

# APOE/C1/C4/C2 Gene Cluster Genotypes, Haplotypes and Lipid Levels in Prospective Coronary Heart Disease Risk Among UK Healthy Men

Gie Ken-Dror, Philippa J Talmud, Steve E Humphries, and Fotios Drenos

Centre for Cardiovascular Genetics, BHF Laboratories, The Rayne Building, Department of Medicine, Royal Free and University College Medical School, London, United Kingdom

The role of common *APOE* variants on plasma lipids, particularly low density lipoprotein (LDL) levels, and coronary heart disease (CHD) risk is well known; the influence of variation in the other nearby apolipoprotein genes *APOC1*, *APOC4* and *APOC2* is unclear. This study examines the association between *APOE/C1/C4/C2* gene cluster variation using tagging SNPs and plasma lipid concentration along with risk of CHD in a prospective cohort. Genotypes for 11 common *APOE/C1/C4/C2* SNPs were determined in 2,767 middle-aged (49 to 64 years) men from the Second Northwick Park Heart Study, with 275 CHD events over a 15-year follow-up period. Seven SNPs showed significant associations with one or more lipid trait in univariate analysis. Multivariate and haplotype analysis showed that the *APOE* genotypes are most strongly associated with effects on LDL-C and apoB concentration (explaining 3.4% of the LDL-C variance) while the other SNPs in this gene cluster explained an additional 1.2%. Haplotypes in *APOC2* and *APOC4* were associated with modest effects on HDL-C and apoAI (explaining respectively 1.4% and 1.2%). Carriers of the *APOE*  $\epsilon$ 2 SNP had a significantly lower risk of CHD hazard ratio (HR) of 0.63 (95% confidence interval (CI): 0.42–0.95), as did carriers of the *APOC2* SNP rs5127 (HR = 0.72, 95% CI: 0.56–0.93), while carriers of *APOC1* SNP rs4803770 had higher risk of CHD (HR = 1.36, 95% CI: 1.04–1.78) compared with noncarriers. While the common *APOE* polymorphism explains the majority of the locus genetic determinants of plasma lipid levels, additional SNPs in the *APOC1/C2* region may contribute to CHD risk, but these effects require confirmation.

© 2010 The Feinstein Institute for Medical Research, [www.feinsteininstitute.org](http://www.feinsteininstitute.org)

Online address: <http://www.molmed.org>

doi: 10.2119/molmed.2010.00044

## INTRODUCTION

Elevated triglycerides (TG), total cholesterol (TC), low density lipoprotein (LDL-C) and apolipoprotein B (apoB), together with decreased levels of high density lipoprotein (HDL-C), and apolipoprotein AI (apoAI), are associated with an increased risk of CHD (1–4). Genetic and environmental factors are believed to regulate both metabolic and cellular function, and the concentration of lipoproteins and apolipoproteins (4,5). The heritability estimates of lipoproteins and apolipoproteins from twin and family studies are in the range of 40–80% (6,7), suggesting a considerable genetic

contribution, and discovery of the genes that contribute to these changes may lead to a better understanding of these processes.

Two of the most well-studied common genotypes that determine plasma lipid levels are the two single-base changes in the coding region of the *APOE* gene that create the  $\epsilon$ 2,  $\epsilon$ 3 and  $\epsilon$ 4 alleles (8,9). ApoE circulates in the blood as a component of a number of lipoprotein particles, including chylomicrons and their remnants; TG-rich very low density lipoprotein (VLDL) particles and their remnants, and HDL (10). ApoE is a ligand for receptors involved in the clearance of these

lipoproteins, particularly in the liver (10), and the three common alleles of *APOE* code for three isoforms which vary in their affinity for binding to the LDL-receptor (11). The *APOE*  $\epsilon$ 2 allele has reduced binding affinity, causing lower efficiency of clearance of VLDL and chylomicron remnants from the blood to the liver, thus slowing the uptake of postprandial lipoprotein particles compared with carriers of the  $\epsilon$ 3 and  $\epsilon$ 4 alleles (11,12). *APOE*  $\epsilon$ 4 and  $\epsilon$ 3 bind with approximately equal affinity to lipoprotein receptors while  $\epsilon$ 2 binds with less than 2% affinity (13). As would be expected from the relationship between *APOE* genotypes and lipid levels, meta analyses also has clearly demonstrated a strong relationship with risk of CHD (14), and compared with subjects with the genotype  $\epsilon$ 3/ $\epsilon$ 3, the odds ratio for coronary disease was 0.80 (95% confidence interval [CI]: 0.70–0.90) in  $\epsilon$ 2 carriers and 1.06 (95% CI: 0.99–1.13) in  $\epsilon$ 4 carriers.

---

**Address correspondence and reprint requests to** Steve E Humphries, Centre for Cardiovascular Genetics, BHF Laboratories, The Rayne Building, Department of Medicine, Royal Free and University College Medical School, 5 University St, London WC1E 6JF, United Kingdom. Phone: +44 20 7679 6968; Fax: +44 20 7679 6212; E-mail: [rnhaseh@ucl.ac.uk](mailto:rnhaseh@ucl.ac.uk). Submitted April 3, 2010; Accepted for publication May 18, 2010; Epub ([www.molmed.org](http://www.molmed.org)) ahead of print May 20, 2010.

The *APOE* gene is located on the long arm of chromosome 19q12–13.2 within a cluster of other genes coding for apolipoprotein CI (*APOC1*), apolipoprotein CIV (*APOC4*) and apolipoprotein CII (*APOC2*). ApoCI is a constituent of VLDL and HDL, inhibiting the binding of VLDL to LDL-related proteins (15,16), as well as the apoE-mediated binding of VLDL to the LDL receptor (17)—leading to reduced levels of LDL-C (18). ApoCII is an essential cofactor of lipoprotein lipase (LPL), the rate-limiting enzyme for the hydrolysis and removal of triglycerides in chylomicrons and VLDL (19), and individuals who are completely deficient in apoCII have chylomicronaemia and grossly elevated plasma TG levels, although their parents, who have half normal plasma levels of apoCII, have normal TG levels (20). The role of apoCIV in human lipoprotein metabolism is unclear, mainly due to the relatively low apoCIV gene expression in the liver and the low levels of the apoCIV protein detectable in plasma (21), where it is associated mainly with VLDL particles. However, mice transgenic for *APOC4* show elevated levels of TG (22,23) and in one study, common missense variants (L96RL, rs5167) were associated with differences in plasma TG levels in women (23).

Despite the fact that each protein encoded by the *APOE/C1/C4/C2* gene cluster has a clear or potential role in plasma lipid metabolism, the association between variation in the entire cluster and plasma lipid concentration is understudied. Similarly, while the association between the *APOE* variants and CHD risk has been well studied and confirmed by meta-analysis (14), the association between the other cluster gene variants and CHD is less clear. The primary purpose of this study is to evaluate the contribution of common SNPs and haplotypes in the *APOE/C1/C4/C2* gene cluster to plasma lipid concentrations in a large group of healthy middle-aged men. A secondary aim was to determine to what extent SNPs and haplotypes, associated with differences in plasma lipid concentrations, were contributing to the  $\epsilon$ 2 low-

ering and  $\epsilon$ 4 raising effect on prospective risk of CHD.

## MATERIALS AND METHODS

### Study Subjects and Data Collection

The prospective Second Northwick Park Heart Study (NPHSII), commenced in 1989, and includes 3,052 middle-aged men (49–64 years) recruited from nine general medical practices in the UK. Participants were free of unstable angina, myocardial infarction, evidence of silent infarcts, coronary surgery, anticoagulant drugs (including aspirin), cerebrovascular disease, malignancy and any condition or disease preventing the attainment of written, informed consent or long-term follow-up. Informed consent was received from all study subjects. Information on lifestyle habits, height, weight, blood pressure and a number of blood biomarkers was recorded at baseline and on subsequent prospective follow-ups. Details of recruitment, measurements, follow-up and definitions of disease are described elsewhere (24–26). Lipids, TC and TG concentrations were measured with automated enzyme procedures, while serum apoAI and apoB concentrations were measured by immunoturbidometry. CHD events taken as end points were fatal (sudden or not) and non-fatal MI, based on World Health Organization criteria (27), plus coronary artery surgery and silent MI on the follow-up ECG or coronary revascularization procedures and sudden unexplained death. Clinical information for each event was assembled by inquiries through the participating practices, hospitals attended and, for fatal events, coroners' offices. This information was collated and submitted to an independent assessor who assigned qualifying events to the appropriate category. A DNA sample was obtained from 2,767 of the men from a blood sample taken at the time of recruitment. Interviews and repeat measurements were conducted annually for surviving participants.

### Genetic Analysis

Nine SNPs were genotyped using an Illumina GoldenGate candidate gene chip

(Illumina Inc., San Diego, CA, USA) (28). *APOE*: rs405509 (promoter), rs439401 (2 kb); *APOC1*: rs4420638 (200bp), rs4800770 (5 kb), rs7259004 (10kb); *APOC4*: rs12691089 (exon 2), rs5167 (exon 3); *APOC2*: rs5127 (3'URT) rs10413089 (3 kb). Tagging SNPs (tSNPs) for the *APOE/C1/C4/C2* genes cluster were selected with Tagger (developed by scientists at Broad Institute, Cambridge, MA, USA) (29) using the CEU panel of HapMap (see International HapMap Project [<http://hapmap.ncbi.nlm.nih.gov/>]), applying an  $r^2$  threshold of 0.8 and a minor allele frequency threshold of 0.04. The tSNPs were optimized for the Illumina platform by preferentially selecting the SNP with the highest genotyping success rate in each block as a tag and reevaluating the  $r^2$  in the sample. Additionally, two SNPs, *APOE*: rs429358 (exon 4), rs7412 (exon 4), which detect the *E2/3/4* protein isoforms were genotyped by polymerase chain reaction (PCR) and restriction enzyme digestion as described previously (30,31).

### Statistical Analysis

Hardy-Weinberg equilibrium (HWE) was assessed with a chi-square test. Pairwise linkage-disequilibrium (LD) between the SNPs was calculated from the genotype data measured as  $D'$  and  $r^2$ . Haplotype frequencies were calculated using the PHASE algorithm (32–34). Missing SNP genotypes were imputed by using the haplotypic reconstruction (35). As for both baseline (measured at the start of the study) and “average” (the mean of all the available annual measurements for each individual in the five-year follow-up period), lipids levels were used (25,26). The lipid concentrations were transformed logarithmically to reduce skewness in their distributions. The Student  $t$  test and chi-square test were used to compare a number of phenotypes between those with and those without a CHD event. General linear model analysis was performed to identify associations between the SNPs or haplotypes and plasma lipid levels using indicator variables.

The common homozygote genotype of each SNP and the most frequent haplotype were used as a reference. Stepwise regression analysis with Akaike's information criterion included all SNPs genotyped and was performed to identify the independent determinants of plasma lipid levels and those with the strongest contribution to overall variance under an additive model. All linear models included adjustment for age and recruiting center (clinics). Since associations were formally examined for 10 SNPs and five lipid traits, some adjustment of the significance level is needed to account for multiple comparisons. However, there are high correlations between pairs of the phenotypes, and many of the SNPs also showed a degree of LD, therefore the Bonferroni correction would be too conservative (36–38). In addition, type I errors are random, and patterns in results that confirm previous reports should be given more weight than isolated results with a single low *P* value (36,38). Also, correction for multiple comparisons largely increases the likelihood of type II errors and important differences are considered nonsignificant (38). To take this into account, for individual tests of association, rather than applying a correction for multiple testing at global significance level of 5% ( $P < 0.05$ ), we defined statistical significance as less than 1% ( $P < 0.01$ ). Cox proportional hazards models were used to estimate hazard ratios (HRs) and 95% confidence intervals (95% CI) describing the association between SNPs, haplotypes and risk of CHD events.

All supplementary materials are available online at [www.molmed.org](http://www.molmed.org).

## RESULTS

A total of 2,767 men were included in the analysis. The mean ( $\pm$  SD) age was  $56.1 \pm 3.44$  and the mean ( $\pm$  SD) follow-up time was  $12.2 \pm 3.53$  years. During the follow-up period (overall 33,271 person years), 275 subjects experienced a CHD event. Table 1 shows the baseline charac-

**Table 1.** Mean baseline characteristics (SD) of event-free subjects and those who developed CHD over 15 years.

	Non-CHD (n <sup>a</sup> = 2,492)	CHD (n = 275)	<i>P</i> value
Age (years)	56.0 (3.42)	56.6 (3.52)	0.007
BMI (kg/m <sup>2</sup> )	26.4 (3.49)	26.9 (3.36)	0.043
SBP (mmHg)	138 (19)	142 (19)	<0.001
DBP (mmHg)	82 (11)	84 (11)	<0.001
Current smokers (%)	27.1	37.5	<0.001
Diabetes mellitus (%)	2.1	7.3	<0.001
Lipoprotein and Apolipoprotein			
TC (mmol/L)	5.70 (1.01)	6.07 (1.03)	<0.001
LDL-C (mmol/L)	3.06 (1.01)	3.37 (0.98)	<0.001
apoB (g/L)	0.89 (0.26)	0.95 (0.23)	<0.001
TG (mmol/L)	2.05 (1.25)	2.35 (1.35)	<0.001
HDL-C (mmol/L)	1.72 (0.59)	1.59 (0.55)	0.003
apoAI (g/L)	1.64 (0.32)	1.60 (0.29)	0.148

<sup>a</sup>n, Sample size.

teristic of the subjects who remained event free and those who developed CHD over 15 years. CHD patients were older and characterized by higher body mass index (BMI); systolic blood pressure (SBP) and diastolic blood pressure (DBP); concentration of TC, TG, LDL-C and apoB; and lower HDL-C. In addition, the CHD group included a higher proportion of smokers and diabetics than the non-CHD group.

### Allele Frequencies and Pairwise LD Structure at the APOE/C1/C4/C2 Gene Cluster

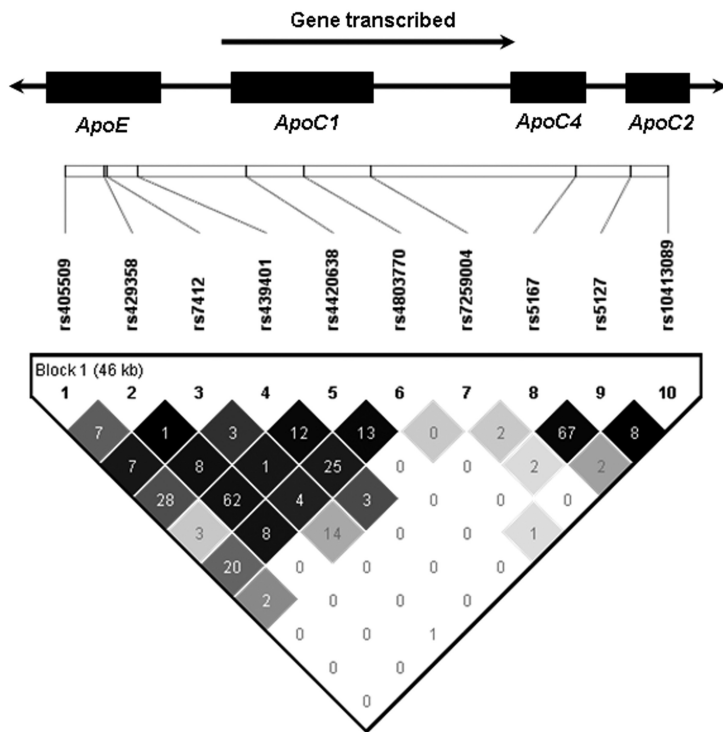
All samples were genotyped for 11 SNPs, 9 using the Illumina platform and 2 by PCR and restriction enzyme digestion. The frequency of the rare allele for the rs12691089 SNP was 0.002%, so it was not considered further. All polymorphisms genotyped were in Hardy–Weinberg equilibrium. The LD structure, expressed as  $D'$  and  $r^2$ , is shown in Figure 1. Although most of the SNPs throughout the APOE/C1/C4/C2 gene cluster did not show statistically significant evidence of LD, weak LD ( $D' \leq 0.05$  and  $r^2 \leq 0.05$ ) was detected between SNPs rs7259004, rs429358, and rs4420638, SNPs rs5167, rs405509, and rs439401; SNPs rs5127, rs4420638 and rs4803770; SNPs rs10413089, rs429358, rs439401, and rs4420638. Moderate LD ( $D' > 0.50$  and  $r^2 \geq 0.50$ ) was detected be-

tween SNPs rs4420638 and rs429358, and SNPs rs5127 and rs5167.

### Association between APOE/C1/C4/C2 Gene SNPs and Plasma Lipid Levels

The genotype frequencies and univariate analysis for each of the 10 SNPs with plasma lipid traits are presented in Table 2. As expected, compared with APOE  $\epsilon 3$  homozygotes, carriage of the APOE  $\epsilon 4$  allele was associated with higher plasma LDL-C; apoB and TG concentration and of the  $\epsilon 2$  allele with lower LDL-C and apoB (but higher TG). The rare alleles of the SNPs rs405509, rs4420638 and rs4803770 were associated with higher plasma LDL-C, and apoB concentration compared with the noncarriers, while the rare alleles of rs7259004 were associated with lower plasma LDL-C, apoB concentration and higher apoAI concentration compared with noncarriers. Carriers of the rare allele of rs4803770 had lower plasma HDL-C concentration, and carriers of rs5127 and rs5167 had higher plasma HDL-C and apoAI concentration.

To determine the independent effect of the cluster SNPs, stepwise analysis with the Akaike's information criterion was carried out as shown in Table 3. For all traits, the presence of APOE polymorphisms was the main predictor of plasma lipid concentrations after adjustment for age, recruiting center, BMI and smoking.



**Figure 1.** Pairwise linkage disequilibrium structure represent as  $D'$  (different color intensities) and  $r^2$  values (numbers) in 10 SNPs. The rs numbers and the relative physical distance between the SNPs are shown above (genes are the larger rectangular boxes). The color gradient indicates relative level of LD from black complete to white no LD.

Smoking had a modest effect on all traits (0.1–0.6% of the residual variance), with significant effects on apoB, TG, HDL-C and apoAI. The common *APOE* polymorphisms explained 3.4% of the residual variance in LDL-C and apoB concentration, and 0.1% and 0.4% of the residual variance in TG, HDL-C and apoAI concentration, respectively. By comparison, the other genotypes explained an additional 1.2% for LDL-C, 1.3% for apoB, 0.05% for TG, but an additional 1.4% for HDL-C and 1.2% for apoAI plasma concentration.

**Association between APOE/C1/C4/C2 Gene Cluster Haplotypes and Plasma Lipid Levels**

Initially, haplotypes were inferred using all 10 SNPs as shown in supplementary Table S2. Since each gene in the cluster might be characterized by a different haplotype, in turn associated with a phenotype linked to the gene function,

the haplotypes were re-estimated for each phenotype using only the SNPs which showed statistically significant association with the trait in the univariate analysis.

SNPs rs429358, rs7412 (*APOE*  $\epsilon 2/\epsilon 3/\epsilon 4$ ), rs439401 (*APOE*), rs4420638 and rs4803770 (*APOC1*), and rs7259004 (between *APOC1* and *APOC4*) were included in the haplotype estimation for the LDL-C and apoB traits. The ten inferred haplotypes that had a relative frequency of more than 1% (which included 97.3% of the inferred haplotypes), were tested for their association with LDL-C and the results are shown in Table 4. The single haplotype bearing the *APOE*  $\epsilon 4$  variant (A3 frequency 10.9%) was associated with higher levels of LDL-C and apoB and the two haplotypes bearing the *APOE*  $\epsilon 2$  variants frequency 4.4% and 4.0% (designated A5 and A7) were significantly associated with lower levels of LDL-C and apoB. No other haplotypes

were significantly associated with effects on these or other traits. The haplotypes explained 2.2% ( $P < 0.001$ ) and 2.3% ( $P < 0.001$ ) of the residual variance in LDL-C and apoB concentration, respectively.

For the HDL-C and apoAI traits, SNPs rs429358, rs7412 (*APOE*  $\epsilon 2/\epsilon 3/\epsilon 4$ ), rs4803770 and rs7259004 (*APOC1*), and rs5167 (*APOC2*) were included in the haplotype estimation. Eight haplotypes with a relative frequency of more than 2% (which included 92.6% of the inferred haplotypes), were analyzed. Compared to the most common haplotype, two haplotypes (B3 and B8), which both carry the rare allele of rs5167 (frequency 11.8% and 2.8%, respectively), were associated with significantly higher HDL-C and apoAI concentration (Table 4). Compared to noncarriers, carriers of this haplotype had, respectively, 4.0%, 6.6% higher levels of HDL-C and 2.4%, 4.1% higher levels of apoAI. Conversely, haplotype B4 (an  $\epsilon 4$ -containing haplotype) was associated with lower HDL-C and apoAI concentration. Compared to noncarriers, carriers of this haplotype had, respectively, 4.3% and 3.2% lower levels of HDL-C and apoAI. The haplotypes explained 0.8% ( $P < 0.001$ ) of the residual variance in HDL-C and apoAI concentration.

**Association between APOE/C1/C4/C2 Cluster SNPs and Haplotypes with CHD**

The association between the *APOE/C1/C4/C2* SNPs and risk of CHD is presented in Table 5. Compared to the *APOE*  $\epsilon 3/\epsilon 3$  genotype, carriers of the *APOE*  $\epsilon 2$  allele had lower risk of CHD with a HR of 0.63 (95% CI: 0.42–0.95) after adjustment for nonlipid classical risk factors, an effect unchanged after adjustment for lipid risk factors (HR = 0.65, 95% CI: 0.43–0.98). In addition there was evidence of a protective effect associated with the rs5127 SNP; where compared with noncarriers, subjects with one or more rare allele had a HR of 0.71 (95% CI: 0.55–0.91), an effect which was relatively unchanged after adjustment for nonlipid and lipid risk factors (HR = 0.72, 95% CI: 0.56–0.93). By contrast, carriers of the



**Table 2.** Univariate analysis between *APOE/C1/C2/C4* genotypes and lipid traits.

		LDL-C		apoB		Average <sup>a</sup> TG		HDL-C		ApoAI	
		n <sup>b</sup>	Mean (SE) <sup>c</sup>	n	Mean (SE)	n	Mean (SE)	n	Mean (SE)	n	Mean (SE)
<i>ApoE</i>	2+ <sup>d</sup>	368	2.69 (0.05)	368	0.77 (0.01)	429	1.91 (0.02)	365	1.71 (0.03)	368	1.65 (0.02)
	33	1,356	3.13 (0.03)	1,358	0.87 (0.01)	1,580	1.81 (0.01)	1,347	1.72 (0.02)	1,358	1.64 (0.01)
	4+ <sup>e</sup>	581	3.26 (0.04) <sup>f</sup>	582	0.90 (0.01) <sup>f</sup>	696	1.88 (0.02) <sup>g</sup>	579	1.66 (0.02)	582	1.61 (0.01)
rs405509	11	622	3.00 (0.04)	623	0.84 (0.01)	724	1.88 (0.02)	614	1.71 (0.02)	623	1.64 (0.01)
	12	1,242	3.09 (0.03)	1,244	0.86 (0.01)	1,471	1.83 (0.01)	1,235	1.70 (0.02)	1,244	1.63 (0.01)
	22	492	3.21 (0.04) <sup>f</sup>	492	0.89 (0.01) <sup>f</sup>	567	1.83 (0.02)	491	1.70 (0.03)	492	1.63 (0.01)
rs439401	11	998	3.11 (0.03)	1,000	0.86 (0.01)	1,176	1.88 (0.01)	988	1.68 (0.02)	1,000	1.62 (0.01)
	12	1,090	3.06 (0.03)	1,090	0.86 (0.01)	1,281	1.83 (0.01)	1,084	1.71 (0.02)	1,090	1.64 (0.01)
	22	268	3.16 (0.06)	269	0.88 (0.02)	305	1.77 (0.03)	268	1.76 (0.04)	269	1.66 (0.02)
rs4420638	11	1,575	3.03 (0.02)	1,576	0.85 (0.01)	1,834	1.82 (0.01)	1,565	1.72 (0.01)	1,576	1.64 (0.01)
	12	696	3.18 (0.04)	698	0.88 (0.01)	823	1.89 (0.02)	690	1.68 (0.02)	698	1.62 (0.01)
	22	85	3.56 (0.11) <sup>f</sup>	85	0.96 (0.03) <sup>f</sup>	105	1.94 (0.04)	85	1.66 (0.06)	85	1.61 (0.03)
rs4803770	11	835	3.06 (0.03)	836	0.85 (0.01)	963	1.84 (0.01)	832	1.75 (0.02)	836	1.65 (0.01)
	12	1,239	3.09 (0.03)	1,241	0.86 (0.01)	1,472	1.85 (0.01)	1,229	1.68 (0.02)	1,241	1.62 (0.01)
	22	282	3.23 (0.06) <sup>g</sup>	282	0.90 (0.02) <sup>f</sup>	327	1.84 (0.02)	279	1.68 (0.03) <sup>g</sup>	282	1.61 (0.02)
rs7259004	11	1,914	3.13 (0.02)	1,917	0.87 (0.01)	2,252	1.85 (0.01)	1,901	1.69 (0.01)	1,917	1.62 (0.01)
	12	420	2.96 (0.05)	420	0.82 (0.01)	485	1.81 (0.02)	417	1.78 (0.03)	420	1.66 (0.01)
	22	22	2.70 (0.21) <sup>f</sup>	22	0.83 (0.05) <sup>f</sup>	25	2.13 (0.09)	22	1.78 (0.12) <sup>g</sup>	22	1.75 (0.07) <sup>f</sup>
rs5167	11	1,147	3.13 (0.03)	1,149	0.87 (0.01)	1,376	1.87 (0.01)	1,137	1.65 (0.02)	1,149	1.60 (0.01)
	12	974	3.07 (0.03)	975	0.86 (0.01)	1,113	1.82 (0.01)	970	1.74 (0.02)	975	1.66 (0.01)
	22	235	3.02 (0.06)	235	0.83 (0.02)	273	1.84 (0.03)	233	1.79 (0.04) <sup>f</sup>	235	1.68 (0.02) <sup>f</sup>
rs5127	11	1,364	3.12 (0.03)	1,366	0.87 (0.01)	1,621	1.87 (0.01)	1,353	1.67 (0.02)	1,366	1.61 (0.01)
	12	860	3.05 (0.03)	861	0.85 (0.01)	981	1.80 (0.01)	857	1.75 (0.02)	861	1.66 (0.01)
	22	132	3.06 (0.09)	132	0.83 (0.02)	160	1.91 (0.03)	130	1.77 (0.05) <sup>f</sup>	132	1.67 (0.03) <sup>f</sup>
rs10413089	11	1,616	3.12 (0.02)	1,618	0.86 (0.01)	1,936	1.84 (0.01)	1,605	1.70 (0.01)	1,618	1.62 (0.01)
	12	660	3.04 (0.04)	661	0.86 (0.01)	735	1.85 (0.02)	657	1.73 (0.02)	661	1.65 (0.01)
	22	80	2.98 (0.11)	80	0.87 (0.03)	91	2.03 (0.05)	78	1.72 (0.07)	80	1.66 (0.03)

<sup>a</sup>Average of available annual measures during the study.<sup>b</sup>n, Sample size.<sup>c</sup>P value adjusted for age and clinic.<sup>d</sup>ε2 Data shown in light gray.<sup>e</sup>ε4 Data shown in dark gray.<sup>f</sup>P < 0.01.<sup>g</sup>P < 0.05.

rs4803770 rare allele had a higher risk of CHD (HR = 1.42, 95% CI: 1.09–1.85) compared with noncarriers, with a similar effect after adjustment (HR = 1.36, 95% CI: 1.04–1.78). There were no significant associations between any other SNPs and risk of CHD.

Haplotype estimation with CHD events was determined using the common *APOE* SNPs rs429358, rs7412, and rs4803770 (*APOC1*) and rs5127 (*APOC2*), present in Table 6. Eight haplotypes with a relative frequency of more than 1% (which included 99.0% of the inferred haplotypes), were identified. The *APOE* ε2 variant occurs on two haplotypes both associated with (nonsignificantly) lower

CHD risk. The rare allele of rs5127 occurs on four different haplotypes and in each case was associated with a lower CHD risk than the comparable haplotype lacking this allele (for example, C3 versus C2, C5 versus C4, C8 versus C6, and C7 versus C1), but because of the low frequency, none of these effects were statistically significant. The *APOE* ε4 variant occurs on two haplotypes, with the haplotype lacking the rare allele of rs5127 showing a nonsignificantly higher risk. The rs4803770 rare allele also occurs on two haplotypes with again the haplotype C2 lacking the common allele of rs5127 showing a significantly higher risk compared with the common C1 hap-

lotype. The consequence of this is that carriers and homozygotes of haplotype C2 show a significantly higher risk of CHD (HR = 1.37, 95% CI: 1.15–1.63, and HR = 1.63, 95% CI: 1.17–2.26) compared with noncarriers, and this estimate was relatively unchanged after adjustment for both nonlipid and lipid risk factors (HR = 1.29, 95% CI: 1.08–1.54, and HR = 1.54, 95% CI: 1.11–2.15). In addition, carriers and homozygotes of haplotype C4 show a significantly higher risk of CHD (HR = 1.32, 95% CI: 1.07–1.62) compared with noncarriers, and, once more, this estimation was relatively unchanged after adjustment for age, clinic, BMI, diabetes, total cholesterol, smoking and systolic

**APOE/C1/C4/C2 GENE CLUSTER AND PLASMA LIPID LEVELS**

**Table 3.** Multivariate regression with stepwise analysis with Akaike’s information criterion (AIC) of the independent contributions of genotype and covariates to lipid traits.

	LDL-C		apoB		TG		HDL-C		apoAI	
	Partial ( $R^2 \times 100$ )	P value	Partial ( $R^2 \times 100$ )	P value	Partial ( $R^2 \times 100$ )	P value	Partial ( $R^2 \times 100$ )	P value	Partial ( $R^2 \times 100$ )	P value
Age	0.00	0.993	0.00	0.798	0.00	0.336	0.00	0.584	0.00	0.584
Clinic	4.32	<0.001	16.16	<0.001	3.76	<0.001	4.32	<0.001	8.76	<0.001
BMI	0.86	<0.001	2.07	<0.001	10.11	<0.001	5.59	<0.001	1.87	<0.001
Smoking	0.01	0.633	0.23	0.010	0.67	<0.001	0.68	<0.001	0.31	<0.001
APOE	3.35	<0.001	3.42	<0.001	0.13	0.020	0.13	0.999	0.37	0.050
rs405509	—	—	—	—	—	—	—	—	0.08	0.111
rs439401	0.52	<0.001	0.62	0.013	—	—	—	—	—	—
rs4420638	0.27	0.009	0.21	0.011	0.05	0.198	—	—	—	—
rs4803770	0.23	0.016	0.22	<0.001	—	—	0.25	0.012	0.23	0.012
rs7259004	—	—	—	—	—	—	0.16	0.045	—	—
rs5167	0.19	0.030	—	—	—	—	0.78	<0.001	0.87	<0.001
rs5127	—	—	0.21	0.013	—	—	—	—	—	—
rs10413089	—	—	—	—	—	—	—	—	0.18	0.193
Genotype $R^2$	4.56		4.68		0.18		1.32		1.56	
Total $R^2$	9.08		22.57		14.31		11.29		12.14	

**Table 4.** Common haplotypes estimation using different SNPs from the APOE/C1/C2/C4 gene cluster and their association with LDL-C, apoB, HDL-C, and apoAI.

Haplotype	rs405509	rs429358	rs7412	rs439401	rs4420638	rs4803770	rs7259004	rs5167	rs5127	rs10413089	Freq. <sup>a</sup>	LDL-C	apoB
											(%)	Mean (SE)	Mean (SE)
A1 (ε3)	1 <sup>b</sup>	1	1	1	1	2 <sup>c</sup>	1	—	—	—	28.93	3.11 (0.03)	0.90 (0.01)
A2 (ε3)	2	1	1	2	1	1	1	—	—	—	27.11	3.09 (0.03)	0.89 (0.01)
A3 (ε4) <sup>d</sup>	2	2	1	1	2	1	1	—	—	—	10.93	3.27 (0.04) <sup>e</sup>	0.92 (0.01)
A4 (ε3)	1	1	1	2	1	1	1	—	—	—	4.81	3.07 (0.06)	0.89 (0.02)
A5 (ε2) <sup>f</sup>	1	1	2	1	1	1	2	—	—	—	4.43	2.67 (0.07) <sup>e</sup>	0.81 (0.02) <sup>e</sup>
A6 (ε3)	2	1	1	1	1	2	1	—	—	—	4.32	3.11 (0.07)	0.91 (0.02)
A7 (ε2) <sup>f</sup>	1	1	2	1	1	1	1	—	—	—	3.98	2.65 (0.07) <sup>e</sup>	0.79 (0.02) <sup>e</sup>
A8 (ε3)	1	1	1	1	2	1	1	—	—	—	3.05	3.23 (0.08)	0.93 (0.02)
A9 (ε3)	1	1	1	1	1	1	1	—	—	—	2.22	3.10 (0.10)	0.88 (0.02)
A10 (ε3)	1	1	1	1	1	2	2	—	—	—	2.20	3.09 (0.09)	0.87 (0.02)
												HDL-C	apoAI
											(%)	Mean (SE)	Mean (SE)
B1 (ε3)	—	1	1	—	—	1	1	1	—	—	28.21	1.70 (0.02)	1.63 (0.01)
B2 (ε3)	—	1	1	—	—	2	1	1	—	—	28.14	1.67 (0.02)	1.61 (0.01)
B3 (ε3)	—	1	1	—	—	1	1	2	—	—	11.82	1.77 (0.02) <sup>g</sup>	1.67 (0.01) <sup>g</sup>
B4 (ε4) <sup>d</sup>	—	2	1	—	—	1	1	1	—	—	6.49	1.63 (0.03) <sup>g</sup>	1.58 (0.02) <sup>e</sup>
B5 (ε4) <sup>d</sup>	—	2	1	—	—	1	1	2	—	—	5.60	1.68 (0.04)	1.63 (0.02)
B6 (ε3)	—	1	1	—	—	2	1	2	—	—	5.28	1.73 (0.04)	1.65 (0.02)
B7 (ε2) <sup>f</sup>	—	1	2	—	—	1	1	1	—	—	4.26	1.69 (0.04)	1.64 (0.02)
B8 (ε2) <sup>f</sup>	—	1	2	—	—	1	2	2	—	—	2.80	1.82 (0.05) <sup>g</sup>	1.70 (0.03) <sup>e</sup>

<sup>a</sup>Freq., frequencies.

<sup>b</sup>1, Present major allele.

<sup>c</sup>2, Present minor allele.

<sup>d</sup>ε4 Data shown in dark gray.

<sup>e</sup>Compared to common haplotype,  $P < 0.01$ .

<sup>f</sup>ε2 Data shown in light gray.

<sup>g</sup>Compared to common haplotype,  $P < 0.05$ .

**Table 5.** Cox proportional hazards models for the association between *APOE/C1/C4/C2* genotypes and risk for CHD event.

Genotype		Non-CHD, n (%)	CHD, n (%)	Adjusted model, <sup>a</sup> HR (95% CI)	Adjusted model, <sup>b</sup> HR (95% CI)	Adjusted model, <sup>c</sup> HR (95% CI)
<i>APOE</i>	33	1423 (58.2)	161 (60.8)	1 (Ref. <sup>d</sup> )	1 (Ref.)	1 (Ref.)
	2+ <sup>e</sup>	401 (16.4)	28 (10.6)	0.63 (0.42–0.95) <sup>f</sup>	0.62 (0.41–0.92) <sup>f</sup>	0.65 (0.43–0.98) <sup>f</sup>
	4+ <sup>g</sup>	621 (25.4)	76 (28.7)	1.02 (0.77–1.34)	1.02 (0.77–1.34)	1.03 (0.78–1.35)
rs405509	11	666 (26.7)	59 (21.5)	1 (Ref.)	1 (Ref.)	1 (Ref.)
	12	1318 (52.9)	157 (51.1)	1.28 (0.95–1.73)	1.32 (0.98–1.78)	1.29 (0.95–1.75)
	22	508 (20.4)	59 (21.5)	1.28 (0.89–1.84)	1.36 (0.95–1.96)	1.33 (0.92–1.92)
rs439401	11	1056 (42.4)	123 (44.7)	1 (Ref.)	1 (Ref.)	1 (Ref.)
	12	1162 (46.6)	121 (44.0)	0.92 (0.72–1.19)	0.95 (0.74–1.23)	0.97 (0.75–1.26)
	22	274 (11.0)	31 (11.3)	1.01 (0.68–1.50)	1.05 (0.71–1.56)	1.04 (0.70–1.54)
rs4420638	11	1660 (66.6)	178 (64.7)	1 (Ref.)	1 (Ref.)	1 (Ref.)
	12 + 22	832 (33.4)	97 (35.3)	1.02 (0.80–1.31)	1.01 (0.79–1.30)	0.99 (0.77–1.27)
rs4803770	11	887 (35.6)	76 (27.6)	1 (Ref.)	1 (Ref.)	1 (Ref.)
	12 + 22	1605 (64.4)	199 (72.4)	1.42 (1.09–1.85) <sup>h</sup>	1.38 (1.06–1.80) <sup>f</sup>	1.36 (1.04–1.78) <sup>f</sup>
rs7259004	11	2025 (81.3)	232 (84.4)	1 (Ref.)	1 (Ref.)	1 (Ref.)
	12 + 22	467 (18.7)	43 (15.6)	0.84 (0.60–1.16)	0.81 (0.59–1.13)	0.85 (0.61–1.19)
rs5167	11	1224 (49.1)	157 (57.1)	1 (Ref.)	1 (Ref.)	1 (Ref.)
	12	1017 (40.8)	96 (34.9)	0.76 (0.59–0.98)	0.76 (0.59–0.98)	0.78 (0.61–1.02)
	22	251 (10.1)	22 (8.0)	0.72 (0.46–1.13)	0.68 (0.44–1.07)	0.73 (0.47–1.15)
rs5127	11	1443 (57.9)	183 (66.6)	1 (Ref.)	1 (Ref.)	1 (Ref.)
	12 + 22	1049 (42.1)	92 (33.5)	0.71 (0.55–0.91) <sup>h</sup>	0.70 (0.54–0.90) <sup>h</sup>	0.72 (0.56–0.93) <sup>f</sup>
rs10413089	11	1755 (70.4)	186 (67.6)	1 (Ref.)	1 (Ref.)	1 (Ref.)
	12 + 22	734 (29.6)	89 (32.4)	1.17 (0.91–1.51)	1.15 (0.89–1.49)	1.15 (0.89–1.49)

<sup>a</sup>Adjusted for age and clinic.<sup>b</sup>Adjusted for age, clinic, BMI, diabetes, smoking and systolic blood pressure.<sup>c</sup>Adjusted for age, clinic, BMI, diabetes, smoking, systolic blood pressure and total cholesterol.<sup>d</sup>Ref., reference group.<sup>e</sup>ε2 Data shown in light gray.<sup>f</sup>*P* < 0.05.<sup>g</sup>ε4 Data shown in dark gray.<sup>h</sup>*P* < 0.01.

blood pressure (HR = 1.38, 95% CI: 1.12–1.70). The analysis was repeated using the THESIAS program and similar results were obtained (not shown).

## DISCUSSION

In this paper we have used a tagging SNPs (tSNPs) approach to capture essentially all of the common genetic variation at the *APOE/C1/C4/C2* gene cluster, allowing a comprehensive analysis of the genetic architecture of this locus and its relationship with plasma lipid levels and prospective CHD risk to be carried out. The major findings are that the *APOE* ε2/ε3/ε4 SNPs have the largest effect on LDL-C and apoB levels, with common SNPs elsewhere in the gene cluster increasing the proportion of sample variance explained by roughly 25%. By con-

trast, other SNPs, particularly one in *APOC4*, have a proportionally larger effect on the locus variance of HDL-C and apoA1 levels, explaining roughly 90% compared with roughly 10% for *APOE*. Finally, we present suggestive evidence that the well-known effects on CHD risk associated with *APOE* ε2/ε3/ε4 SNPs are modified by variants elsewhere in the cluster, particularly in *APOC1* and *APOC2*.

### SNPs, Haplotypes and Lipid Levels

Of the 10 SNPs examined in *APOE/C1/C4/C2* gene cluster, 5 SNPs, 2 in *APOE* and 3 in *APOC1*, had significant effects on plasma LDL-C and apoB levels. Only the common *APOE* polymorphisms were associated with modest but statistically significant effects on plasma

TG levels in univariate analysis, but four SNPs—two in *APOC1*, one in *APOC4* and one in *APOC2*—had significant effects on the plasma HDL-C, and three of these had significant effects on apoA1 levels. Multivariate analysis suggested that the strongest associations between common genetic variations in this gene cluster are with the *APOE* ε2 and ε4 alleles, with their effects mainly being to influence LDL-C and apoB concentration. *APOE* genotype was associated with about a threefold greater effect on LDL-C and apoB concentration than all other SNPs in the gene cluster.

To fully explore the information provided by the SNPs analysis, a haplotype analysis was performed. In general, the genotyped SNPs in the *APOE/C1/C4/C2* gene cluster did not show strong evi-

**Table 6.** Cox proportional hazards models between inferred common haplotypes, using different SNPs from the *APOE/C1/C2/C4* gene cluster, and CHD events. The reference level is noncarriers of the haplotypes.

Haplotype	rs429358	rs7412	rs4803770	rs5127	Non-CHD (%)	CHD (%)	Adjusted model, <sup>a</sup> HR (95% CI)	Adjusted model, <sup>b</sup> HR (95% CI)	Adjusted model, <sup>c</sup> HR (95% CI)
C1 (ε3)	1	1	1	1	36.46	34.36	0.91 (0.76–1.09)	0.93 (0.78–1.11)	0.92 (0.77–1.10)
C2 (ε3)	1	1	2	1	25.94	31.27	1.28 (1.07–1.53) <sup>d</sup>	1.25 (1.04–1.50) <sup>d</sup>	1.24 (1.03–1.49) <sup>d</sup>
C3 (ε3)	1	1	2	2	9.72	7.82	0.83 (0.61–1.14)	0.84 (0.61–1.15)	0.84 (0.62–1.15)
C4 (ε4) <sup>e</sup>	2	1	1	1	8.51	10.73	1.25 (0.95–1.64)	1.31 (1.00–1.72)	1.29 (0.99–1.70)
C5 (ε4) <sup>e</sup>	2	1	1	2	5.56	5.45	0.91 (0.63–1.32)	0.85 (0.59–1.24)	0.82 (0.56–1.20)
C6 (ε2) <sup>f</sup>	1	2	1	1	5.06	4.36	0.85 (0.57–1.29)	0.86 (0.57–1.29)	0.85 (0.55–1.30)
C7 (ε3)	1	1	1	2	4.96	3.82	0.80 (0.52–1.24)	0.81 (0.53–1.26)	0.83 (0.54–1.29)
C8 (ε2) <sup>f</sup>	1	2	1	2	3.79	2.18	0.61 (0.35–1.09)	0.58 (0.33–1.03)	0.68 (0.38–1.21)
Only C2									
11 <sup>g</sup>					54.01	45.09	Ref.	Ref.	Ref.
12 <sup>h</sup>					40.33	47.27	1.37 (1.15–1.63) <sup>i</sup>	1.31 (1.10–1.56) <sup>i</sup>	1.29 (1.08–1.54) <sup>i</sup>
22 <sup>j</sup>					5.66	7.64	1.63 (1.17–2.26) <sup>i</sup>	1.64 (1.18–2.27) <sup>i</sup>	1.54 (1.11–2.15) <sup>i</sup>
Only C4									
11					83.71	78.91	Ref.	Ref.	Ref.
12 + 22					16.29	21.09	1.32 (1.07–1.62) <sup>i</sup>	1.39 (1.13–1.72) <sup>i</sup>	1.38 (1.12–1.70) <sup>i</sup>

<sup>a</sup>Adjusted for age, and clinic.<sup>b</sup>Adjusted for age, clinic, BMI, diabetes, smoking and systolic blood pressure.<sup>c</sup>Adjusted for age, clinic, BMI, diabetes, smoking, systolic blood pressure and total cholesterol.<sup>d</sup> $P < 0.05$ .<sup>e</sup>ε4 Data shown in dark gray.<sup>f</sup>ε2 Data shown in light gray.<sup>g</sup>11, Noncarrier of C2 or C4.<sup>h</sup>12, Heterozygote for C2 or C4.<sup>i</sup> $P < 0.01$ .<sup>j</sup>22, Homozygote for C2 or C4.

dence of LD, since, apart from the ε2/ε3/ε4 SNPs, they were chosen to be tagging SNPs. By use of all 10 SNPs, more than 28 haplotypes were observed with a frequency of greater than 0.5%, but, of these, only some of the SNPs were required to explain the genetic architecture at this locus. Ten haplotypes using seven SNPs were inferred with a frequency greater than 0.5%, representing 92.0% of the subjects, for the analysis of the LDL-C and apoB phenotypes. Compared to the common haplotype H1, only haplotypes with *APOE* ε2 and ε4 alleles had significant effects on the plasma lipid levels. For the HDL-C and apoAI phenotypes, eight haplotypes were inferred with a frequency greater than 0.5%, representing 92.6% of the subjects. Compared to the common haplotype H1, three haplotypes had a significant independent association with HDL-C and apoAI concentration, with one ε4-carrying haplotype having lower levels, and one

common ε3-carrying and one rare ε2-carrying haplotype having higher levels and with both of these haplotypes including the rare allele of the *APOC4* SNP rs5167.

It is possible that these more rare haplotype effects are due to chance, and confirmation of these effects in larger samples is required. This SNP has been associated previously with elevated TG levels in women (but not men) (23) and this gender difference may explain the lack of association with TG levels in the all male NPHSII samples. To attempt to address this in part, we have carried out an *in silico* replication of this SNP effect by analyzing the association in a previously published study (39) with a focus on 3,403 middle-aged (39–63 years) men from the Whitehall II study (WH-II). As shown in supplementary Table S4, the minor allele was associated with significantly higher levels of HDL-C (+2.1%) and apoAI (+1.1%) concentration, with no significant

effect on LDL-C, apoB or TG levels. There were no significant effects of this SNP with any lipid traits in the smaller group of women (n = 1,202). It thus appears that the modest HDL and ApoAI raising effect associated with this SNP is of interest.

These data confirm previous studies that have shown that the *APOE* ε2 and ε4 alleles account for a significant proportion of lipid and apolipoprotein variability, with estimates between 3% and 6% of the LDL-C variation (40,41). While other studies (28,42,43) also have identified that there are other important variants in the *APOE* gene or in other nearby genes that influence variation in LDL-C, the analysis here indicates that the effects are small at best. With a sample of 2,767, the study has the power ( $\alpha = 0.05$ ,  $\beta = 0.80$ ) to detect an effect on LDL-C of 0.15, 0.12, 0.10 mmol/L with SNPs with a MAF of 10%, 20% or 30%, respectively. However while the ε2/ε3/ε4 SNPs are having a small effect on HDL-C and apoAI levels,



a single SNP in *APOC4*, rs5167, is having a significant effect on these traits.

### Mechanism of SNP Effects

The mechanism of the effect of the *APOE*  $\epsilon 2/\epsilon 3/\epsilon 4$  SNPs on LDL-C and apoB levels is well documented and is due to the impact of the resulting amino acid changes causing changes in the affinity of the lipoprotein particles which carry these “ligand” apoproteins for LDL receptor binding (11). The number of apoE molecules per particle is thought to be in the range of 2–4 (10,44), with *in vitro* data suggesting that two apoE molecules may act together to improve binding (44). The lipoprotein particles in the blood of an individual carrying  $\epsilon 2$  or  $\epsilon 4$  will therefore be a spectrum of “ $\epsilon 3$ -only-containing” and “ $\epsilon 3$ -and- $\epsilon 2$ -containing” or “ $\epsilon 3$ -and- $\epsilon 4$ -containing” particles, but a particle with two  $\epsilon 3$  molecules and one  $\epsilon 2$  (or  $\epsilon 4$ ) may be able to be taken up as effectively as an “ $\epsilon 3$ -only” particle. Therefore, the amount of *APOE* gene transcription and thus mRNA and protein being made in the liver and secreted may also have an influence on lipoprotein metabolism. This is supported by studies that have examined the association between *APOE* promoter variants and plasma lipid levels, and, in particular, the rs405509–219 G > T variant (45,46). It has been shown that the –219T allele is associated with decreased transcriptional activity (45) and plasma apoE concentration (47,48), and, as expected, from the known role of apoE in lipoprotein clearance with moderately increased LDL-C concentrations in at least some studies (49). Several studies have also suggested –219T allele carriers have an increased risk of myocardial infarction (47) and premature CHD (50) after adjustment for lipid traits. In this study, the rare allele of rs405509 (that is, –219T) was associated with significantly higher LDL-C and apoB levels in univariate analysis but not in multivariate analysis. By inspection of the haplotype data, the rare allele occurs on three independent haplotypes and only on the  $\epsilon 4$ -carrying haplotype is it associated with a significant LDL and apoB

raising effect. This suggests that the effect, at least on lipid traits may not be statistically independent.

The other relevant potentially functional SNP s5167 is a T > G missense SNP located in exon 3 of gene *APOC4* which changes Leucine at position 96 to Arginine. In both this and the WHII study, the 96R allele was associated with significantly higher levels of HDL-C and apoAI concentration. While this SNP may be in LD with a functional SNP elsewhere in the gene cluster, for example, determining expression of one or more of the apoproteins, the amino acid change from a neutral to a charged amino acid is likely to affect function. The computer prediction program POLYPHEN predicts that this change is likely to be damaging. As shown in Supplementary Figure S1, the amino acid at position 96 is an arginine in mouse, rat, rabbit and cow, and thus the rare allele restores the common sequence. It may thus be that the amino acid-change results in an altered relative affinity of the apoCIV for VLDL versus HDL particles, which affects the rate of lipolysis of VLDL by LPL. However the SNP was not significantly associated with effects on TG levels in either study, so it may be influencing HDL and apoAI levels by a more direct mechanism. Since at this time the exact role of apoCIV in lipoprotein metabolism is unclear, further speculation is unwarranted.

### SNPs, Haplotypes and CHD Risk

We confirm here the well-reported effects of the common *APOE* variants on CHD risk (14), but apart from the reported higher risk associated with rs405509 (47,49), there is little data for other SNPs in the *APOE/C1/C4/C2* gene cluster or for the gene cluster haplotypes. The *APOE*  $\epsilon 2$  allele and the *APOC2* SNP rs5127 were both independently associated with roughly 30% lower prospective CHD risk, effects which remained after adjustment for all other risk factors. By contrast, the *APOC1* SNP rs4803770 was associated with 40% higher prospective CHD risk after adjustment for all other risk factors. None of the other SNPs had

a significant effect on prospective CHD risk. We were not able to confirm previous reports of an association between rs405509 and higher risk of CHD, but the study is underpowered to detect the reported effect of 1.29 (47).

We used haplotype analysis to explore whether the effect of the common *APOE* genotypes were modified by other SNPs. The *APOE*  $\epsilon 2$  allele was found on two haplotypes, one of which also carried the *APOC2* rs5127 rare allele, and both of which were associated with low CHD risk. There was suggestive evidence that the protective effect of  $\epsilon 2$  was enhanced by the *APOC2* rs5127 rare allele, with this haplotype being associated with the lowest risk, although its rarity means the effect was not statistically significant. The *APOE*  $\epsilon 3$  allele occurred on four common haplotypes and one of these, haplotype C2, which also carries the minor allele of rs4803770 and the major allele of rs5127, was associated with a significantly higher (25%) prospective CHD risk after adjustment for all other risk factors. The consequence of this was that individuals who were homozygous for this haplotype, who represented 5.7% of these subjects, had a 63% higher risk while individuals heterozygote for C2 (40% of the subjects) had ~37% higher risk compared with noncarriers of C2, with these effects being only slightly attenuated by adjustment for CHD risk factors. The *APOE*  $\epsilon 4$  allele was found on two haplotypes, but, interestingly, only one of these, haplotype C4, which also carried the risk alleles of rs4803770 and rs5127, was associated with higher risk of CHD. The combined group of homozygotes and those heterozygote for the C4 haplotype had a 38% higher risk compared with noncarriers of C4. Although for all these haplotypes, the risk effect could be at least partly due to the associated effects on plasma lipid levels; based on statistical arguments, this seems unlikely.

The mechanism of the SNP and haplotype effects in the *APOE/C1/C4/C2* gene cluster with CHD risk effect is unclear. The SNP 5127 is a G > T change located in the 3'UTR of *APOC2* and, by inspec-

tion, the T allele creates a new NKkB4 binding site. Whether or not this is functional is currently unknown. The SNP rs4803770 is located 4.5 kb downstream of *APOC1* (3' end), and the minor allele was associated with a significant raising effect of ~5.3% on LDL-C and apoB concentrations, but its impact on CHD was maintained after adjustment for lipid levels. *In silico*, this variant does not occur within a recognized transcription factor binding site. For both of these sites, it is likely that they are markers by LD for functional sites affecting transcription of one or more of the apoproteins in the cluster.

### Limitations

A major limitation of this study is the relatively small number of CHD events, and, clearly, the association of SNPs and haplotypes with CHD risk needs to be interpreted with caution before confirmation in larger studies. However, the prospective nature of the NPHSII study reduces the possibility of the result being confounded by population stratification (51). The study also would be strengthened by measures of plasma apoE and of the other apoproteins in the cluster, since this would allow determination of to what extent the effects seen on lipid traits and CHD risk are being mediated through differences in levels of circulating proteins. As well as having a role in lipoprotein metabolism, apoE has antioxidant properties (47), and measures of antioxidant potential may also be of value.

### ACKNOWLEDGMENTS

We acknowledge the contribution of the late Professor George Miller (1939–2006) who was the PI on the NPHSII study. We thank Sarah Leigh for providing bioinformatics support. The British Heart Foundation support FD and SEH (RG2008/008). The NPHSII study was supported by the Medical Research Council, the US National Institutes of Health (NHLBI 33014) and DuPont Pharma. We also thank all the medical staff and patients who contributed to the

NPHSII study and the Office for National Statistics (NHS) Central Registry for provision of mortality data.

### DISCLOSURE

The authors declare that they have no competing interests as defined by *Molecular Medicine*, or other interests that might be perceived to influence the results and discussion reported in this paper.

### REFERENCES

- Austin MA, Hokanson JE, Edwards KL. (1998) Hypertriglyceridemia as a cardiovascular risk factor. *Am. J. Cardiol.* 81:7B–12B.
- Castelli WP, Anderson K, Wilson PW, Levy D. (1992) Lipids and risk of coronary heart disease. The Framingham Study. *Ann. Epidemiol.* 2:23–28.
- Brunzell JD, Sniderman AD, Albers JJ, Kwiterovich PO Jr. (1984) Apoproteins B and A-I and coronary artery disease in humans. *Arteriosclerosis.* 4:79–83.
- Lusis AJ. (2000) Atherosclerosis. *Nature.* 407:233–41.
- Sing CF, Stengard JH, Kardia SL. (2003) Genes, environment, and cardiovascular disease. *Arterioscler. Thromb. Vasc. Biol.* 23:1190–6.
- Lusis AJ, Mar R, Pajukanta P. (2004) Genetics of atherosclerosis. *Annu. Rev. Genomics Hum. Genet.* 5:189–218.
- Beekman M, et al. (2002) Heritabilities of apolipoprotein and lipid levels in three countries. *Twin Res.* 5:87–97.
- Ehnholm C, Lukka M, Kuusi T, Nikkila E, Utermann G. (1986) Apolipoprotein E polymorphism in the Finnish population: gene frequencies and relation to lipoprotein concentrations. *J. Lipid Res.* 27:227–35.
- Dallongeville J, Lussier-Cacan S, Davignon J. (1992) Modulation of plasma triglyceride levels by apoE phenotype: a meta-analysis. *J. Lipid Res.* 33:447–54.
- Strittmatter WJ, Bova Hill C. (2002) Molecular biology of apolipoprotein E. *Curr. Opin. Lipidol.* 13:119–123.
- Eichner JE, et al. (2002) Apolipoprotein E polymorphism and cardiovascular disease: a HuGE review. *Am. J. Epidemiol.* 155:487–95.
- Boerwinkle E, et al. (1987) The use of measured genotype information in the analysis of quantitative phenotypes in man. II. The role of the apolipoprotein E polymorphism in determining levels, variability, and covariability of cholesterol, betalipoprotein, and triglycerides in a sample of unrelated individuals. *Am. J. Med. Genet.* 27:567–82.
- Siest G, et al. (1995) Apolipoprotein E: an important gene and protein to follow in laboratory medicine. *Clin. Chem.* 41:1068–86.
- Bennet AM, et al. (2007) Association of apolipoprotein E genotypes with lipid levels and coronary risk. *JAMA.* 298:1300–11.
- Jong MC, Hofker MH, Havekes LM. (1999) Role of ApoCs in lipoprotein metabolism: functional differences between ApoC1, ApoC2, and ApoC3. *Arterioscler. Thromb. Vasc. Biol.* 19:472–84.
- Weisgraber KH, et al. (1990) Apolipoprotein C-I modulates the interaction of apolipoprotein E with beta-migrating very low density lipoproteins (beta-VLDL) and inhibits binding of beta-VLDL to low density lipoprotein receptor-related protein. *J. Biol. Chem.* 265:22453–9.
- Sehayek E, Eisenberg S. (1991) Mechanisms of inhibition by apolipoprotein C of apolipoprotein E-dependent cellular metabolism of human triglyceride-rich lipoproteins through the low density lipoprotein receptor pathway. *J. Biol. Chem.* 266:18259–67.
- Pentikainen MO, Oksjoki R, Oorni K, Kovanen PT. (2002) Lipoprotein lipase in the arterial wall: linking LDL to the arterial extracellular matrix and much more. *Arterioscler. Thromb. Vasc. Biol.* 22:211–7.
- Breckenridge WC, Little JA, Steiner G, Chow A, Poapst M. (1978) Hypertriglyceridemia associated with deficiency of apolipoprotein C-II. *N. Engl. J. Med.* 298:1265–73.
- Shen Y, Lookene A, Nilsson S, Olivecrona G. (2002) Functional analyses of human apolipoprotein CII by site-directed mutagenesis: identification of residues important for activation of lipoprotein lipase. *J. Biol. Chem.* 277:4334–42.
- Allan CM, Taylor S, Taylor JM. (1997) Two hepatic enhancers, HCR.1 and HCR.2, coordinate the liver expression of the entire human apolipoprotein E/C-I/C-IV/C-II gene cluster. *J. Biol. Chem.* 272:29113–9.
- Kotite L, Zhang LH, Yu Z, Burlingame AL, Havel RJ. (2003) Human apoC-IV: isolation, characterization, and immunochemical quantification in plasma and plasma lipoproteins. *J. Lipid Res.* 44:1387–94.
- Kamboh MI, Aston CE, Hamman RF. (2000) DNA sequence variation in human apolipoprotein C4 gene and its effect on plasma lipid profile. *Atherosclerosis.* 152:193–201.
- Cooper JA, et al. (2000) Comparison of novel hemostatic factors and conventional risk factors for prediction of coronary heart disease. *Circulation* 102:2816–2822.
- Miller GJ, et al. (1995) The effects of quality and timing of venepuncture on markers of blood coagulation in healthy middle-aged men. *Thromb. Haemost.* 73:82–6.
- Miller GJ, Bauer KA, Barzegar S, Cooper JA, Rosenberg RD. (1996) Increased activation of the haemostatic system in men at high risk of fatal coronary heart disease. *Thromb. Haemost.* 75:767–71.
- World Health Organization Regional Office for Europe. (1976) *Myocardial Infarction Community Register*. Copenhagen: World Health Organization. Public Health in Europe 5.
- Drenos F, et al. (2009) Integrated associations of genotypes with multiple blood biomarkers

- linked to coronary heart disease risk. *Hum. Mol. Genet.* 18:2305–16.
29. de Bakker PI, *et al.* (2005) Efficiency and power in genetic association studies. *Nat. Genet.* 37:1217–23.
  30. Bolla MK, Haddad L, Humphries SE, Winder AF, Day IN. (1995) High-throughput method for determination of apolipoprotein E genotypes with use of restriction digestion analysis by microplate array diagonal gel electrophoresis. *Clin. Chem.* 41:1599–604.
  31. Day IN, Humphries SE. (1994) Electrophoresis for genotyping: microtiter array diagonal gel electrophoresis on horizontal polyacrylamide gels, hydrolink, or agarose. *Anal. Biochem.* 222:389–95.
  32. Stephens M, Scheet P. (2005) Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation. *Am. J. Hum. Genet.* 76:449–62.
  33. Stephens M, Smith NJ, Donnelly P. (2001) A new statistical method for haplotype reconstruction from population data. *Am. J. Hum. Genet.* 68:978–89.
  34. Graham RR, *et al.* (2001) Genetic linkage and transmission disequilibrium of marker haplotypes at chromosome 1q41 in human systemic lupus erythematosus. *Arthritis Res* 3:299–305.
  35. Marchini J, *et al.* (2006) A comparison of phasing algorithms for trios and unrelated individuals. *Am. J. Hum. Genet.* 78:437–50.
  36. Rothman KJ. (1990) No adjustments are needed for multiple comparisons. *Epidemiology* 1:43–6.
  37. Eichenbaum-Voline S, *et al.* (2004) Linkage and association between distinct variants of the APOA1/C3/A4/A5 gene cluster and familial combined hyperlipidemia. *Arterioscler. Thromb. Vasc. Biol.* 24:167–74.
  38. Lai CQ, *et al.* (2004) Influence of the APOA5 locus on plasma triglyceride, lipoprotein subclasses, and CVD risk in the Framingham Heart Study. *J. Lipid Res.* 45:2096–105.
  39. Talmud PJ, *et al.* (2009) Gene-centric association signals for lipids and apolipoproteins identified via the HumanCVD BeadChip. *Am. J. Hum. Genet.* 85:628–42.
  40. Isaacs A, *et al.* (2007) Heritabilities, apolipoprotein E, and effects of inbreeding on plasma lipids in a genetically isolated population: the Erasmus Rucphen Family Study. *Eur. J. Epidemiol.* 22:99–105.
  41. Klos K, *et al.* (2008) APOE/C1/C4/C2 hepatic control region polymorphism influences plasma apoE and LDL cholesterol levels. *Hum. Mol. Genet.* 17:2039–46.
  42. North KE, *et al.* (2006) Linkage analysis of LDL cholesterol in American Indian populations: the Strong Heart Family Study. *J. Lipid Res.* 47:59–66.
  43. Willer CJ, *et al.* (2008) Newly identified loci that influence lipid concentrations and risk of coronary artery disease. *Nat. Genet.* 40:161–9.
  44. Curtiss LK. (2000) ApoE in atherosclerosis : a protein with multiple hats. *Arterioscler. Thromb. Vasc. Biol.* 20:1852–3.
  45. Artiga MJ, *et al.* (1998) Allelic polymorphisms in the transcriptional regulatory region of apolipoprotein E gene. *FEBS Lett.* 421:105–8.
  46. Lambert JC, *et al.* (1998) A new polymorphism in the APOE promoter associated with risk of developing Alzheimer's disease. *Hum. Mol. Genet.* 7:533–40.
  47. Lambert JC, *et al.* (2000) Independent association of an APOE gene promoter polymorphism with increased risk of myocardial infarction and decreased APOE plasma concentrations—the ECTIM study. *Hum. Mol. Genet.* 9:57–61.
  48. Moreno JA, *et al.* (2003) The influence of the apolipoprotein E gene promoter (–219G/ T) polymorphism on postprandial lipoprotein metabolism in young normolipemic males. *J. Lipid Res.* 44:2059–64.
  49. Moreno JA, *et al.* (2005) The apolipoprotein E gene promoter (–219G/T) polymorphism determines insulin sensitivity in response to dietary fat in healthy young adults. *J. Nutr.* 135:2535–40.
  50. Viitanen L, *et al.* (2001) Apolipoprotein E gene promoter (–219G/T) polymorphism is associated with premature coronary heart disease. *J. Mol. Med.* 79:732–7.
  51. Chen XH, *et al.* (2004) Evidence of admixture from haplotyping in an epidemiological study of UK Caucasian males: implications for association analyses. *Hum. Hered.* 57:142–55.