Linkage Between Cholesterol 7α-Hydroxylase and High Plasma Low-density Lipoprotein Cholesterol Concentrations

Jinping Wang,* Dilys J. Freeman,* Scott M. Grundy,* Daniel M. Levine,§ Rudy Guerra,‡ and Jonathan C. Cohen*

**The Center for Human Nutrition, Department of Clinical Nutrition and Department of Internal Medicine, Dallas, Texas 75235-9052;* ‡*The Department of Statistical Science, Southern Methodist University, Dallas, Texas 75275-0332; and* §*The Rogosin Institute, New York, New York 10021*

Abstract

Interindividual differences in plasma low-density lipoprotein cholesterol (LDL-C) levels reflect both environmental variation and genetic polymorphism, but the specific genes involved and their relative contributions to the variance in LDL-C are not known. In this study we investigated the relationship between plasma LDL-C concentrations and three genes with pivotal roles in LDL metabolism: the low-density lipoprotein receptor (*LDLR*), apolipoprotein B (*APOB*), and cholesterol 7α-hydroxylase (*CYP7*). Analysis of 150 nuclear families indicated statistically significant linkage between plasma LDL-C concentrations and *CYP7*, but not *LDLR* or *APOB*. Further sibling pair analyses using individuals with high plasma LDL-C concentrations as probands indicated that the *CYP7* locus was linked to high plasma LDL-C, but not to low plasma LDL-C concentrations. This finding was replicated in an independent sample. DNA sequencing revealed two linked polymorphisms in the 5′ flanking region of *CYP7*. The allele defined by these polymorphisms was associated with increased plasma LDL-C concentrations, both in sibling pairs and in unrelated individuals. Taken together, these findings indicate that polymorphism in *CYP7* contributes to heritable variation in plasma LDL-C concentrations. Common polymorphisms in *LDLR* and *APOB* account for little of the heritable variation in plasma LDL-C concentrations in the general population. (*J. Clin. Invest.* 1998. 101:1283–1291.) Key words: cholesterol 7 alpha-monooxygenase • polymorphism (genetics) • bile acids

Introduction

Elevated plasma concentrations of low-density lipoprotein cholesterol (LDL-C)[1] are a major risk factor for coronary heart disease (1). Accordingly, considerable research efforts have been focused on the factors that determine plasma LDL-C concentrations (2). Data from family and twin studies have indicated that genetic factors account for ~ 50% of the interindividual variation in plasma LDL-C concentrations (3–6). The molecular mechanisms responsible for genetic variation in plasma LDL-C concentrations have been most clearly elucidated in Mendelian disorders of LDL metabolism. In familial hypercholesterolemia (7), mutations in the gene encoding the LDL receptor (*LDLR*) cause markedly elevated plasma LDL-C concentrations. A similar phenotype is observed in familial defective apoB, caused by a G→A mutation in codon 3500 of the gene encoding apolipoprotein B (*APOB*; 8). Several other mutations in *APOB* cause familial hypocholesterolemia in which affected individuals have very low plasma LDL-C concentrations (9). These syndromes, however, are relatively rare in most populations, and therefore account for a minor fraction of the heritable variation in plasma LDL-C concentrations.

Much less is known about the causes of heritable variation in plasma LDL-C concentrations in the general population. In most families, segregation of plasma LDL-C concentrations is not consistent with simple Mendelian inheritance; therefore, heritable variation in plasma LDL-C concentrations presumably reflects common sequence polymorphisms in genes that modulate LDL metabolism. Associations between plasma LDL-C concentrations and common polymorphisms in several genes have been reported, but these effects have generally been small, and have not proven to be consistent (10–13). To date the only common polymorphism that has consistently been shown to be associated with heritable variation in plasma LDL-C concentrations lies in the gene encoding apolipoprotein E (*APOE*), a ligand for receptor-mediated lipoprotein uptake (14). Three common alleles of *APOE* (ε2, ε3, and ε4) have been systematically associated with different plasma LDL-C concentrations in several populations (14), accounting for ~ 5–10% of the population variance in plasma LDL-C concentrations (15). Polymorphism in *APOE* therefore explains a small fraction (10–20%) of the heritable variation in plasma LDL-C concentrations. The primary cause(s) of heritable variation in plasma LDL-C concentrations remains to be determined.

Our laboratory has initiated a systematic study of several candidate genes that may affect LDL-C concentrations. By using allele-sharing methods in nuclear families, the contribution of these genes to heritable variation in plasma LDL-C concentrations can be determined. In this report we focus on *LDLR*, *APOB*, and the gene encoding cholesterol 7α-hydroxylase (*CYP7*), the rate-limiting enzyme in bile acid biosynthesis. Each of these genes has a pivotal role in LDL-C metabolism, and genetic (7–9) or pharmacological (16–20) alterations in their expression leads to marked alteration in plasma LDL-C concentrations. To assess the sensitivity of the analysis, *APOE* was used as a positive control.

Address correspondence to Jonathan C. Cohen, Ph.D., Center for Human Nutrition, UT Southwestern Medical Center, 5323 Harry Hines Blvd., Dallas, TX 75235-9052. Phone: 214-648-4774; FAX: 214-648-4837; E-mail: cohen@crcdec.swmed.edu

Received for publication 29 July 1997 and accepted in revised form 8 January 1998.

1. *Abbreviation used in this paper:* LDL-C, low-density lipoprotein cholesterol.

J. Clin. Invest.
© The American Society for Clinical Investigation, Inc.
0021-9738/98/03/1283/09 $2.00
Volume 101, Number 6, March 1998, 1283–1291
http://www.jci.org

Methods

Subjects

The study protocol was approved by the Institutional Review Board of the University of Texas Southwestern Medical Center at Dallas.

For heritability estimation and initial linkage analysis, 934 individuals from 150 nuclear families in which both parents and three or more children were available for sampling were recruited by advertisements in newspapers, churches, and health centers (Group 1). These families were ascertained solely on the basis of family size with no a priori knowledge of plasma LDL-C concentrations. None of these individuals had familial hypercholesterolemia or familial hypobetalipoproteinemia. Blood samples were drawn from all available family members, but data from 102 individuals (68 parents and 34 offspring) who were diabetic or postmenopausal, or who took lipid-lowering medication or hormone replacement were excluded a priori from the analysis. In addition, since the Friedewald equation provides a less reliable estimate of LDL-C in hypertriglyceridemic sera (22), 22 individuals (15 parents and 7 offspring) who had plasma triglyceride concentrations $>$ 300 mg/dl were also excluded.

Corroborative linkage analyses were performed in a second group (Group 2) comprising 347 individuals from 54 nuclear families ascertained through individuals who had symptomatic, premature coronary artery disease. Association studies were performed in a third group (Group 3) comprising unrelated men and women recruited at health fairs at the Dallas Fire Department and three hospital-based fitness centers. None of these individuals used lipid-lowering medication, and none of the women were postmenopausal.

All individuals in the study identified themselves as white, and not of Hispanic origin. Fasting blood samples were drawn into 10-ml vacuum tubes containing sodium EDTA. Plasma was separated by centrifugation and stored at 4°C until analysis. Genomic DNA was isolated from the white blood cell pellet using an automated DNA extractor (Applied Biosystems, Inc., Foster City, CA). Informed consent was obtained from each study participant, and ascertainment was essentially complete ($>$ 95%) for children over the age of 5 yr. All participants completed a detailed questionnaire to furnish data on family history, past and current health status, diet and exercise habits, alcohol and tobacco use, and medications.

Plasma lipid, lipoprotein, and apolipoprotein B assay

Plasma cholesterol and triglyceride concentrations were determined in duplicate by enzymatic assay using commercial reagents (Cholesterol/HP; Boehringer Mannheim Biochemicals, Indianapolis IN; and Triglycerides GPO-TRINDER; Sigma Chemical Co., St. Louis, MO). High-density lipoprotein cholesterol (HDL-C) was measured in the supernatant after precipitation of apo B–containing lipoproteins with sodium phosphotungstate (21). Samples from all individuals in a family were analyzed in duplicate in the same assay. Intraassay variation was $<$ 3% for plasma cholesterol and triglyceride, and $<$ 5% for plasma HDL-C. Plasma LDL-C concentrations were calculated according to the formula of Friedewald et al. (22). Calculated plasma LDL-C concentrations were strongly correlated (r = 0.88) with plasma concentrations of apolipoprotein B determined by immunoturbidimetric assay (23, 24).

Segregation of candidate genes

In each nuclear family, segregation of the parental alleles of each candidate gene was determined using microsatellite polymorphisms that cosegregated with the gene. For each parental allele, a haplotype was constructed using at least two informative polymorphisms located within 5 centimorgans of the gene. To preclude misspecification of alleles due to recombination between the marker locus and the gene, each haplotype included one marker within the gene, or one marker on each side of the gene. In cases where recombination could not be excluded, the individual data were excluded from the analysis. The microsatellite polymorphisms D19S394, D19S221, D19S411, and D19S226 (for *LDLR*); D2S220, D2S224, D2S1360, and APOB (for *APOB*); D8S1102, D8S1113, D8S507, and D8S285 (for *CYP7*); and D19S178, D19S197, and DM (for *APOE*) were identified using the Genome Database (25). The location of these microsatellite polymorphisms with respect to the candidate genes was confirmed in 20 large nuclear families recruited by our laboratory (data not shown).

DNA sequencing

The coding region, intron/exon boundaries, and the 750 bp upstream of the translation initiation codon of *CYP7* were sequenced using a standard protocol for cycle sequencing (26). DNA regions to be sequenced were PCR-amplified and purified using spin columns to remove excess PCR primers and buffer. Sequencing reactions were performed in 20-μl volumes containing 8 μl terminator mix, 100–200 ng DNA template, 3.2 pmol primer, and 8 μl water. Reactions were overlaid with 50 μl mineral oil, and subjected to 25 amplification cycles. Extension products were purified using Centri-Sep (Princeton Separations, Adelphia, NJ) spin columns, resuspended in loading buffer (containing deionized formamide, 25 mM EDTA, and 10 mg/liter blue dextran), heated to 90°C, and loaded onto an ABI 377 automated sequencer.

Assay of CYP7 polymorphism

A C→A substitution 278 bp upstream of the translation initiation codon destroyed a Bsa I restriction site (5′ GGTCTC 3′). To assay the polymorphism, a 968-bp fragment containing the restriction site was PCR-amplified using the primers 5′ TGGTAGGTAAATTATTAATAGATGT 3′ and 5′ AAATTAAATGGATGAATCAAAGAGC 3′. The fragment was labeled by adding 0.3 pmol [^{32}P]dCTP to the PCR reaction mixture. The PCR-amplified DNA fragment was digested by adding 10 U of Bsa I in 30 μl NEB buffer 4 (New England Biolabs Inc., Beverly, MA) to the PCR reaction, and was electrophoresed on a 5% polyacrylamide gel.

Statistical analysis

Heritability estimation. The heritability index of plasma LDL-C levels was estimated by regressing the average of the offspring plasma LDL-C values on the midparent values using weighted least squares. Weights to adjust for unequal family sizes were calculated as suggested by Falconer (27), with an unweighted least squares estimate used as the initial estimate of heritability in the weights.

Sibling pair linkage analysis. Initial testing for linkage was performed using the sibling-pair method developed by Haseman and Elston (28) as described in detail previously (29). In this method, the squared sibling pair difference of plasma LDL-C concentrations (Δ) is regressed on the proportion (π) of alleles shared by descent at the candidate locus. Evidence for linkage is provided by a statistically significant negative slope. Since our data do not meet the usual assumptions of the t test for the regression coefficient, corroborative testing was performed using a nonparametric Monte Carlo simulation test that mimics the probability mechanism of allele segregation under no linkage. By randomly assigning arbitrarily labeled parental alleles to each individual, the null distribution of the slope estimate can be determined. We replicated this procedure 20,000 times. The P value was calculated as the percentage of times the slope estimate based on the randomly assigned alleles exceeded the slope calculated from the observed data. All possible sibling pairs were used, and were treated as independent observations.

Variance component linkage analysis. To test for linkage under a more complete genetic model that accommodates covariates and nuclear families, we used a robust variance components procedure proposed by Hopper and Mathews (30) and developed by Amos (31). Simulation studies indicate that this method offers greater power to detect linkage than does the Haseman-Elston method (32). The null hypothesis of no linkage was evaluated using a likelihood ratio test that compares a model that allows for genetic correlation due to *CYP7* with one that does not allow for this correlation. To provide corroborative testing to the standard χ^2 test, we performed a Monte Carlo simulation test (as described above) using the likelihood ratio test statistic.

Sibling pair correlations. The intraclass correlation between the plasma LDL-C concentrations of siblings was calculated by ANOVA. To determine whether plasma LDL-C concentrations were more highly correlated among siblings sharing both alleles of a candidate gene than among siblings sharing no alleles of the gene, a Monte Carlo simulation test was performed. Parental alleles were randomly assigned to each individual, and the difference of the intraclass correlation among sibling pairs sharing 2 and 0 alleles ($r_2 - r_0$) was calculated. The procedure was repeated 10,000 times, and the *P* value was calculated as the percentage of times the Monte Carlo differences exceeded the observed difference.

Analysis of selected samples. To determine whether polymorphism in *CYP7* was specifically associated with high plasma LDL-C concentrations, linkage studies were performed in sibships that included at least one individual (proband) with high plasma LDL-C concentrations. If polymorphism in *CYP7* is associated with high plasma LDL-C concentrations, then plasma LDL-C concentrations should be higher in siblings sharing both *CYP7* alleles in common with the proband than in siblings who share neither of the proband's *CYP7* alleles. Sibling pair differences (proband − sibling) for plasma LDL-C concentrations were calculated for all possible pairs in which one member of the pair had plasma LDL-C concentrations exceeding the 90th percentile for age and sex. The mean difference in LDL-C concentrations of siblings sharing both *CYP7* alleles in common was compared with the mean value of siblings sharing no *CYP7* alleles in common using a one-sided, unpaired *t* test. The corresponding median values were compared using Wilcoxon's rank test. These tests were repeated using the 80th percentile as the threshold level for probands. Sibships containing more than one potential proband were analyzed using (*a*) all possible probands, or (*b*) one proband selected at random from among eligible siblings. To test for linkage between CYP7 and low plasma LDL-C concentrations, the analysis was performed using the 10th and 20th percentiles as thresholds for proband selection.

Association studies. The association between a polymorphism 278 bp upstream of the translation initiation codon in *CYP7* and plasma LDL-C concentrations was assessed in two ways.

First, plasma LDL-C concentrations were compared in siblings that were fully discordant for the −278 allele. This procedure allowed us to compare CC and AA genotypes within families. Plasma LDL-C concentrations were ranked within families, and the ranking scores were analyzed across families. Let $S_{i1} < ... < S_{ini}$ denote the n_i AA ranks from the *i*th family, and $W_A^{(i)} = S_{i1} + ... + S_{ini}$ the Wilcoxon rank-sum statistic computed for the *i*th family. For unequal family sizes, the test statistic

$$W_A = \sum [\, W_A^{(i)} / (N_i + 1) \,]$$

(where N_i is the number of siblings in the *i*th family) is appropriate for testing the significance of the effects of the polymorphism on plasma LDL-C concentrations.

Second, plasma LDL-C concentrations of unrelated individuals (Group 3) with the AA genotype were compared with those of individuals with the CC genotype using Wilcoxon's rank test. Men and women were compared separately.

Results

The genetic analyses in this study were performed in three sequential steps comprising linkage analysis in both unselected and selected samples, DNA sequencing, and association studies both within families and in unrelated individuals. Overall heritability estimation and initial linkage studies were performed in subjects from Group 1. The distribution of plasma LDL-C concentrations by age and sex in this sample were similar to those observed in the National Health and Nutrition Education Survey III (data not shown), a nationally representative sample of civilian noninstitutionalized US residents (1).

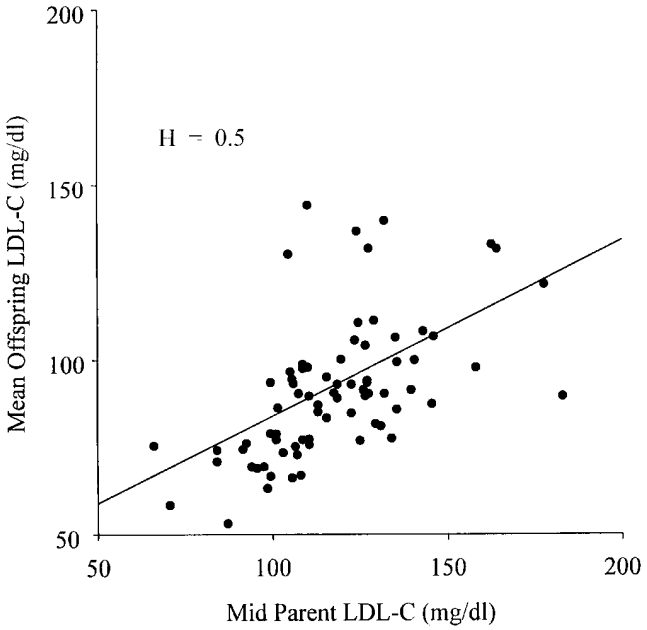

Figure 1. Scatterplot of plasma LDL-C concentrations in parents and their offspring. All families were ascertained solely on the basis of family size (Group 1). *H* is the slope of the regression line.

Heritability estimation. To determine the overall contribution of genetic factors to variation in plasma LDL-C concentrations, a heritability estimate was derived using data from 77 families from Group 1 in which both parents and two or more

*Table I. Linkage Between Plasma LDL-C Concentrations and LDLR, APOB, CYP7, and APOE**

Haseman-Elston sibling pair analysis‡

	All data		Restricted data	
Locus	*t* test	Monte Carlo test	*t* test	Monte Carlo test
APOE	0.003	0.004	0.003	0.010
LDLR	0.150	0.150	0.200	0.230
APOB	0.220	0.220	0.180	0.190
CYP7	0.060	0.060	0.020	0.050

Variance components analysis§

Locus	Likelihood ratio	Monte Carlo test
APOE	0.020	0.045
LDLR	0.140	0.270
APOB	0.460	0.880
CYP7	0.010	0.030

*Table entries are *P* values for linkage. All analyses were performed using data from 150 nuclear families who were selected solely on the basis of family size (Group 1). ‡Haseman-Elston sibling pair linkage analysis was performed using all possible pairs ($n = 1035$) from 599 siblings. To protect against the effects of outliers in the data, a second analysis was performed on a restricted data set. Points exceeding the 97.5 percentile at each π were excluded. §Variance components analysis was performed using individuals in Group 1.

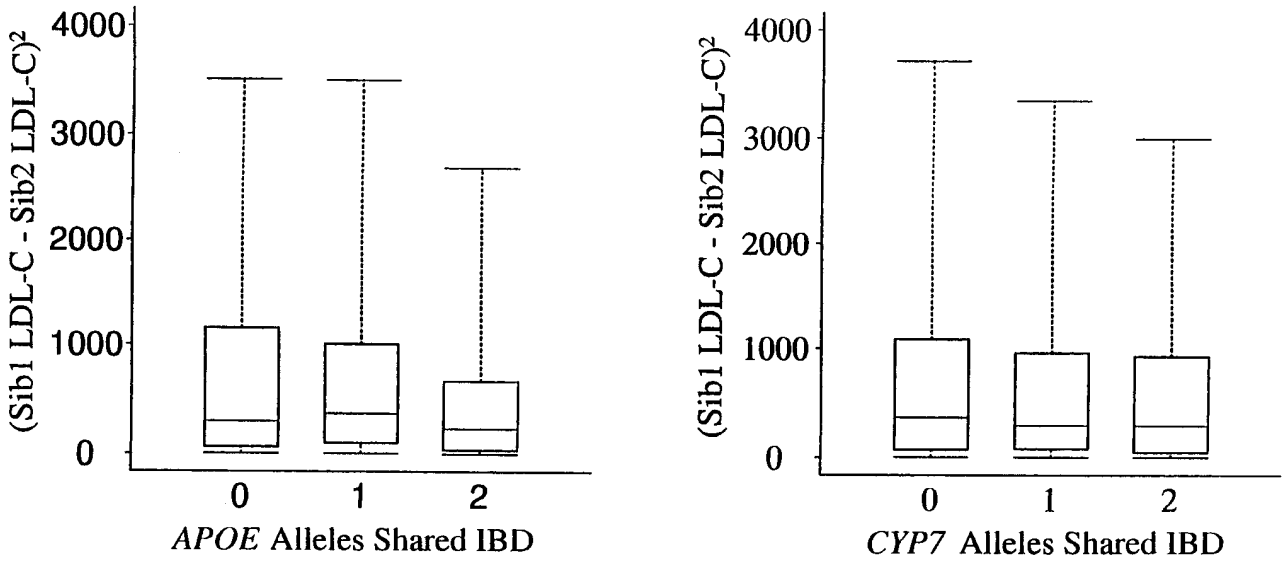

Figure 2. Sibling pair analysis of linkage between plasma LDL-C concentrations and *CYP7* and *APOE*. Data are squared sibling pair differences of observed plasma LDL-C levels. Subjects were unselected with respect to plasma LDL-C concentrations, and all possible sibling pairs (n = 1035) from 599 individuals in 150 nuclear families were included in the analysis. Median values are indicated by *bars* within boxes. Upper and lower quartiles are indicated by the *tops* and *bottoms of the boxes*, respectively. Whiskers indicate the 5th and 95th percentiles of the data.

children met the criteria for inclusion. Regression of mean offspring plasma LDL-C concentrations on the midparent value indicated that 50% of the variation in plasma LDL-C concentrations was heritable (Fig. 1). The heritability estimate was not appreciably different (53%) after z score adjustment of plasma LDL-C concentrations for age and sex. Spousal plasma LDL-C concentrations were not significantly correlated (r = 0.02, P = 0.87).

Sibling pair linkage analysis. Sibling pair linkage analyses were performed to determine whether polymorphism in *LDLR*, *APOB*, or *CYP7* contributes to the heritable variation in plasma LDL-C concentrations in this population. Since polymorphism in *APOE* is known to be associated with variation in plasma LDL-C concentrations, this locus was used as a positive control for the linkage studies. The Haseman-Elston sibling pair linkage procedure was used to test for linkage between each candidate gene and plasma LDL-C concentrations in 599 siblings from 150 families in Group 1. Analyses were performed on all possible pairs using both parametric and nonparametric tests. To protect against the influence of outlying

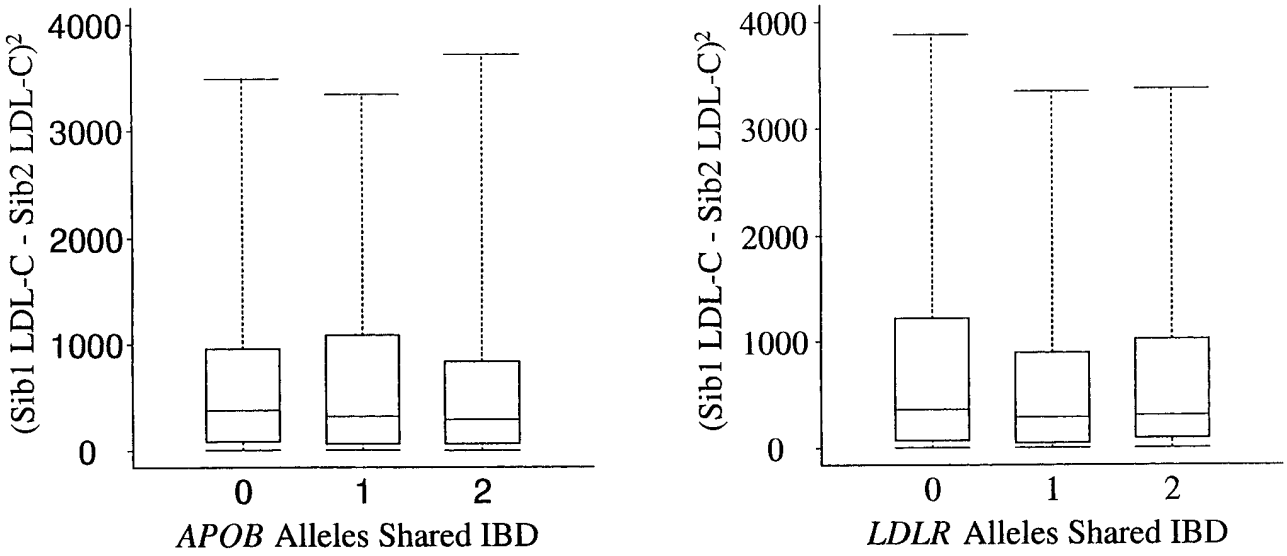

Figure 3. Sibling pair analysis of linkage between plasma LDL-C concentrations and *LDLR* and *APOB*. Data are squared sibling pair differences of observed plasma LDL-C levels. Subjects were unselected with respect to plasma LDL-C concentrations, and all possible pairs (n = 1035) from 599 individuals in 150 nuclear families were included in the analysis. Median values are indicated by *bars within boxes.* Upper and lower quartiles are indicated by the *tops* and *bottoms of the boxes*, respectively; whiskers indicate the 5th and 95th percentiles of the data.

Table II. Intraclass Correlations Between Sibling Pairs

Gene	n_0	r_0	n_2	r_2	*P* value for $r_0 - r_2$
APOE	229	0.46	242	0.60	0.03
LDLR	215	0.45	240	0.48	0.38
APOB	215	0.42	240	0.48	0.20
CYP7	263	0.35	267	0.53	0.01

Data are from families selected solely on the basis of family size (Group 1). n_0 is the number of sibling pairs sharing no alleles at the candidate locus. n_2 is the number of sibling pairs sharing both alleles at the candidate locus. r_0 is the intraclass correlation between siblings sharing no alleles at the candidate locus. r_2 is the intraclass correlation between siblings sharing both alleles at the candidate locus. *P* values were calculated using Monte Carlo simulation.

data points, analyses were repeated after excluding all points beyond the 97.5th percentile at each π. Strong evidence for linkage between *APOE* and plasma LDL-C was obtained under all conditions tested (Table I, Fig. 2). In contrast, no evidence for linkage was obtained between plasma LDL-C and *APOB* or *LDLR* (Table I, Fig. 3). For *CYP7*, the *P* values obtained were below, or very close to the nominal threshold for significance testing (Table I, Fig. 2).

Variance component linkage analysis. To corroborate the results of the sibling pair analyses and to estimate the proportion of variation in plasma LDL-C concentrations that could be accounted for by each candidate gene, a second series of linkage analyses was performed using the robust variance components procedure developed by Amos (31). This procedure indicated statistically significant linkage between plasma LDL-C concentrations and the *APOE* and *CYP7* loci, but not to the *LDLR* or *APOB* loci (Table I). Allelic variation in *APOE* and *CYP7* accounted for 11 and 15%, respectively, of the total variation in plasma LDL-C concentrations.

Sibling pair correlations. Plasma LDL-C concentrations were significantly more highly correlated in siblings sharing both *APOE* alleles than in siblings sharing no *APOE* alleles in common (Table II). Similarly, plasma LDL-C concentrations were more highly correlated among siblings sharing both *CYP7* alleles in common, than among siblings sharing no *CYP7* alleles (Table II). Sibling pair correlations were not significantly different among siblings sharing both, or no *LDLR* or *APOB* alleles in common (Table II).

Analysis of selected samples. Analysis of specific subgroups of sibling pairs was used to determine whether *CYP7* polymorphism was associated with increased or decreased plasma LDL-C concentrations. In Group 1, analysis of sibships including probands with high plasma LDL-C concentrations indicated statistically significant linkage between *CYP7* and elevated plasma LDL-C concentrations (Table III). Mean plasma LDL-C concentrations were higher in siblings sharing both *CYP7* alleles than in siblings sharing no *CYP7* alleles (Fig. 4). No relationship was observed between *CYP7* and low plasma LDL-C concentrations. Analysis of Group 2 individuals also indicated statistically significant linkage between *CYP7* and high plasma LDL-C concentrations (Table III).

DNA sequencing. To identify polymorphisms in *CYP7*, the coding region, intron/exon boundaries, and the 750 bp upstream of the translation initiation codon of *CYP7* were sequenced in 10 unrelated individuals with plasma LDL-C concentrations exceeding the 90th percentile for age and sex who

*Table III. Linkage between CYP7 and Plasma LDL-C Concentrations in Siblings Selected for High or Low Plasma LDL-C Concentrations**

Proband threshold‡	n_0	$P - S_0$	n_2	$P - S_2$	*t* test	Wilcoxon
		mg/dl		*mg/dl*	*P*	*P*
Group 1						
90th percentile	36	53	53	30	0.001	0.001
80th percentile	96	33	110	22	0.014	0.008
20th percentile	94	−23	102	−20	0.200	0.330
10th percentile	43	−25	59	−22	0.280	0.380
Group 2						
90th percentile	7	88	17	44	0.001	0.010
80th percentile	19	62	32	36	0.041	0.042
20th percentile	24	−42	20	−43	0.530	0.620
10th percentile	14	−44	9	−53	0.700	0.850

*Data in the table include all possible probands within each sibship. Similar results were obtained using a single proband selected at random within each sibship. ‡LDL-C percentile for age and sex. n_0 is the number of sibling pairs sharing no *CYP7* alleles. n_2 is the number of sibling pairs sharing both *CYP7* alleles with the proband. $P - S_0$ is the mean difference in age- and sex-adjusted plasma LDL-C concentrations between the proband and siblings sharing no *CYP7* alleles in common with the proband. $P - S_2$ is the mean difference in age- and sex-adjusted plasma LDL-C concentrations between the proband and siblings sharing both *CYP7* alleles in common with the proband.

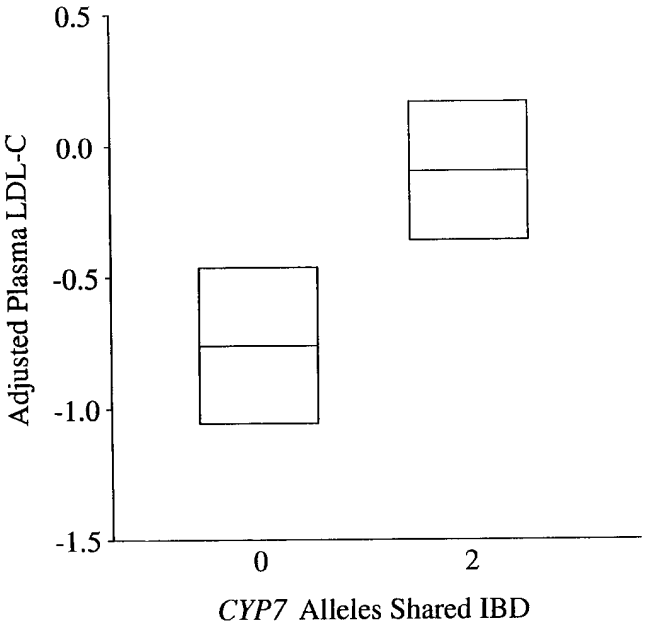

Figure 4. Z score–adjusted plasma LDL-C concentrations (mean± standard deviation) among siblings sharing no or two *CYP7* alleles with probands. Probands had plasma LDL-C concentrations above the 90th percentile for age and sex. *Bars in boxes* are means.

Table IV. Association Between CYP7 Genotype and Plasma Lipoprotein Concentrations

	CYP7 Genotype		
	AA	AC	CC
Men			
n	49	59	24
Age	43±6	44±7	44±8
Total cholesterol	191±27	196±33	203±29
Triglyceride	116±48	113±56	136±61
HDL-C*	36±7	37±9	39±9
LDL-C*	128±23	136±31	140±30
Women			
n	56	74	33
Age	38±6	39±5	42±5
Total cholesterol‡	171±28	180±36	191±34
Triglyceride	98±50	98±55	111±64
HDL-C	46±10	51±13	47±12
LDL-C‡	106±26	110±30	122±30

Data are from unrelated men and women who were not using lipid lowering drugs. None of the women were postmenopausal, and none used hormones. *For AA vs. CC, $P < 0.05$, Wilcoxon's test. ‡For AA vs. CC, $P < 0.01$, Wilcoxon's test. Units are mg/dl.

shared at least one *CYP7* allele with a sibling who also had plasma LDL-C concentrations exceeding the 90th percentile. No sequence polymorphisms were detected in the coding regions or intron/exon boundaries of these individuals. A C→A polymorphism 278 bp upstream of the translation initiation codon, and a T→C polymorphism 554 bp upstream of the initiation codon were identified in three individuals. Of 37 individuals homozygous for the C allele at −278, 35 were homozygous and 2 were heterozygous for the T allele at −554, whereas 17 of 18 individuals homozygous for the A allele at −278 were homozygous for the C allele at −554. These results indicate that the two polymorphisms were in almost complete linkage disequilibrium. Therefore, association studies were performed using the polymorphism at −278 only. In 132 men and 163 women, the frequency of the −278 C allele was 0.42, and the genotype frequencies were in Hardy-Weinberg equilibrium.

Association studies. Association studies were performed to test the relationship between the *CYP7* −278 polymorphism and plasma LDL-C concentrations. In the 24 nuclear families that had both CC and AA offspring, plasma LDL-C concentrations were significantly higher ($P = 0.03$) in CC individuals than in their AA siblings. In unrelated individuals, LDL-C concentrations were higher in CC than in AA homozygotes both in men and in women (Table IV).

Discussion

Several studies (3–6) have indicated that ~ 50% of the variation in plasma LDL-C concentrations in the general population is genetically determined. Polymorphism in the genes controlling LDL metabolism must therefore be an important determinant of interindividual differences in plasma LDL levels, but the number of genes involved and the magnitude of their individual effects are not known. As a first step towards elucidating the genetic architecture of plasma LDL-C concentrations, we used a sequential approach comprising linkage analysis, DNA sequencing, and association studies to assess the overall effect of polymorphism in three candidate genes (*LDLR*, *APOB*, and *CYP7*) on plasma LDL-C concentrations in 150 nuclear families. This analysis resulted in three primary observations: first, statistically significant linkage between plasma LDL-C concentrations and *CYP7* was detected. Allelic variation in *CYP7* accounted for 15% of the overall variation in plasma LDL-C concentrations. Analysis of selected samples revealed linkage between *CYP7* and high plasma LDL-C concentrations. No relationship between *CYP7* and low plasma LDL-C concentrations was observed. Therefore, alleles of *CYP7* (or a closely linked locus) that confer susceptibility to increased plasma LDL-C concentrations must be common in the population. Second, an allele defined by two polymorphisms in the 5′ flanking region of *CYP7* was found to be associated with an increase in plasma LDL-C concentrations both in sibling pairs and in unrelated individuals. This observation strongly supports the linkage results, and indicates that genetic variation in cholesterol 7α-hydroxylase activity contributes to variation in plasma LDL-C concentrations. Third, no evidence for linkage between plasma LDL-C concentrations and *LDLR* or *APOB* was obtained. This finding indicates that common polymorphism in these genes accounts for little of the heritable variation in plasma LDL-C concentrations.

For each of the candidate genes, a comprehensive series of analyses was performed to protect against spurious linkage. Initial testing was performed using the sibling pair method developed by Haseman and Elston. This procedure was chosen because it is based on simple genetic principles (28), is robust to nonindependent sibling-pairs (33), and can readily be adapted to nonparametric statistical tests (34). Furthermore, although the procedure was originally based on a simple biallelic monogenic model (28), we have shown that it can also be applied to more complex traits influenced by multiple genes with two or more alleles (35). Therefore, the method is appropriate for genetic analysis of plasma LDL-C concentrations. Since the Haseman-Elston procedure is sensitive to outliers in the data, analyses were repeated after excluding pairs with extremely discordant plasma LDL-C concentrations. Corroborative testing was then performed using a variance components procedure developed by Amos (31). This procedure offers greater statistical power than the Haseman-Elston method, and addresses some of the limitations of the sibling pair approach. The variance components method uses nuclear family data rather than sibling pairs, can directly accommodate covariates such as age and sex, and allows for the effects of genes other than the candidate gene (residual polygenic effects). The use of both methods reduces the possibility of errors due to violation of the statistical or modelling assumptions on which the tests are based. For each of the genes examined in this study, the two methods of analysis provided highly congruent results.

Pharmacological manipulation of *CYP7* expression by administering bile acid–binding resins leads to a fivefold increase in cholesterol 7α-hydroxylase activity (36, 37) that is associated with a 15–20% decrease in plasma LDL-C concentrations. To date, however, the effects of polymorphism in this gene on plasma LDL-C concentrations have not been reported. In this study, evidence for linkage between *CYP7* and plasma LDL-C concentrations was obtained with both sibling

pair and variance components procedures using parametric and nonparametric tests. Under all conditions examined, the *P* values obtained were below or very close to the nominal significance threshold for linkage (0.05). Further evidence for linkage was provided by the observation that plasma LDL-C concentrations were significantly more highly correlated among siblings sharing both *CYP7* alleles in common than among siblings who shared neither *CYP7* allele. Variance components analysis indicated that polymorphism at the *APOE* and *CYP7* loci accounts for 11 and 15%, respectively, of the variation in plasma LDL-C concentrations in this sample of white Americans. Thus, the overall effect of *CYP7* polymorphism in the population appears to be comparable in magnitude to that of *APOE*. The possibility that the linkage observed between *CYP7* and plasma LDL-C concentrations is due to violation of the statistical or modeling assumptions of the test procedures seems very unlikely, since evidence for linkage was obtained using both parametric and nonparametric statistical procedures under genetically robust models. Furthermore, the observed linkage could not be ascribed to outliers, nonindependent sibling pairs, or to covariates known to affect LDL, including age, sex, and pharmacological drug use.

To determine whether allelic variation at the *CYP7* locus was specifically linked to high plasma LDL-C concentrations, further linkage studies were performed using individuals from Group 1 who had high plasma LDL-C concentrations as probands. Siblings who inherited both *CYP7* alleles in common with these probands had significantly higher plasma LDL-C concentrations than did siblings who shared no *CYP7* alleles in common with the probands. This finding was replicated in an independent group of families (Group 2). In both groups, parametric and nonparametric tests provided evidence for linkage between *CYP7* and high plasma LDL-C using a plasma LDL-C concentration above the 90th percentile as a threshold for proband selection. Evidence for linkage was also significant at the less stringent 80th percentile threshold. A parallel analysis using probands selected for low plasma LDL-C concentrations revealed no evidence for linkage between *CYP7* and low plasma LDL-C concentrations.

The simplest explanation of the linkage observed between plasma LDL-C concentrations and *CYP7* is that polymorphisms in the *CYP7* gene that cause increased plasma LDL-C concentrations are common in the population. To identify alleles that encode a less active cholesterol 7α-hydroxylase protein, we sequenced the coding region and intron/exon boundaries of *CYP7* in 10 individuals with high plasma LDL-C concentrations. By sequencing 10 randomly selected individuals, any allele with a population frequency of 0.1 or more would be detected with a probability of $\sim 88\%$ ($1–0.81^{10}$); by using individuals selected for high LDL-C, the probability of detecting an allele associated with high LDL-C is even greater. Sequencing both the coding and noncoding strands revealed no polymorphism in these individuals. Therefore, it is unlikely that the linkage observed between *CYP7* and high plasma concentrations of LDL-C is due to a common polymorphism in the coding region of this gene.

Sequencing of the 5′ flanking region of *CYP7* identified two polymorphisms (−278 C→A and −554 C→T) that were in almost complete linkage disequilibrium. In three independent samples including a group of families having offspring with both CC and AA genotypes, a group of unrelated men and a group of unrelated women, plasma LDL-C concentrations were significantly higher in −278 C homozygotes than in −278 A homozygotes. Therefore, these polymorphisms define a common *CYP7* allele associated with a systematic increase in plasma LDL-C concentrations. Taken together, the results of the association studies and the linkage studies in unselected and selected samples provide strong evidence that polymorphism in *CYP7* or in a closely linked gene contributes to heritable variation in plasma LDL-C concentrations. Interestingly, plasma HDL-C concentrations were also higher in men who were homozygous for the −278 C allele. The significance of this finding is unclear since the CYP7 genotype did not systematically affect plasma HDL-C concentrations in women. However, Machleder et al. have reported that cholesterol 7α-hydroxylase activity segregates with plasma HDL-C concentrations in mice (38). Further studies will be required to assess the relationship between CYP7 polymorphism and plasma HDL-C concentrations.

Since we did not measure cholesterol 7α-hydroxylase activity directly, we can only speculate on the mechanisms underlying the association observed between *CYP7* and plasma LDL-C concentrations. Since cholesterol 7α-hydroxylase activity is inversely related to plasma LDL-C concentrations, the nucleotide substitutions at −278 or −554 may reduce expression of cholesterol 7α-hydroxylase protein by reducing the rate of transcription from the *CYP7* gene. It is also possible that these polymorphisms do not directly affect expression of cholesterol 7α-hydroxylase protein, but that another linked polymorphism in *CYP7* reduces gene expression. Finally, we cannot exclude the possibility that the −278 allele is in linkage disequilibrium with an LDL-raising allele of another as yet unidentified gene, and that allelic variation in a gene closely linked to *CYP7* rather than in *CYP7* itself causes systematic variation in plasma LDL-C concentrations. Although none of the genes known to regulate LDL metabolism have been mapped to this region (8q11-q12), other as of yet unidentified genes that influence plasma LDL-C concentrations may be found near the *CYP7* locus.

Mutations in *LDLR* and *APOB* account for the overwhelming majority of Mendelian hyper- and hypocholesterolemia. The dyslipidemias associated with these mutations are dominant, indicating that a decrease in the function of either gene by 50% (corresponding to the loss of a single allele) leads to markedly altered plasma LDL-C concentrations. Therefore, it might be expected that polymorphisms associated with more subtle alterations of gene function would lead to moderate changes in plasma LDL-C concentrations. However, the present results indicate that polymorphisms in *LDLR* and *APOB* are not a significant cause of heritable variation in plasma LDL-C concentrations among unselected individuals. Comprehensive analysis using both sibling pair and variance components methods revealed no evidence for linkage between plasma LDL-C concentrations and either of these loci. Since linkage methods based on allele sharing are known to have low power to detect genes with modest effects on trait levels, we assessed the sensitivity of the analyses used in this study by testing for linkage between plasma LDL-C concentrations and *APOE*, a gene that accounts for $\sim 10\%$ of the variance in plasma LDL-C concentrations. Under all conditions tested, statistically significant linkage was detected between plasma LDL-C concentrations and *APOE*, indicating that the sample size and analytical procedures used in this study were sufficient to detect genes with moderate effects on

plasma LDL-C concentrations. Therefore, it seems unlikely that the failure to detect linkage between plasma LDL-C concentrations and allelic variation in *LDLR* and *APOB* reflects an inadequate sample size. The results of the linkage analyses were supported by the observation that plasma LDL-C concentrations were equally well correlated among siblings sharing no alleles of *LDLR* or *APOB*, and among siblings sharing both alleles of either candidate gene. The possibility that linkage between plasma LDL-C concentrations and *LDLR* and/or *APOB* was obscured by inadvertent bias in the recruitment of the study subjects also seems highly unlikely because (*a*) the subjects were recruited without regard to any phenotype, (*b*) the distribution of plasma LDL-C concentrations among individuals in Group 1 match those of a large population-based sample, and (*c*) the heritability estimate of plasma LDL-C concentrations from our data ($\sim 50\%$) is in close agreement with estimates reported by previous investigators. Accordingly, the most likely explanation for our data is that polymorphisms in *LDLR* and *APOB* that are systematically associated with altered plasma concentrations of LDL-C are not common in the general population. This finding is consistent with the results of Ludwig et al. (39), who found functionally significant *APOB* polymorphisms in only a small minority ($< 2\%$) of hypercholesterolemic patients. In contrast, Gavish et al. (40) found evidence for allele-specific differences in plasma apoB concentrations in 2 out of 20 individuals with low (< 85 mg/dl) levels of LDL-C. Therefore, while polymorphism in *APOB* does not appear to contribute appreciably to overall variation in plasma LDL-C concentrations, the effects of allelic variation at this locus may be more significant among individuals with low plasma LDL-C concentrations.

The results of this study have significant implications for the genetic architecture of plasma LDL-C concentrations. The overall heritability of plasma LDL-C concentrations in this study was 50%, consistent with the observations of several previous studies. Genetic factors are therefore an important determinant of interindividual differences in plasma LDL-C concentrations, even among unselected individuals. Surprisingly, *LDLR* and *APOB*, the genes responsible for the overwhelming majority of cases with Mendelian disorders of LDL metabolism, contribute little to the heritable variation in plasma LDL-C concentrations in the general population. A comprehensive series of linkage analyses and association studies indicated a significant role for *CYP7* in determining plasma LDL-C concentrations. Polymorphism at the *APOE* and *CYP7* loci accounts for as much as one half of the genetic variation in plasma LDL-C concentrations. This finding indicates that heritable variation in plasma LDL-C concentrations is not due to allelic variation in a single gene (monogenic), and is also not due to the aggregate effect of subtle variations in a large number of genes (polygenic). Rather, heritable variation in plasma LDL-C concentrations is due primarily to allelic variation in a few key genes (oligogenic), each of which have moderate effects on plasma LDL-C concentrations.

Acknowledgments

We thank Betty Pham and Sijing Niu for excellent technical assistance, Dick Verstraete and Ned Warner for family recruitment, and David Russell, Helen Hobbs, and Ann White for helpful discussions.

This work was supported by National Institutes of Health grants HL-53917 and HL-29252.

References

1. National Cholesterol Education Program. 1993. Second report of the expert panel on detection, evaluation, and treatment of high blood cholesterol in adults. National Institutes of Health Publication No. 93-3095.

2. Grundy, S.M. 1991. George Lyman Duff Memorial Lecture. Multifactorial etiology of hypercholesterolemia. Implications for prevention of coronary heart disease. *Arterioscler. Thromb.* 11:1619–1635.

3. Perusse, L., J.-P. Despres, A. Tremblay, C. Leblanc, J. Talbot, C. Allard, and C. Bouchard. 1989. Genetic and environmental determinants of serum lipids and lipoproteins in French Canadian families. *Arteriosclerosis.* 9:308–318.

4. Rice, T., G.P. Vogler, T.S. Perry, P.M. Laskarzewski, and D.C. Rao. 1991. Familial aggregation of lipids and lipoproteins in families ascertained through random and nonrandom probands in the Iowa Lipid Research Clinics Family Study. *Hum. Hered.* 41:107–121.

5. Austin, M.A., M.C. King, R.D. Bawol, S.B. Hulley, and G.D. Friedman. 1987. Risk factors for coronary heart disease in adult female twins. Genetic heritability and shared environmental influences. *Am. J. Epidemiol.* 125:308–318.

6. Bucher, K.D., Y. Friedlander, E.B. Kaplan, K.K. Namboodiri, J.D. Kark, S. Eisenberg, Y. Stein, and B.M. Rifkind. 1988. Biological and cultural sources of familial resemblance in plasma lipids: a comparison between North America and Israel—the Lipid Research Clinics Program. *Genet. Epidemiol.* 5:17–33.

7. Goldstein, J.L., H.H. Hobbs, and M.S. Brown. 1995. Familial hypercholesterolemia. *In* The Metabolic and Molecular Basis of Inherited Disease. C.R. Scriver, A.L. Beaudet, W.S. Sly, and D. Valle, editors. McGraw-Hill Inc., New York. 1981–2030.

8. Myant, N.B. 1993. Familial defective apolipoprotein B-100: a review, including some comparisons with familial hypercholesterolaemia. *Atherosclerosis.* 104:1–18.

9. Linton, M.F., R.V. Farese, Jr., and S.G. Young. 1993. Familial hypobetalipoproteinemia. *J. Lipid Res.* 34:521–541.

10. Humphries, S., D.A. Coviello, P. Masturzo, R. Balestreri, G. Orecchini, and S. Bertolini. 1991. Variation in the low density lipoprotein receptor gene is associated with differences in plasma low density lipoprotein cholesterol levels in young and old normal individuals from Italy. *Arterioscler. Thromb.* 11:509–516.

11. Gaffney, D., D.J. Freeman, J. Shepherd, and C.J. Packard. 1993. The ins/del polymorphism in the signal sequence of apolipoprotein B has no effect on lipid parameters. *Clin. Chim. Acta.* 218:131–138.

12. Hansen, P.S., L.U. Gerdes, I.C. Klausen, N. Gregersen, and O. Faergeman. 1993. Polymorphisms in the apolipoprotein B-100 gene contributes to normal variation in plasma lipids in 464 Danish men born in 1948. *Hum. Genet.* 91: 45–50.

13. Klausen, I.C., P.S. Hansen, L.U. Gerdes, N. Rudiger, N. Gregersen, and O. Faergeman. 1993. A PvuII polymorphism of the low density lipoprotein receptor gene is not associated with plasma concentrations of low density lipoproteins including LP(a). *Hum. Genet.* 91:193–195.

14. de Knijff, P., A.M. van den Maagdenberg, R.R. Frants, and L.M. Havekes. 1994. Genetic heterogeneity of apolipoprotein E and its influence on plasma lipid and lipoprotein levels. *Hum. Mut.* 4:178–194.

15. Boerwinkle, E., and C.F. Sing. 1987. The use of measured genotype information in the analysis of quantitative phenotypes in man. III. Simultaneous estimation of the frequencies and effects of the apolipoprotein E polymorphism and residual polygenetic effects on cholesterol, betalipoprotein and triglyceride levels. *Ann. Hum. Genet.* 51:211–226.

16. Superko, H.R., P. Greenland, R.A. Manchester, N.A. Andreadis, G. Schectman, N.H. West, D. Hunninghake, W.L. Haskell, and J.L. Probstfield. 1992. Effectiveness of low-dose colestipol therapy in patients with moderate hypercholesterolemia. *Am. J. Cardiol.* 70:135–140.

17. Vessby, B., G. Kostner, H. Lithell, and J. Thomis. 1982. Diverging effects of cholestyramine on apolipoprotein B and lipoprotein Lp(a). A dose-response study of the effects of cholestyramine in hypercholesterolaemia. *Atherosclerosis.* 44:61–71.

18. Glueck, C.J., S. Ford, Jr., D. Scheel, and P. Steiner. 1972. Colestipol and cholestyramine resin. Comparative effects in familial type II hyperlipoproteinemia. *JAMA.* 222:676–681.

19. Jones, R.J., and L. Dobrilovic. 1970. Lipoprotein lipid alterations with cholestyramine administration. *J. Lab. Clin. Med.* 75:953–966.

20. Miller, N.E., P. Clifton-Bligh, and P.J. Nestel. 1973. Effects of colestipol, a new bile-acid-sequestering resin, on cholesterol metabolism in man. *J. Lab. Clin. Med.* 82:876–890.

21. Warnick, G.R., M.C. Cheung, and J.J. Albers. 1983. Comparison of current methods for high-density lipoprotein cholesterol quantitation. *Clin. Chem.* 29:538–542.

22. Friedewald, W.T., R.I. Levy, and D.S. Fredrickson. 1972. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. *Clin. Chem.* 18:499–502.

23. Sandkamp, M., B. Tambyrajah, H. Schriewer, and G. Assmann. 1988. Simplified turbidimetric determination of apolipoproteins A-I, A-II and B using a microtitre method. *J. Clin. Chem. Clin. Biochem.* 26:685–688.

24. Marcovina, S.M., J.J. Albers, H. Kennedy, J.V. Mei, L.O. Henderson, and W.H. Hannon. 1994. International Federation of Clinical Chemistry standardization project for measurements of apolipoproteins A-I and B. IV. Com-

parability of apolipoprotein B values by use of International Reference Material. *Clin. Chem.* 40:586–592.

25. Johns Hopkins University (1997). *The Genome Database* [online]. Available: http://gdbwww.gdb.org

26. Applied Biosystems International PRISM™. 1997. Dye Terminator Cycle sequencing ready reaction kit Protocol. Protocol # 402078.

27. Falconer, D.S. 1989. Introduction to Quantitative Genetics. Longman Group UK Ltd., Harlow, United Kingdom. 148–186.

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

29. Guerra, R., J.P. Wang, S.M. Grundy, and J.C. Cohen. 1997. A hepatic lipase (LIPC) allele associated with high plasma concentrations of high density lipoprotein cholesterol. *Proc. Natl. Acad. Sci. USA.* 94:4532–4537.

30. Hopper, J.L., and J.D. Mathews. 1982. Extensions to multivariate normal models for pedigree analysis. *Ann. Hum. Genet.* 46:373–383.

31. Amos, C.I. 1994. Robust variance-components approach for assessing genetic linkage in pedigrees. *Am. J. Hum. Genet.* 54:535–543.

32. Amos, C.I., D.K. Zhu, and E. Boerwinkle. 1996. Assessing genetic linkage and association with robust components of variance approaches. *Ann. Hum. Genet.* 60:143–160.

33. Blackwelder, W.C., and R.C. Elston. 1982. Power and robustness of sib-pair linkage test and extension to larger sibships. *Commun. Statis.* 11:449–484.

34. Wan, Y., J. Cohen, and R. Guerra. 1997. A permutation test for the robust sib-pair linkage method. *Ann. Hum. Genet.* 61:79–87.

35. Stoesz, M.R., J.C. Cohen, V. Mooser, S. Marcovina, and R. Guerra. 1997. Extension of the Haseman-Elston method to multiple alleles and multiple loci. Theory and practice for candidate genes. *Ann. Hum. Genet.* 61:263–274.

36. Reihner, E., I. Bjorkhem, B. Angelin, S. Ewerth, and K. Einarsson. 1989. Bile acid synthesis in humans: regulation of hepatic microsomal cholesterol 7 alpha-hydroxylase activity. *Gastroenterology.* 97:1498–1505.

37. Bjorkhem, I., E. Reihner, B. Angelin, S. Ewerth, J.E. Akerlund, and K. Einarsson. 1987. On the possible use of the serum level of 7 alpha-hydroxycholesterol as a marker for increased activity of the cholesterol 7 alpha-hydroxylase in humans. *J. Lipid Res.* 28:889–894.

38. Machleder, D., B. Ivandic, C. Welch, L. Castellani, K. Reue, and A.J. Lusis. 1997. Complex genetic control of HDL levels in mice in response to an atherogenic diet. Coordinate regulation of HDL levels and bile acid metabolism. *J. Clin. Invest.* 99:1406–1419.

39. Ludwig, E.H., P.N. Hopkins, A. Allen, L.L. Wu, R.R. Williams, J.L. Anderson, R.H. Ward, J.M. Lalouel, and T.L. Innerarity. 1997. Association of genetic variations in apolipoprotein B with hypercholesterolemia, coronary artery disease, and receptor binding of low density lipoproteins. *J. Lipid Res.* 38: 1361–1373.

40. Gavish, D., E.A. Brinton, and J.L. Breslow. 1989. Heritable allele-specific differences in amounts of apoB and low-density lipoproteins in plasma. *Science.* 244:72–76.