Genome-Wide Linkage Analysis Reveals Evidence of Multiple Regions That Influence Variation in Plasma Lipid and Apolipoprotein Levels Associated With Risk of Coronary Heart Disease

Kathy L. Klos, Sharon L.R. Kardia, Robert E. Ferrell, Stephen T. Turner, Eric Boerwinkle, Charles F. Sing

Abstract—Results of genome-wide linkage analyses to identify chromosomal regions that influence interindividual variation in plasma lipid and apolipoprotein levels in the Rochester, Minn, population are reported. Analyses were conducted for total cholesterol (total-C), triglycerides (TGs), high density lipoprotein cholesterol (HDL-C), apolipoprotein A-I, apolipoprotein A-II, apolipoprotein B, apolipoprotein C-II, apolipoprotein C-III, apolipoprotein E, the total-C/HDL-C ratio, and the TG/HDL-C ratio. Genotypes were measured for 373 genome-wide marker loci on 1484 individuals distributed among 232 multigeneration pedigrees sampled without regard to health status. LOD scores and estimates of additive genetic variance associated with map locations were obtained by using the variance-component method of linkage analysis. No evidence of linkage with genes influencing variation in age served as a negative control. Plasma apolipoprotein E levels and the apolipoprotein E gene served as a positive control (LOD score 4.20). Evidence (LOD score >2.00) was provided that was suggestive of a gene or genes on chromosomes 4 and 5 influencing variation in the apolipoprotein A-II level, on chromosome 12 influencing variation in the apolipoprotein A-I level, and on chromosome 17 influencing variation of total-C/HDL-C. These analyses provide new information about genomic regions in humans that influence interindividual variation in plasma lipid and apolipoprotein levels and serve as a basis for further fine-mapping studies to identify new genes involved in lipid metabolism. (Arterioscler Thromb Vasc Biol. 2001;21:971-978.)

Key Words: genetic linkage ■ apolipoprotein E ■ apolipoprotein A-I ■ apolipoprotein A-II ■ total cholesterol ■ high density lipoproteins

One of the greatest challenges that the medical and public health communities face is to identify and characterize the genes that influence the risk of coronary artery disease (CAD). Only a small fraction of the genes that are expected to be involved in determining the risk of CAD are known. Measures of lipid metabolism are well-established intermediate traits that translate genetic variation into variation in the risk of CAD. Increased risk of developing CAD has been associated with elevated triglyceride (TG),1–3 total cholesterol (total-C),4–5 and LDL cholesterol6,7 levels and with decreased levels of HDL cholesterol (HDL-C).8,9 CAD has also been associated with low levels of plasma apoA-I and apoA-II and high levels of apoB and apoE.9–14 Only a fraction of the genetic variation in intermediate measures of lipid metabolism is explained by variation in known candidate genes. For example, the total genetic contribution to variation in total-C, HDL-C, and other lipid traits is estimated to be between 40% and 80%.15,16 However, single polymorphisms typically explain only 5% to 15% of that variation.16–20 The present study is part of an effort to find new genes that influence the genetic architecture of CAD by carrying out genetic studies of intermediate measures of lipid metabolism that translate genetic variation into variation in risk of CAD. The genetic architecture of a trait is defined by the number of genes, the number of alleles per gene and their relative frequencies, and the gene-gene and gene-environment interactions that combine to influence interindividual variation within a population.21 Genome-wide linkage scans have the potential for identifying regions of the genome not previously recognized as sources of interindividual variation in intermediate traits that influence variation in the risk of CAD.22–26

To identify regions of the genome predictive of interindividual variation in plasma lipid levels for the purpose of identifying new candidate genes, we have applied a variance
component–based linkage method to phenotypic data in pedigrees ascertained without regard to health status. Evidence of the linkage of marker loci with a gene(s) influencing interindividual variation in plasma apoE level was observed in the region of chromosome 19 containing the apoE gene. LOD scores provided suggestive evidence of the linkage of marker loci with previously unknown genes on chromosomes 17 influencing the total-C/HDL-C ratio.

Methods

Sample and Laboratory Methods

As part of the Rochester Family Heart Study, multigeneration pedigrees were ascertained without regard for health through households with ≥2 children enrolled in the primary and secondary schools of Rochester, Minn. Sampling details, the clinic examination protocol, and baseline characteristics have been described by Moll et al.32,33 Marker genotype data were obtained for 1484 individuals (779 females and 705 males) of the 4486 individuals making up the 232 pedigrees (see Table 1). There was an average of 2.94 children per pedigree. Individuals were genotyped for 373 marker loci with previously unknown genes on chromosomes 17 influencing the total-C/HDL-C ratio.

Linkage Analysis

The lipid levels were adjusted for gender-specific age, height, and weight to the third order, as well as body mass index (BMI), before carrying out the linkage analysis. Marker relative allele frequencies were estimated as the sample relative allele frequencies. Genetic map location estimates were taken from the maps prepared by the Center for Medical Genetics at the Marshfield Medical Research Foundation.33 Locations for markers not contained in the Marshfield map were inferred from their map locations in the Genetic Location Database.36

Linkage analyses were conducted by the variance-components method37 with use of the SOLAR computer program. In this method, a linear mixed model was fit to the trait data so that the phenotypic variance about the trait mean is partitioned into a monogenic component ($\sigma^2_{\text{BL}}$) representing the contribution of a quantitative trait locus (QTL) linked to the marker locus, a residual familial component ($\sigma^2_{\text{F}}$) attributable to familial genetic and shared environmental effects, and a component ($\sigma^2_{\text{E}}$) attributable to environmental effects unique to the individual. With the assumption of no recombination between trait and marker loci, the phenotypic variance-covariance matrix ($\Omega$) for individuals in a pedigree may be written as

$$\Omega = \Phi \sigma^2_{\text{BL}} + \Pi \sigma^2_{\text{F}} + 1 \sigma^2_{\text{E}},$$

where $\Phi$ is a matrix of the proportion of alleles shared identical by descent (IBD) at a point in the genome estimated from the genotypic data, $\Pi$ is a matrix of the expected proportion of alleles shared IBD for pairs of relatives, and $I$ is an identity matrix. Multipoint estimates of IBD were obtained as a weighted average of the IBD at each individual marker.34 The estimates of IBD for individual markers were obtained by the computer program SOLAR35 by using the relationship information from the 4486 individuals making up the 232 pedigrees and the genotype data for the subset of 1484 individuals who were genotyped for the present study. Missing genotype values in the subset of 1484 were imputed by using information from flanking markers before full multipoint estimation.37

All LOD scores reported below are based on multipoint analyses. The LOD scores were calculated as the difference between the maximum of the $\log_{10}$ likelihood of the full model, including estimates of $\sigma^2_{\text{BL}}, \sigma^2_{\text{F}},$ and $\sigma^2_{\text{E}},$ and the maximum of the $\log_{10}$ likelihood of the reduced model in which $\sigma^2_{\text{BL}}$ was constrained to equal 0. Twice the difference between the maximum of the log likelihood of these 2 models is a likelihood ratio test statistic asymptotically distributed as a $1/2:1/2$ mixture of a $\chi^2$ with 0 and 1 df. In 3 cases, the linkage analysis yielded evidence of ≥1 QTL influencing interindividual variation in a trait (multipoint LOD >2.00). In these cases, a second linkage analysis was performed with the contribution of the first QTL to trait variation fixed in the full and reduced models.39 All linkage analyses were carried out in SOLAR with use of the t distribution function, which provides a better fit to data that do not meet assumptions of normality. We have used a LOD score of ≥3.00 to indicate adequate evidence of linkage, a LOD threshold of ≥2.00 as suggestive, and a LOD score of ≥1.30 as tentative evidence of linkage.40

Results

The phenotypic characteristics of the 1482 individuals in 232 pedigrees used in these linkage analyses are summarized in Table 2. The average age of males was 36.3 years (range 5.2 to 89.3 years) and of females was 39.0 years (range 5.4 to 90.3 years). The average BMI was 23.4 kg/m$^2$ (range 13.7 to 50.8 kg/m$^2$) for females and 23.7 kg/m$^2$ (range 13.8 to 43.3 kg/m$^2$) for males. On average, 1263 (range 819 to 1392)
individuals were genotyped for a marker locus. The average individual was successfully genotyped for 320 markers.

The multipoint linkage analysis results of the genome scan are presented in Table 3 as the highest LOD score observed for each chromosome for each of 11 plasma lipid and apolipoprotein traits. LOD peaks $>2.00$ are indicated in boldface type. LOD peaks $>2.00$ are indicated in boldface type; LOD scores $>3.00$ were observed for the apoE level on chromosome 19 in the region containing the apoE gene. In the analysis of age, we observed no LOD score $>0.00$ (Table 3).

### Table 3. Peak Multipoint LOD Scores and Positions on Each Chromosome for 11 Plasma Lipid and Apolipoprotein Traits

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>Peak LOD Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>ApoA-I</td>
<td>ApoA-II</td>
</tr>
<tr>
<td>1</td>
<td>0.69</td>
</tr>
<tr>
<td>(76 cM)</td>
<td>(88 cM)</td>
</tr>
<tr>
<td>2</td>
<td>0.75</td>
</tr>
<tr>
<td>(169 cM)</td>
<td>(240 cM)</td>
</tr>
</tbody>
</table>
Figure 1 depicts the linkage results for 4 traits (apoA-I, apoA-II, apoE, and total-C/HDL-C) for the chromosomes for which LOD peaks >1.30 were observed. The highest LOD score observed was 4.20, corresponding to a chromosomal region influencing the plasma apoE level on 19q, near the marker D19S246. An interval 40 cM wide between markers D19S245 and D19S254 had LOD scores >2.00. Tentative evidence for linkage to a region influencing variation in apoE was also found on chromosome 12p near the marker GATA91H06 (LOD score 1.50). To test whether the multiple QTLs identified as influencing variation in apoE levels formed a true multilocus system, we ran a multipoint linkage scan of the genome by using full and reduced models, which both included the contribution of the QTL on chromosome 19. A LOD score from this analysis of 1.30 on chromosome 12 near the marker GATA91H06 supports the initially observed evidence of linkage to a QTL in this region that influences variation in apoE levels.

Suggestive evidence of a QTL near the marker D4S2368 on chromosome 4q (LOD score 2.35) and a QTL on 5p (LOD score 2.13) near the marker D5S2500, which influence variation in apoA-II levels, was detected. Tentative evidence was observed for a third QTL influencing apoA-II variation on chromosome 18 near the marker D18S976 (LOD score 1.53). Linkage analysis in which the full and reduced models included the contribution of the chromosome 4 QTL resulted in a LOD score of 1.55 for a second QTL on chromosome 5 near D5S2500. Suggestive evidence was found of linkage near marker D12S2070 on chromosome 12 (LOD score 2.02) with a QTL influencing plasma apoA-I levels. Evidence suggestive of a QTL near marker D17S928 at the end of the long arm of chromosome 17 (LOD score 2.48) that influences plasma total-C/HDL-C variation was observed. Additional tentative evidence of a QTL influencing variation in total-C/HDL-C was found on chromosome 5q near the marker D5S408 (LOD score 1.57). Linkage analysis with the contribution of the chromosome 17 QTL included in the full and reduced models resulted in a LOD score of 1.26 for a second QTL at the chromosome 5 location.

**Discussion**

The purpose of the present study was to identify genomic regions containing genes influencing interindividual variation in plasma lipid and apolipoprotein levels in nonhispanic white families from Rochester, Minn, selected without regard to health status. The results of our linkage scan provide evidence of regions of the genome that may contain previously unknown genes that influence normal interindividual variation in apoA-I, apoA-II, apoE, and total-C/HDL-C. Positional candidate genes in regions identified by multipoint linkage analyses are summarized in Table 4.

The genome-wide linkage scan for loci influencing age serves as a negative control for the variance-components linkage method in this population. Figure 2 contrasts the
multipoint linkage results for age with those for plasma apoE on chromosome 19. The fact that no LOD score >0.00 was observed for age on any chromosome is in accord with the low type I error rate (0.0025) estimated for this method in the GAW10 problem set of simulated nuclear families.

The apoE gene region on chromosome 19q serves as a positive control for the variance-components linkage method. There is a large body of evidence supporting the major contribution of the apoE gene to variation in plasma apoE level and to the risk of developing CAD.11,14,42–45 The 3 major apoE isoforms (E2, E3, and E4) account for between 11% and 20% of the interindividual variation in levels of apoE in this and other populations.16,33,46 The highest LOD score observed in the present study (LOD score 4.20, near the marker D19S246) identifies a region of chromosome 19 that includes the apoE/C-I/C-II gene cluster. The width of this peak, covering more than one third of chromosome 19 (LOD score >2.00), may be due to the major effect of the apoE gene on interindividual variation in the apoE level. Alternately, the width of the region with LOD scores suggestive of linkage may indicate overlapping peaks that are due to genes other than apoE within this region with effects on lipid metabolism. To test whether variation in the apoE gene was responsible for the multipoint LOD peak observed on chromosome 19, we adjusted plasma apoE levels for apoE genotype means (after gender-specific adjustment for variation in age, height, and weight to the third order and BMI). The resulting peak multipoint LOD, 1.10 at 78 cM (Figure 2), is below our threshold for tentative evidence of linkage but may still reflect variation in the apoE gene not measured by the e2, e3, and e4 alleles (J.H. Stengård, personal communication, 2000).

The ability to detect linkage markers with the apoE gene region, when the plasma apoE level is considered, documents the utility of this method to localize genes influencing interindividual variation in plasma lipid traits of complex inheritance. Simulations by Blangero et al39 indicate that a variance-components linkage analysis of 1000 individuals in randomly ascertained extended pedigrees has a power of 80% to detect a QTL that accounts for as little as 20% of trait variation with a LOD score of ≥3.0. However, the pedigrees used in the study of Blangero et al were much larger than the pedigrees used in the present study, and the power of the variance-components method to detect linkage decreases with declining QTL-specific heritability.48 Thus, the many QTLs that are each expected to have considerably smaller effects on genetic variance may not be detected by a linkage analysis performed with this simple model on small randomly ascertained pedigrees.

Polymorphism in the apoE gene has been identified as a predictor of variation in intermediate risk factors for CAD, apart from apoE level, including total-C, TGs, apoB, and HDL-C.13,45,49 For example, the traditional apoE genotypes explain 5% and 23% of variation in TGs in men and women, respectively, in Rochester, Minn.33 Peak multipoint LOD scores on chromosome 19 for TGs did not exceed 2.00, nor did the peak correspond to the position of the apoE gene (Table 3). However, the effect of apoE on TG level varies in contexts such as age and measures of body size.50–53 The context-dependent effects of apoE genotype on other measures of lipid metabolism, and on the correlations among them, are also well known.52,53 Much of the complexity of the genetic architecture of apoE and other plasma lipid and apolipoprotein traits is ignored by the simple additive variance-components model of linkage used in the present study. We chose to initially examine the first-order effects to maximize the power to detect a QTL by using the full sample in this first genome scan. We expect that QTLs with large marginal effects will be detected by this model but that QTLs whose effects are confined to particular strata of the population may be associated with LOD scores below our threshold for linkage.

Despite the limitations of the simple model used in the present study, several regions suggestive of linkage with lipid and apolipoprotein traits within which lie candidate genes for lipid metabolism were identified (Table 4). In addition to the evidence for linkage with a QTL influencing variation in apoE level on chromosome 19, tentative evidence for linkage was observed on chromosome 12. This region contains the gene encoding the catalytic polypeptide of the apoB mRNA-editing enzyme (APOBEC1), which is involved in the metabolism of apoE. In LDL receptor knockout mice, inactivation of APOBEC1 resulted in a 60% increase in plasma apoE.54 Thus, variation in the APOBEC1 gene may affect the risk of CAD by influencing variation in the apoE level in particular

---


<table>
<thead>
<tr>
<th>Trait</th>
<th>Chromosome</th>
<th>Position (cM)</th>
<th>Candidate Gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>ApoA-I</td>
<td>12</td>
<td>0.00</td>
<td>SREBP1</td>
</tr>
<tr>
<td>ApoA-II</td>
<td>4</td>
<td>1.10</td>
<td>CPE</td>
</tr>
<tr>
<td>ApoA-II</td>
<td>5</td>
<td>3.00</td>
<td>HMGCR</td>
</tr>
<tr>
<td>ApoE</td>
<td>12</td>
<td>4.20</td>
<td>APOBEC1</td>
</tr>
<tr>
<td>Total-C/HDL-C</td>
<td>17</td>
<td>6.00</td>
<td>P4HB</td>
</tr>
</tbody>
</table>

CPE indicates carboxypeptidase E; P4HB, procollagen proline 2-oxoglutarate 4-dioxygenase cellular thyroid hormone binding protein.
contexts indexed by the level of LDL receptor expression. Additionally, evidence of a locus on chromosome 12 influencing the risk of Alzheimer’s disease is stronger in families without the apoE e4 allele,55–57 a risk factor for CAD and Alzheimer’s disease.

Regions on chromosomes 4 and 5 showed suggestive evidence of the influence of QTL on apoA-II variation (LOD score >2.00) (Figure 1). ApoA-II is the second most abundant apolipoprotein in HDL (after apoA-I) and may affect the risk of CAD by influencing the rate of cholesterol efflux from peripheral tissues.58 A gene on chromosome 4 that codes for carboxypeptidase E is a positional candidate gene for influencing apoA-II level. Carboxypeptidase E is an enzyme responsible for prohormone sorting and processing, influencing such signaling molecules as pro-opiomelanocortin and insulin, both of which are involved in energy homoeostasis.59–61 Also near the position of the multipoint LOD peak for apoA-II on chromosome 4 is the gene coding for the fibrinogen-β polypeptide, which has been associated with increased risk of peripheral atherosclerosis.62 On chromosome 5, positional candidate genes include cholesterologenes enzymes 3-hydroxy-3-methylglutaryl-coenzyme A reductase (HMGCGR) and 3-hydroxy-3-methylglutaryl coenzyme A synthase 1 (HMGCS1). Inhibition of HMGCR in humans alters the plasma lipid profile, including a significant reduction in the ratio of cholesterol to HDL-C.63 In rats, administration of tetradeoxythiaoacetate acid decreases apoA-II while increasing HMGCS1 mRNA levels, evidence of a link between HMGCS1 expression and the level of liver apolipoproteins.64 Neither of the 2 regions influencing variation in apoA-II that were identified in this population included the structural gene for apoA-II on chromosome 1q21-q23. Variation in the apoA-II gene can account for 10% to 11% of interindividual variation in the apoA-II level.65 However, in the sample considered here, variation in the apoA-II gene does not contribute to variation in apoA-II at a level detectable by linkage analysis (LOD score <0.50, Table 2).

ApoA-I is the most abundant apolipoprotein in HDL-C and is a better predictor of CAD risk than is HDL-C alone.10,13 Studies of the genetic determination of apoA-I have found evidence of 1 major locus,66 2 major loci,67 or polygenes alone68,69 influencing the variation in different populations. The LOD score of 2.02 at 128 cM in the present study suggests that variation in plasma apoA-I may be influenced by a gene on chromosome 12q. A positional candidate gene, scavenger receptor class B type 1 (SRBI), is located near the position of this LOD peak. SRBI, with a binding affinity for multiple apolipoproteins including apoA-I, apoA-II, and apoC-III, acts as a receptor for the selective uptake of HDL cholesterol ester in the liver and steroidogenic tissues and for the efflux of cholesterol from peripheral cells.70 The overexpression of SRBI in mice is accompanied by a substantial, but transitory, decrease in HDL-C and apoA-I levels.71 It is possible that changes in SRBI expression may also influence apoA-I levels in humans.

Evidence suggestive of the influence of a QTL on variation in total-C/HDL-C was observed from the analyses of chromosomes 5 and 17. Regulation of microsomal transfer protein (MTP) expression and activity may influence plasma cholesterol level. There are no obvious candidate genes near this location of the LOD peak on chromosome 5. A positional candidate gene located in the region of the peak LOD score on chromosome 17 is the procollagen proline 2-oxoglutarate 4-dioxygenase cellular thyroid hormone binding protein, a multifunctional protein that is the small subunit of the MTP heterodimer. MTP catalyzes the transport of TGs, cholesteryl esters, and phospholipids between phospholipid surfaces. Defects in the larger subunit of MTP have been previously associated with abetalipoproteinemia.72

Three regions identified in our genome-wide linkage scan, although associated with LOD scores <2.00, are worthy of further mention as supporting evidence for previously published reports of linkage. Aouizerat et al observed strong evidence of linkage on chromosome 11p with a gene contributing to familial combined hyperlipidemia in Dutch pedigrees. The region, near the marker D11S1324 (35 cM), was associated with a combined lipid phenotype that included elevated total-C (>90th percentile). Tentative evidence for linkage (LOD score 1.84 at 34 cM) with a gene(s) on chromosome 11p (34 cM) influencing plasma total-C level was observed in the present study. Knoblauch et al observed strong evidence of linkage with a region of chromosome 13q near the markers D13S1241 and D13S786 (76 to 77 cM) influencing HDL-C, LDL-C, total-C, and BMI in an Arab family from Israel ascertained for familial hypercholesterolemia and in an independent sample of healthy white monozygotic and dizygotic twins from Germany. This region of 13q is very near the QLTs identified in the present study, which may influence TG (LOD score 1.64 at 86 cM) and TG/HDL-C (LOD score 1.37 at 64 cM) levels.

The results of our genome-wide linkage analyses of 11 lipid and apolipoprotein traits provide a focus for future genetic studies. This scan has identified chromosomal regions hypothesized to contain new genes influencing interindividual variation in apoA-I, apoA-II, and total-C/HDL-C and has provided supporting evidence of regions that may contain genes influencing variation in apoE, total-C, TGs, and TG/HDL-C. These results suggest a list of positional candidate genes for detailed DNA sequence analysis aimed at identifying functional mutations that affect variation in the plasma lipid profile. Identification of genes that influence lipid metabolism will lead to a better understanding of the etiology of CAD and provide suggestions for the development of new therapies for the treatment and prevention of disease. Finally, a comparison of our results with the findings of other studies suggests to us the importance of gene-environment interactions and the need for context-dependent linkage analyses to identify regions of the genome that contain genes that influence interindividual variation in plasma lipid and apolipoprotein levels only in particular subdivisions of the population at large.

Acknowledgments

Funding for this research was provided by National Institutes of Health grants HL-39107, HL-51021, and HD-32197. We would like to thank Paul Kopee, Ken Weiss, and Sara Hamon at the University of Michigan and Terry Bertin and Kim Lawson at the University of Texas for their technical assistance and critical input.

References


Genome-Wide Linkage Analysis Reveals Evidence of Multiple Regions That Influence Variation in Plasma Lipid and Apolipoprotein Levels Associated With Risk of Coronary Heart Disease
Kathy L. Klos, Sharon L. R. Kardia, Robert E. Ferrell, Stephen T. Turner, Eric Boerwinkle and Charles F. Sing

doi: 10.1161/01.ATV.21.6.971
Arteriosclerosis, Thrombosis, and Vascular Biology is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2001 American Heart Association, Inc. All rights reserved.
Print ISSN: 1079-5642. Online ISSN: 1524-4636

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://atvb.ahajournals.org/content/21/6/971

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Arteriosclerosis, Thrombosis, and Vascular Biology can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to Arteriosclerosis, Thrombosis, and Vascular Biology is online at:
http://atvb.ahajournals.org//subscriptions/