical Research Council of Canada (MA-5427), the Medical Research Council of Canada/ CIBA Geigy Canada Ltd (UI-0029), L'Institut de Recherche en Santé et en Securité du Travail (IRSST) du Québec, La Succession J. A. DeSeve and Hydro-Quebec, and the NIH (HL39107). A. M. K. was supported as a postdoctoral fellow under a fellowship of the Canadian Medical Research Council.

References

Aburatani H, Matsumoto A, Itoh H, Murase T, Takaku F, Itakura H (1988) Deoxyribonucleic acid (DNA) polymor-

phism in the apolipoprotein AI gene: a study in a Japanese population. Jpn J Med 27:56–59

Antonarakis SE, Oettgen P, Chakravarti A, Halloran SL, Hudson RR, Feisee L, Karathanasis SK (1988) DNA polymorphism haplotypes of the human apolipoprotein ApoA1-ApoC3-ApoA4 gene cluster. Hum Genet 80:265–273

Berg K (1983) Genetics of coronary heart disease. Prog Med Genet 5:35–90

Bodmer W, Thomson G (1977) Population genetics and evolution of the HLA system. In: Dausset J, Svejgaard A (eds) HLA and disease. Williams & Wilkins, Baltimore, pp 280–295

Borresen AL, Moller P, Berg K (1988) Linkage disequilibrium analyses and restriction mapping of four RFLPs at the proα2(I) collagen locus: lack of correlation between linkage disequilibrium and physical distance. Hum Genet 78:216–221

Breslow JL (1988) Apolipoprotein genetic variation and human disease. Physiol Rev 68:85–132

Chakraborty R, Hanis CL, Boerwinkle E (1986) Effect of a marker locus on the quantitative variability of a risk factor to chronic diseases. Am J Hum Genet 39:A231

Cohen T, Karathanasis SK, Kazazian HH, Antonarakis SE (1986) DNA polymorphic sites in the human ApoAI-CIII-AIV cluster: TaqI and AvaI. Nucleic Acids Res 14: 1924

Coleman RT, Gonzalez PA, Funke H, Assmann G, Levy-Wilson B, Frossard PM (1986) Polymorphisms in the apolipoprotein AI-CIII gene complex. Mol Biol Med 3: 213–228

Cooper DN, Clayton JF (1988) DNA polymorphism and the study of disease associations. Hum Genet 78:299–312

Davignon J, Dufour R, Cantin M (1983) Atherosclerosis and hypertension. In: Genest J, Kuchel O, Hamet P, Cantin M (eds) Hypertension: physiopathology and treatment, 2d ed. McGraw-Hill, New York, pp 810–852

Ferns GAA, Galton DJ (1986) Haplotypes of the human apoprotein AI-CIII-AIV gene cluster in coronary atherosclerosis. Hum Genet 73:245–249

Ferns GAA, Stocks J, Ritchie C, Galton DJ (1985) Genetic polymorphisms of apolipoprotein C-III and insulin in survivors of myocardial infarction. Lancet 2:300–303

Fielding CJ, Shore VG, Fielding PE (1972) A protein cofactor of lecithin:cholesterol acyltransferase. Biochem Biophys Res Commun 46:1493–1498

Fisher EA, Coates PM, Cortner JA (1989) Gene polymorphisms and variability of human apolipoproteins. Annu Rev Nutr 9:139–160

Gordon T, Castelli WP, Hjortland MC, Kannel WB, Dawber TR (1977) High density lipoprotein as a protective factor against coronary heart disease. Am J Med 62: 707–714

Gotto AM, Pownall HJ, Havel RJ (1986) Introduction to the plasma lipoproteins. Methods Enzymol 128:3–41

Hamsten A, Iselius L, Dahlen G, deFaire U (1986) Genetic and cultural inheritance of serum lipids, low and high density lipoprotein cholesterol and serum apolipoproteins A-I, A-II and B. Atherosclerosis 60:199–208

Hanis CL, Hewett-Emmett D, Douglas TC, Bertin TK, Schull WJ (1991). Effects of the apolipoprotein E polymorphism on levels of lipids, lipoproteins and apolipoproteins among Mexican-Americans in Starr County, Texas. Arteriosclerosis Thromb 11:362–370

Hegele RA, Plaetke R, Lalouel JM (1990) Linkage disequilibrium between DNA markers at the low-density lipoprotein receptor gene. Genet Epidemiol 7:69–81

Herbert PN, Assmann G, Gotto AM, Fredrickson DS (1983) Familial lipoprotein deficiency: abetalipoproteinemia, hypobetalipoproteinemia, and Tangier disease. In: Stanbury JB, Wyngaarden JB, Fredrickson DS, Goldstein JL, Brown MS (eds) The metabolic basis of inherited disease, 5th ed. McGraw-Hill, New York, pp 589–621

Higgins MW, Luepker RV (eds) (1989) Trends and determinants of CHD mortality: international comparisons. Int J Epidemiol [Suppl 1]: S1–S232

Hill WG (1974) Estimation of linkage disequilibrium in randomly mating populations. Heredity 33:229–239

Kannel WB, McGee D, Gordon T (1976) A general cardiovascular risk profile: the Framingham study. Am J Cardiol 38:46–51

Karathanasis SK, Ferris E, Haddad IA (1987) DNA inversion within the apolipoproteins AI/CIII/AIV-encoding gene cluster of certain patients with premature atherosclerosis. Proc Natl Acad Sci USA 84:7198–7202

Karathanasis SK, McPherson J, Zannis VI, Breslow JL (1983*a*) Linkage of human apolipoproteins A-I and C-III genes. Nature 304:371–373

Karathanasis SK, Yunis I (1986) Structure, evolution, and tissue-specific synthesis of human apolipoprotein AIV. Biochem 25:3962–3970

Karathanasis SK, Zannis VI, Breslow JL (1983*b*) A DNA insertion in the apolipoprotein A-I gene of patients with premature atherosclerosis. Nature 305:823–825

Kessling AM, Berg K, Mockleby E, Humphries SE (1986) DNA polymorphisms around the apo AI gene in normal and hyperlipidaemic individuals selected for a twin study. Clin Genet 29:485–490

Kessling AM, Horsthemke B, Humphries SE (1985) A study of DNA polymorphisms around the human apolipoprotein AI gene in hyperlipidaemic and normal individuals. Clin Genet 28:296–306

Kessling AM, Ouellette S, Bouffard O, Chamberland A, Betard C, Selinger E, Xhignesse M, et al. Patterns of association between genetic variability in apolipoprotein (apo) B, apo AI-CIII-AIV and cholesterol ester transfer protein gene regions and quantitative variation in lipid and lipoprotein traits: influence of gender and exogenous hormones (submitted)

Kessling AM, Rajput-Williams J, Bainton D, Scott J, Miller NE, Baker I, Humphries SE (1988*a*) DNA polymor-

phisms of the apolipoprotein AII and AI-CIII-AIV genes: a study in men selected for differences in high-density-lipoprotein cholesterol concentration. Am J Hum Genet 42:458–467

Kessling AM, Taylor R, Temple A, Hutson J, Hidalgo A, Humphries SE (1988*b*) A *Pvu*II polymorphism in the 5′ flanking region of the apolipoprotein AIV gene: its use to study genetic variation determining serum lipid and apolipoprotein concentration. Hum Genet 78:237–239

Kottke BA (1986) Lipid markers for atherosclerosis. Am J Cardiol 57:11C–17C

Kottke BA, Zinsmeister AR, Holmes DR, Kneller RW, Hallaway BJ, Mao SJT (1986) Apolipoproteins and coronary artery disease. Mayo Clin Proc 61:313–320

Laberge C (1966) Prospectus for genetic studies in the French Canadians, with preliminary data on blood groups and consanguinity. Bull Johns Hopkins Hosp 118:52–68

Law SW, Gray G, Brewer HB, Sakaquchi AY, Naylor SL (1984) Human apolipoprotein A-I and C-III genes reside in the p11-p13 region of chromosome 11. Biochem Biophys Res Commun 118:934–942

Litt M, Jorde LB (1986) Linkage disequilibria between pairs of loci within a highly polymorphic region of chromosome 2q. Am J Hum Genet 39:166–178

Lusis AJ (1988) Genetic factors affecting blood lipoproteins: the candidate gene approach. J Lipid Res 29:397–429

Lussier-Cacan S, Xhignesse M, Dubreuil-Quidoz S, Richard A, Kessling AM, Davignon J. Estimates of the distribution of plasma lipid and lipoprotein traits in a Quebec population selected for health (submitted)

Maciejko JJ, Holmes DR, Kottke BA, Zinsmeister AR, Dinh DM, Mao SJT (1983) Apolipoprotein A-I as a marker of angiographically assessed coronary artery disease. N Engl J Med 309:385–389

Miller NE (1983) High density lipoprotein, atherosclerosis, and ischaemic heart disease. In: Miller NE (ed) Atherosclerosis: mechanisms and approaches to therapy. Raven, New York, pp 153–168

Monsalve MV, Young R, Wiseman A, Dhamu S, Powell JT, Greenhalgh RM, Humphries SE (1989) Study of DNA polymorphisms of the apolipoprotein AI-CIII-AIV gene cluster in patients with peripheral arterial disease. Clin Sci 76:221–228

Oettgen P, Antonarakis SE, Karathanasis SK (1986) PvuII polymorphic site upstream to the human ApoCIII gene. Nucleic Acids Res 14:5571

Oram JF, Albers JJ, Cheung MC, Bierman EL (1981) The effects of subfractions of high density lipoprotein on cholesterol efflux from cultured fibroblasts. J Biol Chem 256: 8348–8356

Ordovas JM, Schaefer EJ, Salem D, Ward RH, Glueck CJ, Vergani C, Wilson PWF et al (1986) Apolipoprotein A-I gene polymorphism associated with premature coronary artery disease and familial hypoalphalipoproteinemia. N Engl J Med 314:671–677

Paul H, Galton D, Stocks J (1987) DNA polymorphic patterns and haplotype arrangements of the apo A-I, apo C-III, apo A-IV gene cluster in different ethnic groups. Hum Genet 75:264–268

Paulweber B, Friedl W, Krempler F, Humphries SE, Sandhofer F (1988) Genetic variation in the apolipoprotein AI-CIII-AIV gene cluster and coronary heart disease. Atherosclerosis 73:125–133

Pownall HJ, Gotto AM (1983) Plasma apolipoproteins: composition, structure, and function. In: Miller NE (ed) Atherosclerosis: mechanisms and approaches to therapy. Raven, New York, pp 125–141

Rees A, Shoulders CC, Stocks J, Galton DJ, Baralle FE (1983) DNA polymorphism adjacent to human apoprotein A-1 gene: relation to hypertriglyceridaemia. Lancet 1:444–446

Rees A, Stocks J, Sharpe CR, Vella MA, Shoulders CC, Katz J, Jowett NI, et al (1985*a*) Deoxyribonucleic acid polymorphism in the apolipoprotein A-I-C-III gene cluster: association with hypertriglyceridemia. J Clin Invest 76:1090–1095

Rees A, Stocks J, Williams LG, Caplin JŁ, Jowett NI, Camm AJ, Galton DJ (1985*b*) DNA polymorphisms in the apolipoprotein C-III and insulin genes and atherosclerosis. Atherosclerosis 58:269–275

Rothman KJ (1990) No adjustments are needed for multiple comparisons. Epidemiology 1:43–46

SAS Institute Inc (1985) SAS/IML User's guide, version 5 ed. SAS Institute, Cary, NC

Schaefer EJ, Eisenberg S, Levy RI (1978) Lipoprotein apoprotein metabolism. J Lipid Res 19:667–687

Schamaun O, Olaisen B, Mevag B, Gedde-Dahl T, Ehnholm C, Teiseberg P (1984) The two apolipoprotein loci apo A-I and apo A-IV are closely linked in man. Hum Genet 68:181–184

Schull WJ, Hanis CL (1990) Genetics and public health in the 1990s. Annu Rev Public Health 11:105–125

Schwartz RC, Sarfarazi M, Tsipouras P (1990) Linkage disequilibrium between two restriction sites in the COL2A1 gene. Am J Hum Genet 46:1214–1215

Shoulders CC, Ball MJ, Mann JI, Baralle FE (1986) Genetic marker in apolipoprotein AI/CIII gene complex associated with hypercholesterolaemia. Lancet 2:1286

Sing CF, Haviland MB, Zerba KE, Templeton AR. Application of cladistics to the analysis of genotype-phenotype relationships. Eur J Epidemiol (in press)

Sing CF, Moll PP (1990) Genetics of atherosclerosis. Annu Rev Genet 24:171–187

Sinnock P, Sing CF (1972) Analysis of multilocus genetic systems in Tecumseh, Michigan. II. Consideration of the correlation between nonalleles in gametes. Am J Hum Genet 24:393–415

Sokal RR, Rohlf FJ (1981) Biometry. WH Freeman, New York

Templeton AR, Boerwinkle E, Sing CF (1987) A cladistic

analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. I. Basic theory and an analysis of alcohol dehydrogenase activity in Drosophila. Genetics 117:343–351

Templeton AR, Sing CF, Kessling A, Humphries S (1988) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. II. The analysis of natural populations. Genetics 120:1145–1154

Thompson EA, Deeb S, Walker D, Motulsky AG (1988) The detection of linkage disequilibrium between closely linked markers: RFLPs at the AI-CIII apolipoprotein genes. Am J Hum Genet 42:113–124

Weir BS, Cockerham CC (1979) Estimation of linkage disequilibrium in randomly mating populations. Heredity 42:105–111

——— (1989) Complete characterization of disequilibrium at two loci. In: Feldman MW (ed) Mathematical evolutionary theory. Princeton University Press, Princeton, NJ, pp 86–110

Wile DB, Barbir M, Gallagher J, Myant NB, Ritchie CD, Thompson GR, Humphries SE (1989) Apolipoprotein A-I gene polymorphisms: frequency in patients with coronary artery disease and healthy controls and association with serum apo A-I and HDL-cholesterol concentration. Atherosclerosis 78:9–18

Xhignesse M, Lussier-Cacan S, Sing CF, Kessling AM, Davignon J. Influences of common variants of apolipoprotein E on measures of lipid metabolism in a sample selected for health. Arteriosclerosis (in press)

Zerba KE, Kessling AM, Davignon J, Sing CF. Genetic structure and the search for genotype-phenotype relationships: an example from disequilibrium in the apo B gene region. Genetics (in press)