

## ORIGINAL INVESTIGATION

Kim E. Zerba · Robert E. Ferrell · Charles F. Sing

**Genetic structure of five susceptibility gene regions for coronary artery disease: disequilibria within and among regions**

Received: 15 January 1998 / Accepted: 24 June 1998

**Abstract** We analyzed two-locus disequilibria for 16 polymorphic loci of seven susceptibility genes for coronary artery disease located in five chromosomal regions distributed across four chromosomes. Included were the genes coding for apolipoprotein B (*ApoB*, chromosome 2, four marker loci), lipoprotein lipase (*LPL*, chromosome 8, three marker loci), apolipoproteins AI, CIII, AIV (*ApoAI-CIII-AIV*, chromosome 11, three marker loci), apolipoprotein E (*ApoE*, chromosome 19, two marker loci), and the low density lipoprotein receptor (*LDLR*, chromosome 19, four marker loci). Our sample included 540 unrelated individuals from the Rochester, Minn. population. There were no statistically significant deviations of single-locus genotypes from Hardy-Weinberg equilibrium. The strongest associations within genes were for composite diallelic disequilibria; 17/19 were significant (13 at  $P < 0.001$ , 1 at  $P < 0.01$ , 3 at  $P < 0.05$ ). These observations suggest marker alleles within genes have a shared evolutionary history reflected by disequilibria that have not been dissipated by recombination. Disequilibrium was not generally concordant with the physical orderings of markers. Only two significant higher-order disequilibria were observed although 12 triallelic disequilibria were at maximum possible values. We observed 19 statistically significant disequilibria ( $P < 0.05$ ; 4 composite diallelic, 13 triallelic, and 2 quadriallelic) between 101 pairs of marker loci, where each locus in a pair was from a different unlinked region. These unexpected results are most likely ex-

plained by recent historical factors, including worldwide population expansion and amalgamation with continuous admixture, that influence the genetic structure (organization of alleles and non-alleles into genotypes) of a population. We conclude that disequilibria between loci from unlinked regions may be more extensive than is commonly assumed. Our findings also suggest that it is, on average, at least 15 times more likely to not detect significant disequilibrium among unlinked loci when it is really present than to make a false positive inference. Disequilibria between functional loci within or between regions will impact estimates of genetic variance associated with particular functional mutations within a susceptibility gene region.

**Introduction**

There is great interest in the vast number of polymorphic markers being discovered at the DNA level for the study of complex, multifactorial human diseases (e.g., Brower 1998; Brown and Hartwell 1998; Clark et al. 1998; Ghosh 1995; Kahn 1996; Murray et al. 1994; Nickerson et al. 1998; Weissman 1995). The ultimate goal of many of these studies is to identify and characterize the effects of functional mutations that influence phenotypic variation in traits that determine interindividual variation in susceptibility to complex diseases in the population at large. For many complex diseases few gene regions involved in the disease are known and new candidate regions must first be identified (Bickeböllner et al. 1995; Briscoe et al. 1994; Devlin and Risch 1995; Ewens and Spielman 1995; Hill and Weir 1994; Jorde 1995a; Kaplan 1995; Kaplan et al. 1995; Kruglyak 1997; Risch and Merikangas 1996; Stephens et al. 1994; Tanksley 1993; Terwilliger 1995; Weeks and Lathrop 1995; Weir 1996). The situation is different for coronary artery disease (CAD), however, since numerous candidate susceptibility genes have already been identified because their products are involved in lipid metabolism and are known to influence risk of disease, or al-

K.E. Zerba (✉)<sup>1</sup> · C.F. Sing  
Department of Human Genetics, University of Michigan,  
Ann Arbor, MI 48109, USA  
e-mail: kzerba@mail.colgate.edu,  
Tel.: +1-315-2287705, Fax: +1-315-2287997

R.E. Ferrell  
Department of Human Genetics, University of Pittsburgh,  
Pittsburgh, PA 15262, USA

<sup>1</sup>Present address:  
Biology Department, Colgate University, 13 Oak Drive,  
Hamilton, NY 13346, USA

lelic variations have been associated with interindividual variation in susceptibility to CAD.

Many studies of the statistical relationships between variation of multiple marker loci within candidate CAD susceptibility genes and variation in biological risk factors are carried out with the intent of identifying mutations with functional phenotypic effects. Separate analyses of variance (ANOVAs) are often used on each marker to detect associations of marker genotypes with phenotypic variation. Inferences are then made about the possible nearby location of an unknown functional mutation, which is assumed to be associated with variability at a marker locus (linkage disequilibrium). Such multiple-marker association studies make the implicit assumption that the marker allele frequencies are independent. Marker allele frequencies are not generally independent, however, since statistically significant associations are commonly observed between the relative allele frequencies of the polymorphic marker loci within susceptibility gene regions (e.g., Benlian et al. 1991; Borresen et al. 1988; Haviland et al. 1991; Hegele et al. 1990; Thompson et al. 1988; Zerba et al. 1991).

Historical evolutionary processes, including recombination, mutation, selection, admixture, migration, and finite population size, and the sampling design specific to a particular study, influence the associations between relative allele frequencies of marker loci. The rare historical events that influence the reproduction of a population are unpredictable in time and the chromosomes they will affect. Moreover, the relative frequency of a new mutation may be of the same order as that of recombination in short DNA regions of less than 50 kb that are typical of many candidate gene regions. A predictable relationship between the magnitude of linkage disequilibrium and physical distance is not expected in such regions (Jorde 1995a). Inferences about the location of a functional mutation based on the magnitude of a single marker genotype–phenotype association are thus unreliable (Zerba et al. 1991).

Cladistic analysis has been developed, which incorporates the shared evolutionary history among multiple marker alleles in short segments of the same chromosome, to identify haplotypes which may be carrying functional mutations (Haviland et al. 1995, 1997; Sing et al. 1992b; Templeton 1995, 1996; Templeton and Sing 1993; Templeton et al. 1987, 1992, 1988). The task is then to identify the functional mutation, or mutations, responsible for a phenotypic effect that is associated with a haplotype is not trivial. Once a functional mutation in a gene has been identified, however, its impact on quantitative phenotypic variation in the population can then be evaluated.

Phenotypic variation associated with a functional mutation is dependent on at least four factors related to frequency and scale: (1) the frequency distribution of alleles at the locus of the functional mutation; (2) the deviations of the average phenotype of the individuals bearing the functional mutation from the phenotypic average in the population (scale); (3) phenotypic correlations of the trait with other traits in the population that are under the influence of other genes (scale); and (4) allelic frequency

associations (disequilibria) within and among other loci in the population. A functional mutation in a candidate gene region for a quantitative trait involved in susceptibility to a complex, multifactorial disease is likely to be only one of many possible functional mutations that may influence such a trait. There may be multiple functional mutations at different sites within a region as well as in different regions on the same or on different chromosomes. Since there can be associations between the relative allele frequencies of unlinked genes in the population (Sinnock and Sing 1972), meaningful inferences about the impact of any particular functional mutation will be dependent on the knowledge about such associations.

There have been no published studies of the genetic structure (organization of alleles and non-alleles into genotypes) of multiple markers within and among multiple unlinked candidate CAD susceptibility genes in the same sample from a single population. The strength of inferences about genetic structure is of concern since the power to detect disequilibrium given the sample sizes used in most studies may be low (Thompson et al. 1988; Ward and Sing 1970; Zerba et al. 1991). The complement to power, the type II error probability of not detecting statistically significant disequilibrium when it is present, may therefore be high in most studies.

This paper estimates the genetic structure using two-locus disequilibria statistics for 16 polymorphic marker loci distributed within and among five regions located on four chromosomes in a sample of 540 unrelated individuals from the Rochester, Minn. population. These regions include *the ApoB*, *LPL*, *ApoAI*, *ApoCIII*, *ApoAIV*, *ApoE*, and *LDLR* genes. We find that disequilibrium between loci within a gene region is not concordant with their physical ordering and that disequilibrium between unlinked regions is more extensive than is commonly assumed. We also estimated the type II error probabilities for disequilibria among unlinked loci. Our findings suggest that, at levels of disequilibria observed in our sample, it is, on average, at least 15 times more likely to not detect significant disequilibrium among unlinked loci when it is really present than to make a false positive inference.

---

## Materials and methods

### Sample

Individuals were sampled from 285 three- and four-generation pedigrees ascertained without regard to health status through elementary school children of the Rochester, Minn. population as part of the Rochester Family Heart Study (described by Moll et al. 1989; Turner et al. 1989). The sample that we studied was constructed in the following way. For each of the 285 unrelated spouse pairs from the middle generation of parents, we chose one of the following combinations of unrelated individuals with the most available data on the 16 markers: (1) spouse pair's parents; (2) wife and husband's parents; (3) husband and wife's parents; or (4) spouse pair. When two or more of these combinations of unrelated individuals would have yielded the largest number of individuals for the pedigree being considered, one combination was selected at random. Also, there were eight spouses who had married twice. In these cases, there were two

possible spouse pairs from which to choose combinations of unrelated individuals (1–4 above). We chose the spouse pair yielding the most data. This procedure resulted in a sample of 540 unrelated individuals, including 298 females and 242 males.

#### Regions and markers

The gene, its markers and their locations are presented in Table 1. The products of these genes play a central role in lipid metabolism and allelic variation in each gene has been implicated in determining interindividual variation in susceptibility to CAD. The gene coding for apolipoprotein B (*ApoB*) is on the short arm of chromosome 2, 43 kb long, and includes 29 exons (Blackhart et al. 1986). ApoB is important in the assembly of triglyceride-rich lipoproteins (Leighton et al. 1990), a major component of low density lipoproteins (LDL), and the ligand for the LDL receptor (LDLR; Brown and Goldstein 1986). Increased plasma LDL cholesterol levels convey increased risk to CAD.

The gene coding for lipoprotein lipase (*LPL*) is on the short arm of chromosome 8, 30 kb long, and includes ten exons (Deeb and Peng 1989). LPL is a catalyst in triacylglycerol metabolism, influencing the catabolism of triglyceride-containing lipoproteins and delivery of fatty acids to tissues (Kirchgeßner et al. 1989; Oka et al. 1991).

Three genes coding for apolipoproteins AI (*ApoAI*), CIII (*ApoCIII*), and AIV (*ApoAIV*) occur in a 22 kb region on the long arm of chromosome 11 (Karathanasis 1985). ApoAI is a key part of high density lipoprotein (HDL; Schaefer et al. 1978) and a cofactor of lecithin cholesterol acyltransferase (Fielding et al. 1972) in reverse cholesterol transport. ApoCIII and ApoAIV are also a part of HDL. Low plasma levels of HDL cholesterol are associated with increased risk to CAD.

The gene coding for apolipoprotein E (*ApoE*) is on the long arm of chromosome 19, 3.6 kb in length, and includes four exons. ApoE influences the metabolism of Apo B-containing lipoproteins (Mahley 1988). There are three common alleles in most populations ( $\epsilon_2$ ,  $\epsilon_3$ ,  $\epsilon_4$ ) that combine to produce six genotypes (Davignon et al. 1988).

This is one of the few examples where there is evidence for variability in the functional characteristics of the gene product among common allelic variations (Rall et al. 1982). At the organismic level, allelic variation in *ApoE* contributes to predicting interindividual phenotypic differences in several traits that are measures of lipid metabolism (Davignon et al. 1988; Kaprio et al. 1991; Sing and Davignon 1985; Sing et al. 1992a; Xhignesse et al. 1991).

The gene coding for the LDLR is on the short arm of chromosome 19, 45 kb long, and includes 18 exons (Südhof et al. 1985). LDL cholesterol is removed from the plasma by non-receptor-mediated pathways and by receptor-mediated endocytosis through the LDLR (Goldstein and Brown 1989). Particular mutations in the *LDLR* gene have been associated with excessive plasma cholesterol levels and premature CAD (for review see Hobbs et al. 1990).

#### Marker typings

Blood samples were collected by venipuncture into EDTA. Markers of all but the *ApoE* gene were typed by DNA amplification using primers flanking each polymorphic site followed by digestion of the amplification product with the appropriate restriction endonuclease and electrophoresis in 1% agarose gels. The three common ApoE protein isoforms coded by the three common *ApoE* alleles were determined by isoelectric focusing using the method described by Kamboh et al. (1988). Since these isoforms correspond directly to differences at amino acid residue sites 112 and 158 (Rall et al. 1982), we used the isoform typings to infer the common *ApoE* genotype of each individual (Reilly et al. 1992). Parents and children of the unrelated sample of individuals used in this study were also typed for the purpose of checking for typing errors. We used the computer program PATCH (Wijsman 1987) to identify inheritance inconsistencies. When an inconsistency was found, all members of a pedigree were retyped to resolve the error.

**Table 1** Gene regions and marker locations (AA amino acid)

Gene region	Chromosome /locus	Marker	Location	Reference
<i>ApoB</i>	2/p23–p24	<i>I/D</i>	Signal peptide region (5')	Boerwinkle and Chan (1989)
		<i>XbaI</i>	Exon 26	Berg et al. (1986)
		<i>MspI</i>	Exon 26	Priestley et al. (1985)
		<i>EcoRI</i>	Exon 29	Shoulders et al. (1985)
<i>LPL</i>	8/p22	<i>PvuII</i>	Intron 6	Fisher et al. (1987); Li et al. (1988)
		<i>HindIII</i>	Intron 8	Heinzmann et al. (1987)
		<i>BstNI</i>	Intron 8	Funke et al. (1988)
<i>ApoAI–CIII–AIV</i>	11/q23.3	<i>I/D</i>	5' to <i>ApoAI</i>	Coleman et al. (1986)
		<i>PstI</i>	3' to <i>ApoAI</i>	Kessling et al. (1985)
		<i>SstI</i>	5' to <i>ApoCIII</i>	Rees et al. (1983)
<i>ApoE</i>	19/q13.2	AA 112	AA 112	Rall et al. (1982)
		AA 158	AA 158	Rall et al. (1982)
<i>LDLR</i>	19/p13.3	<i>TaqI</i>	Intron 4	Yamakawa et al. (1987)
		<i>StuI</i>	Exon 8	Kotze et al. (1986)
		<i>HincII</i>	Exon 12	Leitersdorf and Hobbs (1988)
		<i>AvaII</i>	Exon 13	Hobbs et al. (1987); Leitersdorf and Hobbs (1988); Yamakawa et al. (1987)

Statistical analyses

The most frequent alleles of a pair of markers are designated A and B. All markers have only two codominant alleles. The observed relative frequencies of the single marker and two-marker genotypes were taken to be estimates of the relative genotype frequencies. Double heterozygotes were not distinguished, since complete phase information was not available for all individuals for all pairs of markers. There were nine two-marker genotype classes for each pair of marker loci. Variances of relative allele frequencies were estimated as described by Weir (1996). We characterized the genetic structure within and among regions using the approach and methodology outlined by Weir and Cockerham (1989). For each marker, we estimated the deviations of the three genotype frequencies from Hardy-Weinberg equilibrium expectations ( $D_A$ ). For marker pairs we estimated the composite diallelic ( $\Delta_{AB}$ ), two triallelic ( $D_{AAB}, D_{ABB}$ ), and quadriallelic disequilibria ( $\Delta_{AABB}$ ).

We estimated the bounds for each measure of disequilibrium. Bounds for  $D_A$  were determined according to Weir and Brooks (1986). The bound for  $D_{AB}$  is the same as that for  $D_{A/B}$  when their signs are the same. The signs of  $D_{AB}$  and  $D_{A/B}$  are more likely the same the more  $\Delta_{AB}$  is significantly different from zero. Their signs may be different, however, when  $\Delta_{AB}$  is not significantly different from zero. Differences in sign in this case depend on the distribution of individuals in the two possible double heterozygote classes; the signs are the same only at or near an equal distribution of individuals. Bounds for  $\Delta_{AB}$  were therefore estimated as twice the values for  $D_{AB}$  defined by Weir and Cockerham (1989), assuming a nearly equal distribution of individuals among the two double heterozygote classes. Bounds for the triallelic disequilibria were estimated according to Weir and Cockerham (1989). Bounds for quadriallelic disequilibrium are unknown. We also computed the standardized disequilibrium,  $\Delta'_{AB}$ , as the disequilibrium estimate divided by its positive upper or negative lower bound, corresponding to the sign of the estimate of  $\Delta_{AB}$ .

For each disequilibrium estimate, we tested the null hypothesis that a particular disequilibrium parameter estimate,  $\hat{\delta}$ , for the population was zero by

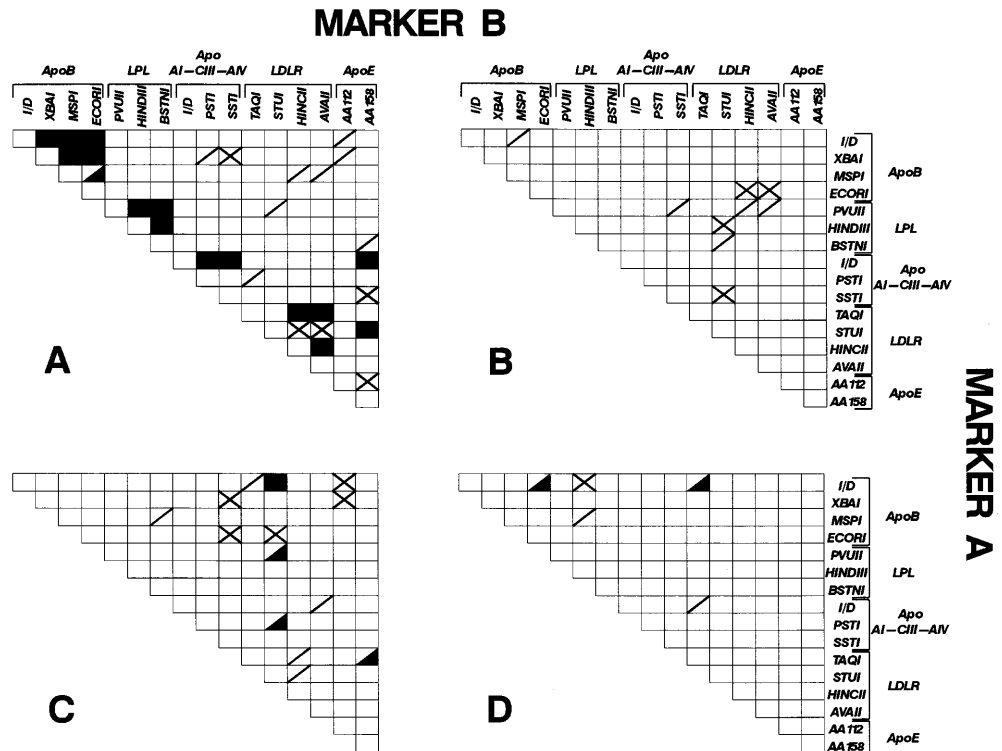
$$X^2 = \hat{\delta}^2 / \text{var}(\hat{\delta})_0 \quad (1)$$

where  $\text{var}(\hat{\delta})_0$  is the variance of  $\hat{\delta}$  estimated under the null hypothesis ( $\delta$  set to zero) and all other parameters, relative genotype frequencies and disequilibria, are at their respective point estimates. The variance of each disequilibrium was estimated according to Weir and Cockerham (1989). The  $X^2$  is approximately distributed as a  $\chi^2$  with one degree of freedom. We use the conventional statistical significance level of  $\alpha=0.05$  for single tests throughout and did not adjust the significance level for multiple tests. This is a conservative procedure for estimating patterns of significant disequilibrium given the low power to detect disequilibrium in general (Zerba et al. 1991).

We were interested in testing for differences in average composite diallelic disequilibrium among regions. It is well known that disequilibrium is a function of allele frequencies (Hedrick 1987; Lewontin 1988). It is not known in our study, however, whether the degree to which variability in the observed disequilibria among regions is due to real differences in disequilibrium among regions and/or differences in allele frequencies among sites/regions. Consequently, we tested homogeneity of the relative allele frequencies among sites, among regions, and among sites within regions using likelihood ratio statistics ( $G^2$ ; Weir 1996). We found significant heterogeneity of relative allele frequencies at all levels. Heterogeneity in allele frequencies among regions could result in significant differences in average disequilibrium among gene regions. Consequently, we performed linear regression analyses of composite diallelic disequilibrium estimates on common relative allele frequency estimates at the two loci involved in each disequilibrium estimate to obtain estimates of the relationships between variability in disequilibria and variability in common allele frequencies. Equality of average composite diallelic disequilibria among regions was then tested by ANOVA.

For each pair of unlinked marker loci (101 pairs altogether) and each of the four disequilibria (composite diallelic, two triallelic and quadriallelic) for that pair, we used computer simulation to determine the type II error probability of not detecting disequilibrium of the magnitude observed in our sample. We constructed a population with the relative genotype and allele frequencies and hence disequilibria observed in our sample. Each observed disequilibrium served as the alternative to the null hypothesis of that disequilibrium equal to zero for the type II error simulations. For each marker pair and each of the four disequilibria for that pair,

**Fig. 1A–D** Patterns of significant disequilibria within and among regions. **A** Composite diallelic. **B** Triallelic,  $D_{AAB}$ . **C** Triallelic,  $D_{ABB}$ . **D** Quadriallelic. Significance levels are indicated by slashes  $0.1 > Pr \geq 0.05$ , crosses  $0.05 > Pr \geq 0.01$ , half-filled squares  $0.01 > Pr \geq 0.001$ , filled squares  $Pr < 0.001$



**Table 2** Most frequent relative allele frequencies. In no case did the observed relative genotype frequencies deviate significantly from the Hardy-Weinberg expectation (AA amino acid)

Region	Marker	$\hat{P}_A$	$\text{var}(\hat{P}_A) \times 10^{-4}$
<i>ApoB</i>	<i>I/D</i>	0.67	1.87
	<i>XbaI</i>	0.54	2.23
	<i>MspI</i>	0.89	0.85
	<i>EcoRI</i>	0.88	0.99
<i>LPL</i>	<i>HindIII</i>	0.74	1.84
	<i>PvuII</i>	0.53	0.24
	<i>BstNI</i>	0.72	1.86
<i>ApoAI-CIII-AIV</i>	<i>I/D</i>	0.56	2.43
	<i>PstI</i>	0.94	0.55
	<i>SstI</i>	0.90	0.84
<i>ApoE</i>	AA 112	0.83	1.32
	AA 158	0.92	0.65
<i>LDLR</i>	<i>TaqI</i>	0.93	0.62
	<i>StuI</i>	0.95	0.41
	<i>HincII</i>	0.53	2.34
	<i>AvaII</i>	0.54	2.28

10,000 bootstrap samples were drawn from this population with replacement. For each replicate sample we estimated the disequilibrium coefficient of interest and tested the null hypothesis using (1). Type II error probability was determined as the proportion of the simulations in which the  $\chi^2$  was less than or equal to the critical value of 3.84 for a single degree of freedom test at  $\alpha=0.05$ . For each of the four types of disequilibria we then estimated the average type II error probability over the 101 pairs of unlinked marker loci for our study

**Table 3** Composite diallelic disequilibria within regions (AA amino acid)

Region: marker pair (A-B)	$\hat{\Delta}_{AB} \times 10$	$\hat{\Delta}'_{AB} \times 100$	$\chi^2$	<i>Pr</i>
<i>ApoB</i> :				
<i>I/D-XbaI</i>	0.80	22.35	71.69	<0.001
<i>I/D-MspI</i>	0.40	27.99	65.26	<0.001
<i>I/D-EcoRI</i>	-0.24	30.53	14.19	<0.001
<i>XbaI-MspI</i>	0.40	34.06	40.34	<0.001
<i>XbaI-EcoRI</i>	-0.54	48.26	56.44	<0.001
<i>MspI-EcoRI</i>	-0.12	46.70	7.64	0.006
<i>LPL</i> :				
<i>PvuII-HindIII</i>	0.96	34.72	93.92	<0.001
<i>PvuII-BstNI</i>	1.02	34.55	104.74	<0.001
<i>HindIII-BstNI</i>	1.88	49.92	441.73	<0.001
<i>ApoAI-CIII-AIV: I/D-PstI</i>				
<i>I/D-SstI</i>	0.34	49.96	33.46	<0.001
<i>PstI-SstI</i>	0.41	35.67	34.70	<0.001
<i>PstI-SstI</i>	-0.01	11.55	0.24	0.625
<i>ApoE</i> :				
AA 112-AA 158	-0.11	40.70	6.48	0.011
<i>LDLR</i> :				
<i>TaqI-StuI</i>	-0.02	32.13	0.89	0.344
<i>TaqI-HincII</i>	-0.41	61.46	48.92	<0.001
<i>TaqI-AvaII</i>	-0.38	57.16	43.89	<0.001
<i>StuI-HincII</i>	-0.09	20.01	4.34	0.035
<i>StuI-AvaII</i>	-0.09	19.69	4.28	0.039
<i>HincII-AvaII</i>	2.44	24.53	521.56	<0.001

## Results

The most frequent relative allele frequencies are presented in Table 2. For each region, markers are listed in order 5' to 3'. These relative allele frequencies ranged from 0.53, for *PvuII* and *HincII* in the *LPL* and *LDLR* gene regions, to 0.95, for *StuI* in the *LDLR* gene region.

There were no significant deviations of single-locus genotype frequencies from the Hardy-Weinberg expectation and none of the values of  $D_A$  were at the maximum possible values. Estimates of composite diallelic disequilibrium ( $\Delta_{AB}$ ) within regions,  $\Delta'_{AB} \times 100$ , and significance tests are presented in Table 3. Most estimates were significant (17/19). The exceptions were *PstI-SstI* in *ApoAI-CIII-AIV* and *TaqI-StuI* in *LDLR*. Note that for these two marker pairs, the 5' marker has a significant  $\Delta_{AB}$  with all other 3' markers. We observed no significant triallelic disequilibria, although 12 estimates were at the maximum possible values. We observed one significant quadriallelic disequilibrium (*I/D-EcoRI* in *ApoB*:  $\Delta_{AABB} = 1.10 \times 10^{-3}$ ,  $\chi^2 = 7.4$ ,  $Pr = 0.006$ ).

Patterns of significant disequilibria within and among regions are presented in Fig. 1. The most significant composite diallelic disequilibria were most frequently observed within regions (Fig. 1A). There were 4/101 composite diallelic disequilibria among regions, however, that were significant. There were 13/202 significant triallelic disequilibria among regions and only one within regions (Fig. 1B, C). We observed 2/101 significant quadriallelic disequilibria among regions (Fig. 1D). If the significance

**Table 4** Type II error probabilities for unlinked loci ( $n=101$  pairs of marker loci). Each of the 101 type II error probabilities was based on 10,000 bootstrap samples

	$\Delta_{AB}$	$D_{AAB}$	$D_{ABB}$	$\Delta_{AABB}$
Mean	0.83	0.83	0.81	0.79
SD	0.16	0.15	0.19	0.23
Minimum	0.05	0.36	0.18	0.00
Maximum	0.95	0.99	0.99	0.97

level is extended to include  $\alpha=0.1$ , there are an additional eight significant composite diallelic, eight triallelic, and two quadriallelic disequilibria among regions, and there an additional two significant triallelic disequilibria within regions.

We observed significant heterogeneity of relative allele frequencies among sites, among regions, among sites within regions, and among sites within every region specifically ( $Pr<0.001$  in all cases). Variability in composite diallelic disequilibrium estimates was significantly negatively associated with variability in common relative allele frequencies at the two loci ( $F=7.42$ ,  $Pr=0.0052$ ,  $df=2,16$ :  $\hat{\beta}_0=4.1$ ,  $\hat{\beta}_{pA}=-3.1$ ,  $t=-3.56$ ,  $Pr=0.0026$ ,  $\hat{\beta}_{pB}=-1.9$ ,  $t=-2.34$ ,  $Pr=0.0328$ ). Means of raw and standardized composite diallelic disequilibrium, however, did not differ significantly among regions ( $F=1.47$ ,  $Pr=0.2644$ ,  $F=0.12$ ,  $Pr=0.9736$ , respectively,  $df=4,14$ ).

Estimates of type II error probabilities for the observed disequilibria among unlinked loci in our sample are presented in Table 4. The type II error probabilities were quite high; the estimates, on average, were at least 15 times greater than the standard type I error probability of 0.05.

## Discussion

We observed statistically significant estimates of disequilibrium within all of the gene regions studied. Moreover, patterns of significant composite diallelic disequilibrium within regions indicate that the relative allele frequencies of pairs of loci are highly associated. These results suggest that particular marker allelic combinations that occur close together on chromosomes have a shared evolutionary history, reflected in diallelic disequilibrium that recombination has not dissipated. In the *ApoAI-CIII-AIV* region, the estimate of composite diallelic disequilibrium was not significant for the *PstI-SstI* marker pair although both markers were in significant disequilibrium with the *I/D* polymorphism 5' to *PstI*. A similar relationship was observed in the *LDLR* region. The estimate of composite diallelic disequilibrium was not significant for the *TaqI-StuI* marker pair although both markers were in significant disequilibrium with all other 3' markers in the region. These two marker pairs, *PstI-SstI* and *TaqI-StuI*, may each have had a longer shared evolutionary history than other marker pairs, such that there has been sufficient time for recombination events to randomize the relative allelic frequencies.

These patterns of disequilibrium in the *ApoAI-CIII-AIV* and *LDLR* regions also indicate that disequilibrium cannot be used to predict the physical distance relationships among loci in these two regions. A similar conclusion can be reached for the *ApoB* region. The magnitude of the standardized estimates of disequilibrium,  $\Delta'_{AB}$ , and the physical orderings of the markers are not correlated (Table 3). Magnitudes of standardized estimates for the *LPL* region were highest for the *HindIII-BstNI* marker pair closest together in the same intron and about the same for both of these marker pairings with the 5' marker (*PvuII*). A conclusion that disequilibrium might be related to physical distance between loci in the *LPL* gene region cannot be made with confidence from data on only three polymorphisms.

Our study adds to a substantial body of empirical evidence that disequilibrium cannot be used to infer physical distance relationships among loci in short DNA regions (but see also Jorde 1995b; see review by Jorde et al. 1994; Zapata and Visedo 1995). All of the regions in our study were less than 50 kb. Jorde (1994) argues that the variance of standardized disequilibrium measure,  $r$ , is too large in regions less than 50–100 kb for there to be a reliable disequilibrium–distance relationship. This argument assumes all polymorphisms are of equal age and recombination is the only force affecting the disequilibrium between loci. There cannot be a reliable disequilibrium–distance relationship in such regions for three additional reasons. First, although markers in a region share an evolutionary history, the unpredictable nature of mutation means that each polymorphism probably originated at a different time. Second, mutation rate may be of the same order of magnitude as that of recombination in these short DNA regions. Third, the rate of recombination is likely to vary along the chromosome resulting in heterogeneity in the decay of disequilibrium between markers.

Thompson and Neel (1997) suggest that there is no need to invoke selection as an explanation for the high degree of disequilibrium observed in closely linked regions. Instead, they argue that most polymorphisms are likely of recent origin and unique to what were once isolated tribal populations. The recent extensive expansion and amalgamation of most human populations is thus expected to have generated considerable disequilibrium among loci over regions as large as 0.5 cM. This could also explain the heterogeneity in allele frequencies we observed within and among regions. Moreover, although variability in composite diallelic disequilibrium was negatively associated with variability in relative allele frequencies, this relationship was not strong enough to result in statistically detectable differences in average estimates of composite diallelic disequilibrium among regions. This suggests that these recent population historical effects on disequilibrium are distributed about equally across the five regions of this study and probably across much of the genome.

It is well known that admixture among populations can generate disequilibrium even among loosely linked loci. The significant composite diallelic disequilibrium among regions that we observed also included higher-order triallelic

elic and quadriallelic disequilibria. Such higher-order disequilibria represent additional dependencies among alleles in the way they combine to form haplotypes and genotypes which may be caused by the same forces responsible for diallelic gametic disequilibrium. Assuming no further genetic contact, however, diallelic gametic disequilibrium among unlinked loci that is caused by admixture is expected to decay rapidly within two to four generations (Stephens et al. 1994). Current worldwide changes in the genetic structure of the human population due to population expansion and amalgamation, as described by Thompson and Neel (1997), and the high degree of mobility of many individuals among populations in modern society suggest that continuous admixture is more likely. Moreover, the population of Rochester, Minn. increased from 57,890 to 70,745 individuals (22.2% increase) between 1980 and 1990 alone (United States Bureau of the Census 1983, 1993). This population increase was during the time which the sample for this study was collected. This population increase is not an isolated event, but parallels the generally recognized worldwide changes in population growth and amalgamation that have occurred since the advent of modern agriculture. We conclude that recent population historical factors are the most likely explanation for the considerable disequilibrium among unlinked regions we observed and suggest that disequilibrium among unlinked loci may therefore be much more extensive than is commonly assumed.

Our findings suggest caution in using information about associations between marker loci and phenotypic variation for inferring the locations of functional mutations. The 16 loci in the five regions we studied represent only a tiny fraction of the possible set of loci in these regions that may also be in disequilibrium. Moreover, assuming functional mutations can be identified in known candidate susceptibility gene regions, estimating their impacts on quantitative trait variation in the population will be affected by other functional loci, within and among regions on the same chromosome as well as on different chromosomes, in disequilibrium with the functional mutation. The portion of genetic variance of a quantitative trait attributable to a particular functional mutation will be either enhanced or diminished depending on the signs of the disequilibria involved (Falconer and Mackay 1996).

One possible interpretation of the number of significant disequilibria among unlinked loci we observed is that it is not much different than might be expected by chance; it could be just a series of type I errors. Type II error probability estimates, however, were at least 15 times higher, on average, than the standard type I error probability of 0.05. We suggest that real disequilibrium is a much more likely interpretation since hypothesis tests about disequilibria are so conservative in general (low power, high type II error probability) that it is much more likely to miss real disequilibria and hence make a type II error using the standard  $\alpha$  level of 0.05 (Ward and Sing 1970; Zerba et al. 1991). Increasing the  $\alpha$  level to 0.10 increases the acceptable type I error rate by only 5%, in contrast to the average observed type II error probability of about 0.85 at the standard  $\alpha$  level. It is indeed

possible that we are picking up additional signals of real disequilibrium in this  $\alpha$  range just above 0.05.

It is commonly assumed that the contributions of variations in polygenes to variations in complex quantitative traits are independent and additive. Considering the potential cumulative impact of large numbers of unlinked loci in disequilibrium across the genome on estimates of the genetic variance of quantitative traits associated with known functional mutations, discounting the possibility of making type II errors and writing off these results solely due to chance could have much more serious consequences than allowing for a slightly higher possibility of making type I errors. We conclude that the patterns of disequilibria within and among the many genes that are involved in determining interindividual variation in susceptibility to a common disease having a complex multifactorial etiology make the characterization of the effects of a particular functional mutation one of the most difficult biomedical research challenges for the twenty-first century.

**Acknowledgements** We thank Dr. Andy Clark for critical comments that improved the paper and Ken G. Weiss for the graphics. Supported by National Institutes of Health grant HL39107. The experiments conducted in this study comply with the current laws of the United States of America.

## References

- Benlian P, Boileau C, Loux N, Pastier D, Maslah J, Coulon M, Nigou M, Ragab A, Guimard J, Ruidavets JB, Bona-Pellié C, Fruchart JC, Douste-Blazy P, Bereziat G, Junien C (1991) Extended haplotypes and linkage disequilibrium between 11 markers at the APOAI-C3-A4 gene cluster on chromosome 11. *Am J Hum Genet* 48:903–910
- Berg K, Powell LM, Wallis SC, Pease R, Knott TJ, Scott J (1986) Genetic linkage between the antigenic group (Ag) and the apolipoprotein B gene: assignment of the Ag locus. *Proc Natl Acad Sci USA* 83:7367–7370
- Bickeböller H, Margaritte-Jeannin P, Babron M-C, Clerget-Darpoux F (1995) Systematic search of susceptibility loci with methods using gametic disequilibrium. *Genet Epidemiol* 12:577–582
- Blackhart BD, Ludwig EM, Pierotti VR, Caiati L, Onasch MA, Wallis SC, Powell L, Pease R, Knott TJ, Chu M-L, Mahley RW, Scott J, McCarthy BJ, Levy-Wilson B (1986) Structure of the human apolipoprotein B gene. *J Biol Chem* 261:15364–15367
- Boerwinkle E, Chan L (1989) A three codon insertion/deletion polymorphism in the signal peptide region of the human apolipoprotein B (APOB) gene directly typed by the polymerase chain reaction. *Nucleic Acids Res* 17:4003
- Borresen AL, Moller P, Berg K (1988) Linkage disequilibrium analyses and restriction mapping of four RFLPs at the pro $\alpha$ 2(I) collagen locus: lack of correlation between linkage disequilibrium and physical distance. *Hum Genet* 78:216–221
- Briscoe D, Stephens JC, O'Brien SJ (1994) Linkage disequilibrium in admixed populations: applications in gene mapping. *J Hered* 85:59–63
- Brower V (1998) Mining the genetic riches of human populations. *Nat Biotechnol* 16:337–339
- Brown MS, Goldstein JL (1986) A receptor-mediated pathway for cholesterol homeostasis. *Science* 232:34–47
- Brown PO, Hartwell L (1998) Genomics and human disease variations – variations on variation. *Nat Genet* 18:91–93

- Clark AG, Weiss KM, Nickerson DA, Taylor SL, Buchanan A, Stengård J, Salomaa V, Vartiainen E, Perola M, Boerwinkle E, Sing CF (1998) Haplotype structure and population genetic inferences from nucleotide sequence variation in human lipoprotein lipase. *Am J Hum Genet* 63:595–612
- 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
- Davignon J, Gregg RE, Sing CF (1988) Apolipoprotein E polymorphism and atherosclerosis. *Arteriosclerosis* 8:1–21
- Deeb SS, Peng R (1989) Structure of the human lipoprotein lipase gene. *Biochemistry* 28:4131–4135
- Devlin B, Risch N (1995) A comparison of linkage disequilibrium measures for fine-scale mapping. *Genomics* 29:311–322
- Ewens WJ, Spielman RS (1995) The transmission/disequilibrium test: history, subdivision, and admixture. *Am J Hum Genet* 57:455–464
- Falconer DS, Mackay TFC (1996) Introduction to quantitative genetics, 4th edn. Longman Group, Essex
- Fielding CJ, Shore VG, Fielding PE (1972) A protein cofactor of lecithin:cholesterol acyltransferase. *Biochem Biophys Res Commun* 46:1493–1498
- Fisher KL, Fitzgerald GA, Lawn RM (1987) Two polymorphisms in the human lipoprotein lipase (LPL) gene. *Nucleic Acids Res* 15:7657
- Funke H, Reckwerth A, Stapenhorst D, Schluzer Beiering M, Jansen M, Assman G (1988) *Bst*NI (*Eco*RII) RFLP in the lipoprotein lipase gene (*LPL*). *Nucleic Acids Res* 16:2741
- Ghosh S (1995) Probability and complex disease genes. *Nat Genet* 9:223–224
- Goldstein JL, Brown MS (1989) Familial hypercholesterolemia. In: Scriver CR, Beaudet AL, Sly WS, Valle D (eds) *The metabolic basis of inherited disease*, 6th edn. McGraw-Hill, New York, pp 1215–1250
- Haviland MB, Kessling AM, Davignon J, Sing CF (1991) Estimation of Hardy-Weinberg and pairwise disequilibrium in the apo AI-CIII-AIV gene cluster. *Am J Hum Genet* 49:350–365
- Haviland MB, Kessling AM, Davignon J, Sing CF (1995) Cladistic analysis of the apolipoprotein AI-CIII-AIV gene cluster using a healthy French Canadian sample. I. Haploid analysis. *Ann Hum Genet* 59: 211–231
- Haviland MB, Ferrell RE, Sing CF (1997) Association between common alleles of the low-density lipoprotein receptor gene region and interindividual variation in plasma lipid and apolipoprotein levels in a population-based sample from Rochester, Minnesota. *Hum Genet* 99: 108–114
- Hedrick PW (1987) Gametic disequilibrium measures: proceed with caution. *Genetics* 117:331–341
- Hegele RA, Plaetke R, Lalouel J-M (1990) Linkage disequilibrium between DNA markers at the low-density lipoprotein receptor gene. *Genet Epidemiol* 7:69–81
- Heinzmann C, Ladas J, Antonarakis S, Kirchgesner T, Schotz M, Lusic AJ (1987) RFLP for the human lipoprotein lipase (LPL) gene: *Hind*III. *Nucleic Acids Res* 15:6763
- Hill WG, Weir BS (1994) Maximum-likelihood estimation of gene location by linkage disequilibrium. *Am J Hum Genet* 54:705–714
- Hobbs HH, Esser V, Russell DW (1987) *Ava*II polymorphism in the human LDL receptor gene. *Nucleic Acids Res* 15:379
- Hobbs HH, Russell DW, Brown MS, Goldstein JL (1990) The LDL receptor locus in familial hypercholesterolemia: mutational analysis of a membrane protein. *Annu Rev Genet* 24:133–170
- Jorde LB (1995a) Linkage disequilibrium as a gene-mapping tool. *Am J Hum Genet* 56:11–14
- Jorde LB (1995b) Reply to Zapata and Visado. *Am J Hum Genet* 57:191–193
- Jorde LB, Watkins WS, Carlson M, Groden J, Albertsen H, Thliveris A, Leppert M (1994) Linkage disequilibrium predicts physical distance in the adenomatous polyposis coli region [see comments]. *Am J Hum Genet* 54:884–898
- Kahn P (1996) Gene hunters close in on elusive prey. *Science* 271:1352–1354
- Kamboh MI, Ferrell RE, Kottke BA (1988) Genetic studies of human apolipoproteins. V. A novel rapid procedure to screen apolipoprotein E polymorphism. *J Lipid Res* 29:1535–1543
- Kaplan NL (1995) Are moment bounds on the recombination fraction between a marker and a disease locus too good to be true? Allelic association mapping revisited for simple genetic diseases in the Finnish population. *Am J Hum Genet* 57:1486–1498
- Kaplan NL, Hill WG, Weir BS (1995) Likelihood methods for locating disease genes in nonequilibrium populations. *Am J Hum Genet* 56:18–32
- Kaprio J, Ferrell RE, Kottke BA, Kamboh MI, Sing CF (1991) Effects of polymorphisms in apolipoproteins E, AIV, and H on quantitative traits related to risk for cardiovascular disease. *Arterioscler Thromb Vasc Biol* 11:1330–1348
- Karathanasis SK (1985) Apolipoprotein multigene family: tandem organization of human apolipoprotein AI, CIII, and AIV genes. *Proc Natl Acad Sci USA* 82:6374–6378
- 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
- Kirchgesner T, Chuat J-C, Heinzmann C, Etienne J, Guilhot S, Svensson K, Ameis D, Pilon C, d'Auriol L, Andalibi A, Schotz MC, Galibert F, Lusic AJ (1989) Organization of the human lipoprotein lipase gene and evolution of the lipase gene family. *Proc Natl Acad Sci USA* 86:9647–9651
- Kotze MJ, Retief AE, Brink PA, Weich HF (1986) A DNA polymorphism in the human low-density lipoprotein receptor gene. *S Afr Med J* 70:77–79
- Kruglyak L (1997) What is significant in whole-genome linkage disequilibrium studies? *Am J Hum Genet* 61:810–812
- Leighton JK, Joyner J, Zamarripa J, Deines M, Davis RA (1990) Fasting decreases apolipoprotein B mRNA editing and the secretion of small molecular weight apoB by rat hepatocytes: evidence that the total amount of apoB secreted is regulated post-transcriptionally. *J Lipid Res* 31:1663–1668
- Leitersdorf E, Hobbs HH (1988) Human LDL receptor gene: *Hinc*II polymorphism detected by gene amplification. *Nucleic Acids Res* 16:7215
- Lewontin RC (1988) On measures of gametic disequilibrium. *Genetics* 120:849–852
- Li S, Oka K, Galton D, Stocks J (1988) *Pvu*II RFLP at the human lipoprotein lipase (LPL) gene locus. *Nucleic Acids Res* 16:2358
- Mahley RW (1988) Apolipoprotein E: cholesterol transport protein with expanding role in cell biology. *Science* 240:622–630
- Moll PP, Michels VV, Weidman WH, Kottke BA (1989) Genetic determination of plasma apolipoprotein AI in a population-based sample. *Am J Hum Genet* 44:124–139
- Murray JC, Buetow KH, Weber JL, Ludwigsen S, Scherpbier-Heddema T, Manion F, Quillen J, Sheffield VC, Sunden S, Duyk GM, Weissenbach J, Gyapay G, Dib C, Morrissette J, Lathrop JM, Vignal A, White R, Matsunami N, Geken S, Melis R, Albertsen H, Plaetke R, Odelberg S, Ward D, Dausset J, Cohen D, Cann H (1994) A comprehensive human linkage map with centimorgan density. *Science* 265:2049–2054
- Nickerson DA, Taylor SL, Weiss KM, Clark AG, Hutchinson RG, Stengård J, Salomaa V, Vartiainen E, Boerwinkle E, Sing CF (1998) DNA sequence diversity in a 9.7 kb region of the human lipoprotein lipase gene. *Nat Genet* 19:233–240
- Oka K, Tkalcovic G, Nakano T, Tucker H, Ishimura-Oka K, Virgil W (1991) Structure and polymorphic map of human lipoprotein lipase gene. *Biochim Biophys Acta* 1090:357
- Priestley L, Knott T, Wallis S, Powell L, Pease R, Scott J (1985) RFLP for the human apolipoprotein B gene. IV *Msp*I. *Nucleic Acids Res* 13:6792
- Rall SC Jr, Weisgraber KH, Mahley RW (1982) Human apolipoprotein E: the complete amino acid sequence. *J Biol Chem* 257:4171–4178
- Rees A, Shoulders CC, Stocks J, Galton DJ, Baralle FE (1983) DNA polymorphisms adjacent to human apoprotein A-1 gene: relation to hypertriglyceridaemia. *Lancet* 1:444–446



- Reilly SL, Ferrell RE, Kottke BA, Sing CF (1992) The gender specific apolipoprotein E genotype influence on the distribution of lipids and apolipoproteins in the population of Rochester, Minnesota. II. Regression relationships with concomitants. *Am J Hum Genet* 51:1311–1324
- Risch N, Merikangas K (1996) The future of genetic studies of complex human diseases. *Science* 273:1516–1517
- Schaefer EJ, Eisenberg S, Levy RI (1978) Lipoprotein apoprotein metabolism. *J Lipid Res* 19:667–687
- Shoulders CC, Myant NB, Sidoli A, Rodriguez JC, Cortese C, Baralle FE, Cortese R (1985) Molecular cloning of human LDL apolipoprotein B cDNA. *Atherosclerosis* 58:277–289
- Sing CF, Davignon J (1985) Role of the apolipoprotein E polymorphism in determining normal plasma lipid and lipoprotein variation. *Am J Hum Genet* 37:268–285
- Sing CF, Haviland MB, Templeton AR, Zerba KE, Reilly SL (1992a) Biological complexity and strategies for finding DNA variations responsible for inter-individual variation in risk of a common chronic disease, coronary artery disease. *Ann Med* 24:539–547
- Sing CF, Haviland MB, Zerba KE, Templeton AR (1992b) Application of cladistics to the analysis of genotype–phenotype relationships. *Eur J Epidemiol* 8 [Suppl 1]:3–9
- Sinnock P, Sing CF (1972) Analysis of multilocus genetic systems in Tecumseh, Michigan. II. Consideration of the correlation between non-alleles in gametes. *Am J Hum Genet* 24:393–415
- Stephens JC, Briscoe D, O'Brien SJ (1994) Mapping by admixture linkage disequilibrium in human populations: limits and guidelines. *Am J Hum Genet* 55:809–824
- Südhof TC, Goldstein JL, Brown MS, Russell DW (1985) The LDL receptor gene: a mosaic of exons shared with different proteins. *Science* 228:815–822
- Tanksley SD (1993) Mapping polygenes. *Annu Rev Genet* 27:205–233
- Templeton A (1995) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping or DNA sequencing. V. Analysis of case/control sampling designs: Alzheimer's disease and the apoprotein E locus. *Genetics* 140:403–409
- Templeton AR (1996) Cladistic approaches to identifying determinants of variability in multifactorial phenotypes and the evolutionary significance of variation in the human genome. In: Chadwick D, Cardew G (eds) *Variation in the human genome*. Ciba Foundation Symposium 197. Wiley, Chichester, pp 259–283
- Templeton AR, Sing CF (1993) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. IV. Nested analyses under cladogram uncertainty and recombination. *Genetics* 134:659–669
- 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 phenotype associations with haplotypes inferred from restriction endonuclease mapping. II. The analysis of natural populations. *Genetics* 120:1145–1154
- Templeton AR, Crandall KA, Sing CF (1992) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics* 132:619–633
- Terwilliger JD (1995) A powerful likelihood method for the analysis of linkage disequilibrium between trait loci and one or more polymorphic loci. *Am J Hum Genet* 56:777–787
- Thompson EA, Neel JV (1997) Allelic disequilibrium as a function of social and demographic history. *Am J Hum Genet* 60:197–204
- 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
- Turner ST, Weidman WH, Michels VV, Reed TJ, Ormson CL, Fuller T, Sing CF (1989) Distribution of sodium–lithium countertransport and blood pressure in Caucasians five to eighty-nine years of age. *Hypertension* 13:378–391
- United States Bureau of the Census (1983) 1980 census of population and housing. Census tracts. Rochester, Minn. US Department of Commerce, Bureau of the Census, Washington, DC
- United States Bureau of the Census (1993) 1990 census of population and housing. Population and housing characteristics for census tracts and block numbering areas. Rochester, Minn. US Department of Commerce Economics and Statistics, Administration Bureau of the Census, Washington, DC
- Ward RH, Sing CF (1970) A consideration of the power of the  $\chi^2$  test to detect inbreeding effects in natural populations. *Am Nat* 104:355–365
- Weeks DE, Lathrop M (1995) Polygenic disease: methods for mapping complex disease traits. *Trends Genet* 11:513–519
- Weir BS (1996) *Genetic data analysis II: methods for discrete population genetic data*, 2nd edn. Sinauer Associates, Sunderland
- Weir BS, Brooks LD (1986) Disequilibrium on human chromosome 11p. *Genet Epidemiol Suppl* 1:177–183
- Weir BS, Cockerham CC (1989) Complete characterization of disequilibrium at two loci. In: Feldman MW (ed) *Mathematical evolutionary theory*. Princeton University Press, Princeton, pp 86–110
- Weissman SM (1995) Genetic bases for common polygenic diseases. *Proc Natl Acad Sci USA* 92:8543–8544
- Wijsman EM (1987) A deductive method of haplotype analysis in pedigrees. *Am J Hum Genet* 41:356–373
- Xhignesse M, Lussier-Cacan S, Sing CF, Kessling AM, Davignon J (1991) Influences of common variants of apolipoprotein E on measures of lipid metabolism in a sample selected for health. *Arterioscler Thromb Vasc Biol* 11:1100–1110
- Yamakawa K, Okafuji T, Iwamura Y, Russell DW, Hamaguchi H (1987) *TaqI* polymorphism in the human LDL receptor gene. *Nucleic Acids Res* 15:7659
- Zapata C, Visedo G (1995) Gametic disequilibrium and physical distance. *Am J Hum Genet* 57:190–191
- Zerba KE, Kessling AM, Davignon J, Sing CF (1991) Genetic structure and the search for genotype–phenotype relationships: an example from disequilibrium in the Apo B gene region. *Genetics* 129:525–33