

ORIGINAL ARTICLE

Genome- and exome-wide association study of serum lipoprotein (a) in the Jackson Heart Study

Jin Li¹, Leslie A Lange¹, Jeremy Sabourin², Qing Duan¹, William Valdar¹, Monte S Willis³, Yun Li¹, James G Wilson⁴ and Ethan M Lange¹

Lipoprotein (a) (Lp(a)) is an independent risk factor for cardiovascular disease. Lp(a) levels in African Americans (AAs) are much higher compared with that in European Americans. We conducted a genome- and an exome-wide association study of Lp(a) among 2895 AAs participating in the Jackson Heart Study. We observed that local ancestry at 6q25.3 was an important risk factor for Lp(a) in AAs, and that multiple single-nucleotide polymorphisms (SNPs) at the well-established *LPA* locus were significantly associated with Lp(a) ($P < 5 \times 10^{-8}$) after adjusting for the local ancestry at 6q25.3. Interestingly, before adjusting for local ancestry, we observed significant ($P < 5 \times 10^{-8}$) associations for hundreds of SNPs spanning ~10 Mb region on 6q surrounding the *LPA* gene, whereas after adjusting for local ancestry, the region containing significantly associated SNPs got much narrower and was centered over the *LPA* gene (<1 Mb). We observed a single nonsynonymous SNP in *APOE* significantly associated with Lp(a) ($P < 5 \times 10^{-8}$). A high burden of coding variants in *LPA* and *APOE* were also associated with higher Lp(a) levels. Our study provides evidence that ancestry-specific causal variant(s) resides in or near *LPA* and that most of the observed associations outside this narrower region are spurious associations.

Journal of Human Genetics (2015) 60, 755–761; doi:10.1038/jhg.2015.107; published online 17 September 2015

INTRODUCTION

Lipoprotein (a) (Lp(a)) is an independent risk factor for cardiovascular disease.^{1,2} Genetic variants in *LPA* are strongly associated with both an increased level of Lp(a) and an increased risk of coronary disease,³ suggesting a causal role of Lp(a) in coronary disease. Lp(a) is a low-density lipoprotein (LDL)-like particle that consists of an apolipoprotein(a) covalently linked to apolipoprotein B100 by a disulfide bond.⁴ Genetic factors have a large impact on the variation of Lp(a) levels and ~70–90% of the total variance of Lp(a) can be attributed to variation within the *LPA* locus across worldwide populations.^{5,6} A copy number variant, the Kringle(IV) type 2 domain, accounts for approximately half of the variance explained by the *LPA* locus.^{7,8} Recent genome-wide association studies (GWAS) in subjects of European descent have identified multiple polymorphisms spanning 12.5 Mb on chromosome 6q26–27, which includes *LPA*, that are significantly associated with Lp(a) levels independent of each other and of the Kringle IV size polymorphism in *LPA* ($P < 5 \times 10^{-8}$).⁷ A candidate gene study on multiethnic populations suggested both SNPs at 6q26–27 and the Kringle IV copy number variant were genomic determinants of Lp(a) level, and the proportion of total variance explained by each determinant differs across ethnic groups.⁹

Lp(a) levels in populations of African ancestry are much higher (2–4-fold) compared with that in populations of European ancestry.¹⁰ A genome-wide admixture study on a population of African

Americans (AAs) suggested that local ancestry at 6q25.3 was significantly associated with Lp(a) after adjustment for the Kringle IV copy number variant.¹¹ However, so far, no genome- or exome-wide association studies of common and uncommon-coding variants, respectively, have been conducted in populations of African ancestry to assess the importance of other genomic regions on Lp(a) levels.

Here, we present a genome- and an exome-wide association study of Lp(a) among AAs participating in the Jackson Heart Study (JHS). We observed numerous SNPs at the well-established *LPA* locus and a single SNP in *APOE* significantly associated with Lp(a) ($P < 5 \times 10^{-8}$). A high burden of coding variants in *LPA* and *APOE* were also associated with higher Lp(a) levels.

MATERIALS AND METHODS

Study subjects and phenotypes

This study included 1106 AA male and 1789 AA female participants with measured Lp(a) levels and available genome-wide genotype data from the JHS, a longitudinal, population-based cohort from Jackson, Mississippi, USA.¹² The design, recruitment and initial characterization of this study was described in detail elsewhere.¹³ Serum Lp(a) levels (mg dl⁻¹) were measured using a Diasorin nephelometric assay on a Roche Cobas FARA analyzer (Roche Diagnostics Corporation, Indianapolis, IN, USA).¹⁴ Fasting LDL, high-density lipoprotein (HDL), triglyceride (TG) and total cholesterol (TC) were measured as described previously.¹⁵ For each individual treated with lipid-lowering therapies, the observed lipid value was multiplied by a correction

¹Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA; ²Department of Statistics and Operations Research, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA; ³Department of Pathology and Laboratory Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA and ⁴Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, MS, USA
Correspondence: Dr EM Lange, Department of Genetics, University of North Carolina at Chapel Hill, 5111 Genetics Medicine Building, 120 Mason Farm Road, Chapel Hill, NC 27599-7064, USA.
E-mail: ethan_lange@med.unc.edu

factor (1.352 for LDL, 0.949 for HDL, 1.210 for TG and 1.271 for TC).¹⁶ The study protocol was approved by the University of Mississippi Medical Center Institutional Review Board, and written informed consent was obtained from all JHS participants. Descriptive characteristics of the JHS participants in this study were summarized in Table 1.

Genome-wide genotype data and genotype imputation

A total of 3030 JHS participants were genotyped using the Affymetrix Genome-Wide Human SNP Array 6.0 (Affymetrix, Santa Clara, CA, USA). Genotyping quality control was conducted using PLINK v.1.07.¹⁷ A total of 874 712 SNPs with a call rate >0.95, minor allele frequency (MAF) >0.01 and Hardy–Weinberg equilibrium P -value $>1 \times 10^{-6}$ were included in the genotype imputation target panel. Thirty-eight million SNPs were imputed, using MACH 2.0,¹⁸ based on a reference panel consisting of the complete sample of 1000 Genome Project participants (November 2010, Version 3); only SNPs with imputation quality of $r^2 > 0.3$ were included in further analysis. The Kringle IV copy number variant was not available in JHS participants.

Exome array

A total of 2790 JHS participants, including 2448 with Affymetrix 6.0 genotype data, were genotyped using the Illumina Human Exome Beadchip (version 12v1_rev5; Illumina Inc., San Diego, CA, USA), consisting of >200 000 putative functional variants selected from >12 000 individual exome and whole-genome sequences across diverse populations and a range of common complex traits.

Statistical analyses

Association of local ancestry estimates. Ethnic differences in Lp(a) levels have been observed between subjects of European and African descent, with some populations of AAs having, on average, almost fourfold higher Lp(a) levels compared with European Americans.¹⁰ To assess the impact of genetic admixture within our AA population on Lp(a) levels, we first estimated the average, or global, African ancestry proportion across the genome for all subjects using the software ADMIXTURE,¹⁹ assuming $K=2$ contributing populations. We then tested whether this estimated proportion was associated with natural-log-transformed Lp(a) levels after adjusting for age, sex and body mass index (BMI). Next, ANCESTRYMAP²⁰ was used to estimate local ancestry (probabilities of whether an individual has 0, 1 or 2 alleles of Caucasian ancestry) at 738 831 autosomal SNPs across the genome for each participant in JHS as described previously.¹⁵ In brief, local ancestry was inferred using hidden Markov models based on the genotypes from a panel of densely spaced markers with highly differential allele frequencies between African and European populations. We performed admixture mapping across the whole genome by regressing natural-log-transformed Lp(a) levels, adjusting for age, sex, BMI and global ancestry, on the local ancestry estimate at each SNP location. The LOD score for association, defined as the log base10 ratio of the likelihood of the data under a model including local ancestry divided by the likelihood of the data under the model excluding local ancestry, was computed at each of these local ancestry informative marker locations across the genome. For regions showing

association of increased African ancestry with higher levels of Lp(a), the LOD scores were assigned positive values, and for regions showing association of increased African ancestry with lower levels of Lp(a), the LOD scores were assigned negative values. LOD scores were plotted across the whole genome, and an LOD score of 5 was assumed to be the threshold of statistical significance.²⁰

Genome-wide association analysis. Lp(a) levels were naturally log-transformed to approximate normality of residuals after accounting for age, sex and BMI. The association between Lp(a) and imputed SNPs were tested using multivariable linear regression models in MACH2QTL v.1.08,¹⁸ adjusting for age, sex, BMI and the first 10 principal components generated from EIGENSOFT²¹ based on a linkage disequilibrium pruned set of SNPs with MAF >0.05. An additive mode-of-inheritance model was assumed for genotype; β -coefficients, representing the estimated change in transformed trait value associated with each additional copy of the effect allele, and their corresponding standard errors were reported.

Manhattan plots were made to illustrate the association results across the genome. Quantile–quantile (Q–Q) plots of observed versus expected $-\log_{10}$ (P -values) were made to assess any systematic inflation of the regression test statistics across the genome before and after removing the SNPs in the 6q25.3 region widely reported to be significantly associated with Lp(a). The genomic inflation factor (λ), defined as observed median value of the χ^2 statistic divided by 0.456, was calculated excluding 6q25.3 SNP results. The observed χ^2 statistics were divided by λ to obtain the corrected χ^2 statistics and corrected P -values. In regions with significant evidence for association, multivariable ‘conditional’ regression models that included the imputed genotype data of the most strongly associated SNP as an additional covariate were performed to assess the evidence for multiple independently associated SNPs in the region. If a second signal also reached genome-wide significance after conditioning on the top variant, multivariable regression models were repeated to include the genotypes of both SNPs as covariates. Region-specific plots were made to show the magnitude of association between all SNPs and Lp(a) levels, as well as the estimated LD between each SNP in the region and the most strongly associated SNP. Finally, to control our association results for possible confounding due to ancestry, for any observed associated SNP, we identified the genetic position of the most strongly associated SNP in the associated region, selected the local ancestry estimate at the location closest to that SNP and performed multivariable regression models as described above, but now including estimated local ancestry proportion as an additional covariate.

Identifying the most important 6q associated SNPs using LASSO-based resample model averaging. To identify the SNPs that most likely contribute to the observed association at 6q, we applied the LASSO local automatic regularization resample model averaging (LLARRMA) method, a method that combines LASSO variable shrinkage and selection with resample model averaging and multiple imputation,²² to estimate the probability of each SNP to be included in a multi-SNP model that best explains the Lp(a) outcome across alternative realizations of the data. We first extracted the genotypes for SNPs on the Affy 6.0 array that mapped within the 1 Mb region centered around the top genotyped SNP (rs9457986) identified in the initial GWAS scan. Then, we used fastPHASE²³ to impute the missing genotypes for SNPs in this region, which failed genotyping on Affy 6.0. Finally, we fitted the LASSO models by regressing the residuals of Lp(a) adjusting for age, gender, BMI, 10 principal components and the local ancestry estimate of rs9457986 on the genotypes using the LLARRMA package in R, and calculated a resample model inclusion probability (RMIP) score for each SNP in this 1 Mb region on chromosome 6. LD statistics (R^2 and D' based on 1000G YRI subjects) between variants with RMIP >0.75 were calculated using GOLD²⁴ and plotted using Haploview.²⁵

Haplotype analyses. Haplotype analyses were conducted using the ‘haplo.stats’ R package,²⁶ to examine specific combinations of allelic variants and whether the observed association signal is likely attributable to a less common unmeasured genetic variant. Haplotypes were constructed among the five genotyped variants in the *LPA* gene with RMIP score >0.75. The ‘haplo.glm’ function implemented in the ‘haplo.stats’ R package was used to calculate effect coefficients (β), standard errors (SE) and P -values for each haplotype relative to

Table 1 Descriptive statistics of JHS participants

	Total	Male	Female
<i>N</i>	2896	1107	1789
Age (years)	54.4 ± 12.9	53.8 ± 13.0	54.8 ± 12.8
BMI	32.1 ± 7.5	30.0 ± 6.3	33.3 ± 7.9
Lp(a) (mg dl ⁻¹)	47 (25, 80)	42 (23, 74)	50 (26, 84)
LDL-c (mg dl ⁻¹)	128 (103, 153)	130 (106, 155)	127 (101, 151)
HDL-c (mg dl ⁻¹)	49 (40, 59)	43 (37, 51)	52 (44, 62)
TG (mg dl ⁻¹)	93 (66, 131)	98 (70, 139)	89 (62, 125)
TC (mg dl ⁻¹)	199 (175, 229)	198 (175, 228)	200 (175, 230)

Abbreviations: BMI, body mass index; HDL-c, high-density lipoprotein; JHS, Jackson Heart Study; LDL-c, low-density lipoprotein; TC, total cholesterol; TG, triglyceride.

Note: Data are mean ± s.d., median (25th, 75th percentiles); lipid levels are untreated.

the most common reference haplotype. The 'haplo.score' function was used to calculate the global score statistic to test the overall association between haplotypes and log-transformed Lp(a). The same set of covariates used in the genotype analyses were used in the haplotype analyses.

Single variant and gene-based analysis of SNPs on exome array. Single variant and multivariate, gene-set or 'gene-burden', association analyses were performed for variants appearing on the exome array. Gene-burden tests performed included the Madsen–Browning test²⁷ and the SKAT-Optimal (SKAT-O)²⁸ test, which picks the 'best' combination of the SKAT²⁹ and a Madsen–Browning²⁷ test for gene-based testing. We performed three levels of gene-based analyses: (level 1) the combination of stop-loss, stop-gain and splice-site regardless of MAF; (level 2) the combination of SNPs in level 1 and all variants, which are predicted to be 'damaging' using PolyPhen³⁰ and SIFT³¹ regardless of MAF; (level 3) the combination of stop-loss, stop-gain, splice-site and nonsynonymous variants with MAF upper limits of 3%.

To investigate how much common or rare SNPs associated with Lp(a) in the single-variant analyses explain the gene-based test results, conditional analyses were performed by including the allele count at these lead SNPs as covariates. To control for the possible confounding of Lp(a) association results because of an association between both SNP and Lp(a) with other lipid traits (HDL, LDL, TG and TC), we performed regression models as described above but now including the lipid traits an additional covariate.

Statistical significance. A significance threshold of $P < 5 \times 10^{-8}$ was used to define genome-wide significance for all individual SNP results (both GWAS and exome array). Test statistics for all individual SNP results, both GWAS and Exome-chip, were adjusted by the genome inflation factor (λ) before calculating significance (P -values) to account for any possible systematic bias in results. A gene-based association result was defined to be significant if $P < 0.05$ per number of genes. The number of genes for level 1, level 2 and level 3 inclusion criteria were 4752, 13 658 and 15 963, resulting in significance thresholds of 1.1×10^{-5} , 3.7×10^{-6} and 3.1×10^{-6} , respectively.

RESULTS

Descriptive statistics of the JHS participants in this study were summarized in Table 1.

Admixture mapping for determinants of Lp(a)

Consistent with observed higher levels of Lp(a) in AAs versus EAs, higher levels of estimated global, or average, African ancestry was significantly associated with higher levels of log-transformed Lp(a)

($\beta = 0.76$, $P = 1.3 \times 10^{-9}$). The estimated global proportion of African ancestry obtained from ADMIXTURE was highly correlated with the first principal component from EIGENSTRAT (correlation = 0.998). Admixture mapping showed a highly significant association between increased African ancestry at chromosome 6q25.3 and increased Lp(a) levels (Supplementary Figure 1) after adjusting for the global proportion of African ancestry. The estimated global proportion of African ancestry became nonsignificant after including the estimated local ancestry at 6q25.3 in the regression models. SNP rs505000, upstream of *SLC22A3*, was the local ancestry informative SNP most strongly associated with Lp(a) levels ($P = 8.2 \times 10^{-27}$, LOD = 24.95).

GWAS SNPs associated with Lp(a) at $P < 5 \times 10^{-8}$

Q–Q plots revealed evidence for systematic inflated association results likely because of left-censoring of the Lp(a) measures in a subset of JHS participants (Supplementary Figure 2a). After excluding all the SNPs on chromosome 6, the distribution of the remaining P -values across the genome still demonstrated inflation (Supplementary Figure 2b). After controlling for the genomic inflation factor ($\lambda = 1.14$), Q–Q plots revealed no substantial evidence for inflation (Supplementary Figures 2c and d). The genomic inflation factor was applied to all individual SNP results. Only the chromosome 6q region reached genome-wide significance for Lp(a) levels (Supplementary Figure 2e) after adjusting for the inflation factor. Overall, 804 SNPs reached genome-wide significance ($P < 5 \times 10^{-8}$) (all on chromosome 6), and the top SNPs that reached genome-wide significance are listed in Table 2.

Lp(a) GWAS results on chromosome 6

All 804 significant SNPs on chromosome 6 mapped to the 6q region, spanning from 153 917 144 to 163 745 411 bp and containing more than 10 genes (Figure 1 and Supplementary Table 1). The strongest signal (rs115848955, $P = 3.1 \times 10^{-55}$, MAF = 0.05) mapped to the *LPA* gene. *LPA* encodes a modified form of low-density lipoprotein, in which a large glycoprotein (Apo(a)) is covalently bound to apolipoprotein B by a disulfide bridge,³² and structurally, the Apo(a) chain contains a region homologous with plasminogen, which gives Lp(a) antifibrinolysis activity by competing with plasminogen's

Table 2 Top SNPs that reached genome-wide significance ($P < 5 \times 10^{-8}$) after controlling for genomic inflation factor

Chr.	No. of SNPs	Most significant SNP	Pos(hg19)	EA	EAF	RSQR	β	SE	P	Nearest gene	Function
6	5	Rs9322428	153 917 946	A	0.85	0.95	0.19	0.03	1.3E–09	<i>RGS17-OPRM1</i>	Intergenic
6	6	Rs17539620	154 896 235	T	0.05	0.85	–0.33	0.05	3.7E–10	<i>CNKSR3-SCAF8-TIAM2</i>	Intergenic
6	24	6:156789295:T_TAC	15 678 929 5	I	0.16	0.74	–0.22	0.03	4.7E–11	<i>NOX3-ARID1B</i>	Intergenic
6	5	Rs6909229	158 572 379	C	0.07	0.92	–0.27	0.04	3.7E–10	<i>SYNJ2-SERAC1-GTF2H5-TULP4</i>	Intronic
6	19	6:159106232:G_GTC	159 106 232	I	0.60	0.95	0.18	0.02	4.8E–15	<i>SYTL3</i>	Intronic
6	41	Rs926657	159 463 452	T	0.42	0.99	0.14	0.02	4.6E–11	<i>C6orf99-RSPH3-TAGAP-FNDC1-SOD2-PNLDC1-MAS1-IGF2R</i>	Intergenic
6	261	Rs149565105	160 878 078	A	0.02	0.97	0.78	0.06	1.7E–34	<i>SLC22A1-SLC22A2-SLC22A3</i>	Intronic
6	55	Rs185414370	160 889 898	C	0.03	0.92	0.82	0.06	2.9E–40	<i>LPL2</i>	ncRNA_intronic
6	177	Rs115848955	161 031 660	T	0.05	0.90	0.79	0.05	1.3E–62	<i>LPA</i>	Intronic
6	130	Rs144788267	161 181 875	A	0.03	0.75	0.83	0.07	4.6E–30	<i>PLG</i>	Intergenic
6	25	Rs142799378	161 305 763	G	0.02	0.85	0.81	0.08	6.8E–23	<i>MAP3K4</i>	Intergenic
6	24	Rs3757037	161 697 400	G	0.66	0.68	0.20	0.03	3.3E–13	<i>AGPAT4</i>	Intergenic
6	27	Rs7769089	162 147 727	T	0.67	1.00	0.16	0.02	4.6E–12	<i>PARK2</i>	Intronic
6	5	Rs6927207	163 740 089	A	0.22	0.75	–0.19	0.03	1.3E–11	<i>PACRG-AS1</i>	ncRNA_intronic

Abbreviations: β , β coefficients representing the estimated change in the log-Lp(a) level associated with each additional copy of the effect allele; Chr, chromosome; EA, effect allele; EAF, effect allele frequency; no. of SNPs, number of SNPs that reached genome-wide significance at each locus; P, P-value after genomic control; Pos(hg19), physical position of the SNP according to human genome build version 19; RSQR, the imputation quality provided by MACH. SNPs with RSQR < 0.3 were excluded from analyses.

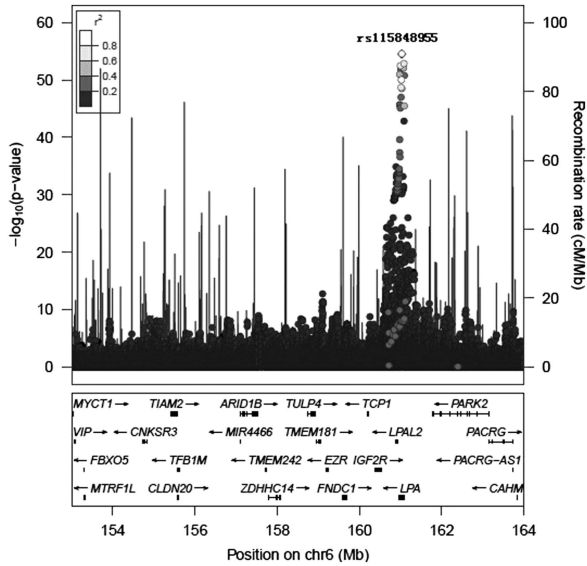


Figure 1 Regional plot of the $-\log_{10}(P)$ values for the single-nucleotide polymorphisms (SNPs) in the chromosome 6q region for Lp(a). The X axis shows the human genome build 19 coordinates (Mb) and the genes in the region. The Y axis shows the $-\log_{10}$ association P -values of SNPs on the left, and recombination rates in cM per Mb on the right. Different colors of shading indicate the strength of linkage disequilibrium (LD) (R^2) between the top SNP and the other SNPs tested in the region.

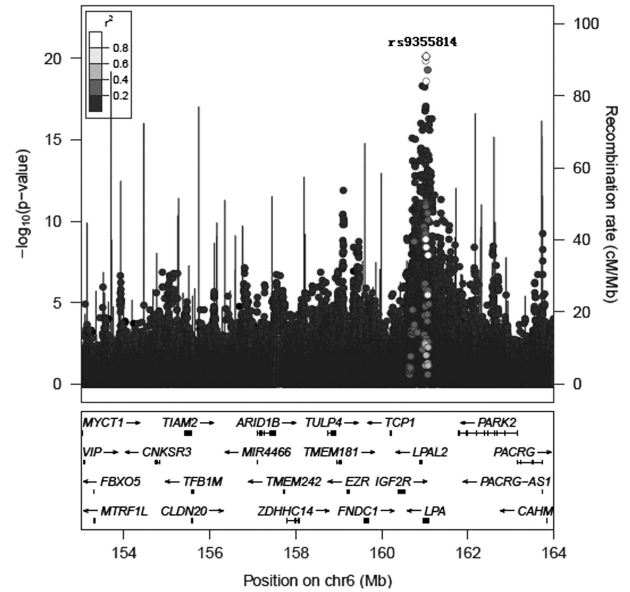


Figure 2 Regional plot of the $-\log_{10}(P)$ values for the single-nucleotide polymorphisms (SNPs) in the chromosome 6q region for Lp(a) after adjusting for the top SNP rs115848955 at this locus. The top SNP was rs9355814 after adjusting for rs115848955.

binding to fibrin. After adjusting for the top SNP (rs115848955), the top associated variant was rs9355814 ($P=7.8 \times 10^{-21}$) (Figure 2 and Supplementary Table 2) and 469 SNPs in the 6q region spanning from 159 092 125 to 163 745 411 bp remained genome-wide significant. However, after adjusting for the local ancestry informative marker closest to rs115848955 (i.e. rs6923917), only 406 SNPs in a relatively narrow region spanning from 160 633 560 to 161 342 219 bp remained significant ($P < 5 \times 10^{-8}$) (Supplementary Table 3). The top SNP in the conditional analysis including local ancestry was rs138429428 ($P=4.2 \times 10^{-50}$, Figure 3). The local ancestry estimate closest to rs115848955 explained an estimated 4.4% of the total variation of Lp(a) after accounting for age, gender, BMI and the first 10 PCs. SNP rs115848955 explained an estimated 8.8% of the total variation of Lp(a) after accounting for age, gender, BMI, the first 10 PCs, and the local ancestry estimate nearest this location. In a series of iterative conditional (forward step-wise) analyses that also contained the local ancestry estimate at rs6923917, 11 SNPs remained genome-wide significant ($P < 5 \times 10^{-8}$) in the *LPA* region (see Supplementary Table 4 for the 11 SNP model results). In aggregate, these 11 SNPs plus the local ancestry estimate at rs6923917 explained 28.8% of the total variation in Lp(a) after accounting for age, gender, BMI and the first 10 PCs. Many more *LPA* region SNPs remained nominally significant ($P < 0.05$) in conditional models including these 11 SNPs.

Evidence for multiple associated SNPs in the *LPA* region on chromosome 6

Fifteen directly genotyped SNPs in the 1 Mb region surrounding the top associated SNP (rs9457986) on chromosome 6 were identified as having stable, independent associations with Lp(a) (RMIP score > 0.75) in multivariate models using LLARRMA, as shown in Supplementary Table 5. The linkage disequilibrium estimates between these SNPs were not high, as shown in Supplementary Figure 3. Five of the SNPs (rs6415084 (RMIP = 0.93), rs3798221 (RMIP = 1),

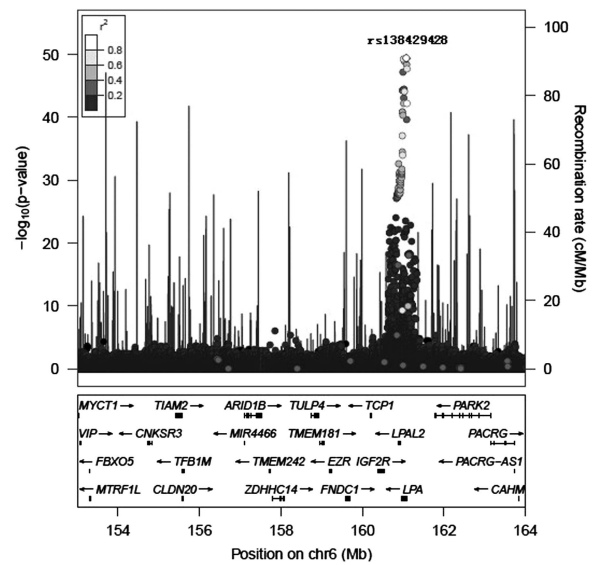


Figure 3 Regional plot of the $-\log_{10}(P)$ values for the single-nucleotide polymorphisms (SNPs) in the chromosome 6q region for Lp(a) after adjusting for the local ancestry estimate at rs6923917. The top SNP was rs138429428, and only SNPs in a relatively narrow region spanning from 160 633 560 to 161 342 219 bp remained significant ($P < 5 \times 10^{-8}$).

rs9457986 (RMIP = 1), rs1367211 (RMIP = 0.97) and rs1406888 (RMIP = 1)) were in the *LPA* gene. The five Lp(a) SNPs together explained an estimated 7.3% of the total variance of Lp(a) after accounting for age, gender, BMI, the first 10 PCs and local ancestry estimate at rs9457986. These five SNPs were used to construct haplotypes and to estimate the effect on Lp(a) levels for each additional copy of a particular haplotype, compared with the reference haplotype (Supplementary Table 6). There was significant evidence

for an overall association between haplotypes and Lp(a) (global $P = 1.12 \times 10^{-55}$); six individual haplotypes were minimally nominally associated ($P < 0.05$) with Lp(a) levels.

Single variant and gene-based analysis for exome array

Seven SNPs on the exome array were associated ($P < 5 \times 10^{-8}$) with Lp(a) (Table 3) after adjustment for the genomic inflation factor, six of them were in the chromosome 6q *LPA* region, and the other was the widely reported chromosome 19 *APOE* SNP rs7412. Five of the six associated chromosome 6q SNPs remained genome-wide significant after adjusting for top GWAS SNP rs115848955 (Table 3). Five of the six SNPs (excepting rs41272114) were intronic or in intergenic regions. Of the seven SNPs, only rs7412 was not significant in the 1000G imputed GWAS results. Rs7412 ($P = 3.3 \times 10^{-8}$) was poorly imputed based on 1000G data and was not included in the original GWAS analyses. In the gene-based analyses, the *LPA* gene reached exome-wide significance for all three levels of SNP inclusion when using the Madsen–Browning test, as well as for level 2 and level 3 SNPs in the SKAT-O test. The *APOE* gene reached exome-wide significance for level 2 SNPs using the Madsen–Browning test (Table 4). Q–Q plots for the single-variant (Supplementary Figure 4) and gene-based analyses (Supplementary Figure 5) showed no evidence for systematic inflation in results.

Uncommon functional variants in *LPA* and *APOE*

All splice-site, stop altering and nonsynonymous SNPs in *LPA* and *APOE* on the exome array are listed in Supplementary Table 7 and Supplementary Table 8, respectively. The most significant individual SNP in *LPA* on the exome array was rs41272114 ($P = 6.5 \times 10^{-12}$, MAF = 0.01), which is a splicing-altering variant. *LPA* SNPs were significantly associated with Lp(a) levels in the gene-burden tests ($P = 1.2 \times 10^{-21}$ for level 3 SNPs using Madsen–Browning test). The level 3 gene-based Madsen–Browning test remained significant after removing the top two individual SNP results (i.e. rs41272114 and rs41272110) (Supplementary Table 9). Only three *APOE* SNPs, including the aforementioned rs7412, were successfully genotyped and informative on the exome array. A second nonsynonymous *APOE* SNP (rs769455, MAF = 0.02), which is specific to populations of African descent, demonstrated strong nominal results that were independent of rs7412 ($P = 2.0 \times 10^{-5}$ before adjustment for rs7412 and $P = 3.5 \times 10^{-5}$ after adjustment for rs7412). Taken together, rs7412 and rs769455 explain ~2.3% of the total variation of Lp(a) after adjusting for age, gender, BMI and the first 10 PCs.

As *APOE* is a strong risk factor for lipid traits (LDL and TC), we investigated whether the association between *APOE* SNPs and Lp(a) were attenuated when adjusting for lipid traits. After adjusting for the untreated LDL and TC levels, we found that the evidence for

Table 3 SNPs on Exome-chip that reached exome-wide significance ($P < 5 \times 10^{-8}$) after controlling for genomic inflation factor

Chr.	Gene	Function	SNP	Pos(hg19)	MAF	β	SE	P	After adjusting for rs115848955			RSQR.in.	
									β .adj	SE.adj	P.adj	GWAS	P.in.GWAS
6	MIR1202-ARID1B	Intergenic	Rs9478712	156 858 484	0.14	-0.23	0.03	7.7E-11	-0.20	0.03	6.4E-09	0.99	1.3E-08
6	SYTL3	Intronic	Rs894124	159 096 121	0.39	-0.15	0.02	8.8E-10	-0.14	0.02	3.1E-09	0.99	7.1E-12
6	LPA	Intronic	Rs6919346	160 960 359	0.03	-0.43	0.07	1.5E-09	-0.39	0.06	6.0E-09	0.99	2.1E-09
6	LPA	Splicing	Rs41272114	161 006 077	0.01	-0.80	0.11	6.5E-12	-0.72	0.11	8.8E-11	0.55	3.2E-11
6	LPA	Intronic	Rs1652507	16 1082 461	0.08	-0.36	0.04	4.4E-16	-0.30	0.04	2.6E-12	0.99	6.3E-19
6	PARK2	Intronic	Rs6455767	162 148 335	0.28	-0.15	0.03	1.4E-08	-0.12	0.03	1.7E-06	0.94	5.2E-10
19	APOE	Exonic; nonsynonymous	Rs7412	45 412 079	0.11	-0.21	0.04	3.3E-08	-0.21	0.04	3.3E-08	0.28	3.3E-03

Abbreviations: β , β coefficients representing the estimated change in the log-Lp(a) level associated with each additional copy of the effect allele; Chr, chromosome; P, P-value after genomic control; PC, prostate cancer; P.in.GWAS, the P-value after genomic control in GWAS; Pos(hg19), physical position of the SNP according to human genome build version 19; RSQR.in.GWAS, the imputation quality of the Exome-chip SNP in GWAS 1000G; SNP, single-nucleotide polymorphism.
 β .adj, SE.adj and P.adj.GC were reported after adjusting for the lead SNP from GWAS signal, rs115848955.
Models were adjusted for age, gender, BMI and 10 PCs.

Table 4 Genes on Exome-chip associated with Lp(a) in gene-based analysis

	Gene	P-value	Madsen–Browning test				SKAT-O test			
			β	SE	cmafTotal ^a	cmafUsed ^b	Gene	P-value	cmafUsed ^b	No. of SNPs
Level 1	LPA	7.4E-14	-0.76	0.10	0.01	0.01				
Level 2	LPA	1.5E-13	-0.44	0.06	0.04	0.04	LPA	1.0E-06	0.04	14
	APOE	1.1E-12	-0.24	0.03	0.13	0.13				
Level 3	LPA	1.2E-21	-0.40	0.04	0.55	0.08	LPA	6.2E-17	0.08	24

Abbreviations: BMI, body mass index; MAF, minor allele frequency; PC, prostate cancer; SKAT-O, SKAT-Optimal; SNP, single-nucleotide polymorphism.
Note: Lp(a) levels were log transformed and residuals were adjusted for age, gender, BMI, 10 PCs and family relatedness. Gene-based tests were carried out using Madsen–Browning test and SKAT-O test, respectively. Level 1 is the combination of stop-loss, stop-gain and splice-site regardless of MAF. Level 2 is the combination of SNPs in level 1 and all variants that are predicted to be 'damaging' using PolyPhen regardless of MAF. Level 3 is the combination of stop-loss, stop-gain, splice-site and nonsynonymous variants with MAF upper limits of 3%. A gene-based association was defined to be significant if $P < 0.05$ per number of genes. Number of genes for level 1, level 2 and level 3 were 4752, 13 658 and 15 963, corresponding to a P-value of 1.05×10^{-5} , 3.66×10^{-6} and 3.13×10^{-6} .
^acmafTotal, cumulative MAF of total Exome-chip variants in the *LPA* gene.
^bcmafUsed, cumulative MAF of Exome-chip variants in the *LPA* gene that were included in the gene-based test.

association decreased, but the signal did not totally disappear (Supplementary Table 8, P -value for rs7412 dropped from 3.2×10^{-8} to 7.6×10^{-4}).

DISCUSSION

We conducted genome- and exome-wide association studies for Lp(a) in 2896 AA participating in the JHS. A higher level of estimated global African ancestry was significantly associated with a higher level of Lp(a), and this association of global ancestry was largely explained by the association between Lp(a) and local ancestry in the chromosome 6q25.3 region. We observed significant ($P < 5 \times 10^{-8}$) associations for hundreds of SNPs spanning ~ 10 Mb region on 6q surrounding the *LPA* gene. Interestingly, after adjusting for local ancestry, the region containing significantly associated SNPs got much narrower and was centered over the *LPA* gene (< 1 Mb). Significant haplotypic effects were also detected in the *LPA* region that implicates numerous causal variants. A single *APOE* SNP, rs7412 on the exome array, also reached genome-wide significance. Gene-burden tests found significant associations between Lp(a) and aggregate collections of SNPs in *LPA* and *APOE*.

Previously, Deo *et al.*¹¹ performed a targeted study of the Lp(a) region, using haplotype tagging SNPs, in 4464 JHS participants and 1726 AA participants from the Dallas Heart Study.¹¹ This study also performed a genome-wide admixture analyses based on a panel of 1447 ancestry informative markers, including subjects in the upper and lower quintile of the Lp(a) distribution. Herein, we markedly expand the coverage of both common and rare variants across the entire *LPA* region; estimate local genetic admixture at 738 831 autosomal SNP locations and perform admixture mapping including all subjects with Lp(a) measures. Eight-hundred and four SNPs in the 6q region (spanning ~ 10 Mb) were significantly associated with Lp(a) levels ($P < 5 \times 10^{-8}$). After adjusting for the most strongly associated SNP rs115848955, multiple signals at 6q region spanning ~ 5 Mb remained significantly associated with Lp(a). After adjusting for the local ancestry at 6q25.3, the region that harbors SNPs significantly associated with Lp(a) ($P < 5 \times 10^{-8}$) became much narrower (from 9.8 to 0.7 Mb) and was centered around the three genes *SLC22A*, *LPL2* and *LPA*. This result suggests confounding between local ancestry and SNPs spanning the larger 6q region identified to be associated with Lp(a). Given the relatively recent admixture in the AA population, local ancestry can confound associations across a relatively large region surrounding the population-specific or population-enriched causal variant(s).^{33,34} The observation that the associations in and near *LPA* remains robust after adjustment for local ancestry at *LPA*, whereas the evidence for association further away markedly declines, suggests that the ancestry-specific (or highly-enriched) causal risk variant(s) resides in or near *LPA* and that most, if not all, of the observed associations outside this narrower region are spurious associations. Interestingly, a similar extended region of association with Lp(a) surrounding *LPA* has been observed in more homogeneous European populations. An obvious candidate to explain some of the differences in association results between European and African populations is the Kringle IV polymorphism, which has not been measured in JHS participants. Deo *et al.*¹¹ demonstrated that the Kringle IV polymorphism explains some of the ancestry effect differences, but noted that several associated common SNPs, with strong allele frequency differences between populations of African and European ancestry, in and around *LPA* explain the majority of the population differences in Lp(a) levels. To our knowledge, there is no evidence suggesting the allele frequency differences between populations at these, or other nearby *LPA* variants, are due to selective pressures. Future functional studies will be

necessary to better understand the significance of these allele frequency differences.

A common nonsynonymous variant at *APOE* on exome array, rs7412, was identified to be significantly associated with Lp(a) in single-variant analysis (MAF=0.11, leading to an Arg to Cys substitution, $P = 3.2 \times 10^{-8}$). Another low-frequency non-synonymous variant at *APOE*, rs769455, was also nominally associated with Lp(a) (MAF=0.02, leading to an Arg to Cys substitution, $P = 2.0 \times 10^{-5}$). Prior studies have investigated the relationship between *APOE* genotypes and Lp(a) levels, but the results were inconsistent. Some studies reported no impact of *APOE* genotypes on Lp(a) levels,^{35–39} whereas others reported significant associations between them.^{40–45} A recent study found that among AAs, lower Lp(a) levels were observed in *APOE* $\epsilon 2$ carriers, and this association was only observed in subjects with large apoA size (defined as > 26 Kringle IV repeats) but not in the subjects with small apoA size.⁴⁶ Another study on Caucasian males also reported that the effect of *APOE* on Lp(a) levels was only observed in subjects of largest quartile of apoA size, but with lower Lp(a) levels for *APOE* $\epsilon 4$ carriers.⁴² In our study, *APOE* $\epsilon 2$ genotype was associated with lower Lp(a) levels, which is consistent with the prior study on AAs. We found no evidence supporting any such interaction between *APOE* genotype and any of our top *LPA* region SNPs (data not shown).

In summary, we observed that local ancestry at 6q25.3 was an important risk factor for Lp(a) in AA, and that SNPs at the well-established *LPA* locus were significantly associated with Lp(a) ($P < 5 \times 10^{-8}$) after adjusting for the local ancestry at 6q25.3. Before covariate adjustment for local ancestry at 6q25.3, the observed region containing associated SNPs spanned ~ 10 Mb. After covariate adjustment, the associated region was only 700 kb. We also observed a significant association for a nonsynonymous variant in *APOE*. Future large multiethnic studies which include high-coverage sequence data, and the Kringle IV polymorphism, would be ideally suited to better understand the complex genetic architecture of the *LPA* region that leads to strong population differences in Lp(a) levels.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We express our gratitude to JHS participants who made this research possible. JHS is supported by contracts HHSN268201300046C, HHSN268201300047C, HHSN268201300048C, HHSN268201300049C and HHSN268201300050C from the National Heart, Lung and Blood Institute and the National Institute on Minority Health and Health Disparities.

- Nordestgaard, B. G., Chapman, M. J., Ray, K., Boren, J., Andreotti, F., Watts, G. F. *et al.* Lipoprotein(a) as a cardiovascular risk factor: current status. *Eur. Heart J.* **31**, 2844–2853 (2010).
- Erqou, S., Kaptoge, S., Perry, P. L., Di Angelantonio, E., Thompson, A., White, I. R. *et al.* Lipoprotein(a) concentration and the risk of coronary heart disease, stroke, and nonvascular mortality. *JAMA* **302**, 412–423 (2009).
- Clarke, R., Peden, J. F., Hopewell, J. C., Kyriakou, T., Goel, A., Heath, S. C. *et al.* Genetic variants associated with Lp(a) lipoprotein level and coronary disease. *N. Engl. J. Med.* **361**, 2518–2528 (2009).
- Koschinsky, M. L., Cote, G. P., Gabel, B. & van der Hoek, Y. Y. Identification of the cysteine residue in apolipoprotein(a) that mediates extracellular coupling with apolipoprotein B-100. *J. Biol. Chem.* **268**, 19819–19825 (1993).
- Boerwinkle, E., Leffert, C. C., Lin, J., Lackner, C., Chiesa, G. & Hobbs, H. H. Apolipoprotein(a) gene accounts for greater than 90% of the variation in plasma lipoprotein(a) concentrations. *J. Clin. Invest.* **90**, 52–60 (1992).
- Mooser, V., Scheer, D., Marcovina, S. M., Wang, J., Guerra, R., Cohen, J. *et al.* The Apo(a) gene is the major determinant of variation in plasma Lp(a) levels in African Americans. *Am. J. Hum. Genet.* **61**, 402–417 (1997).

- 7 Ober, C., Nord, A. S., Thompson, E. E., Pan, L., Tan, Z., Cusanovich, D. *et al*. Genome-wide association study of plasma lipoprotein(a) levels identifies multiple genes on chromosome 6q. *J. Lipid Res.* **50**, 798–806 (2009).
- 8 Sandholzer, C., Hallman, D. M., Saha, N., Sigurdsson, G., Lackner, C., Csaszar, A. *et al*. Effects of the apolipoprotein(a) size polymorphism on the lipoprotein(a) concentration in 7 ethnic groups. *Hum. Genet.* **86**, 607–614 (1991).
- 9 Lanktree, M. B., Anand, S. S., Yusuf, S. & Hegele, R. A. Comprehensive analysis of genomic variation in the LPA locus and its relationship to plasma lipoprotein(a) in South Asians, Chinese, and European Caucasians. *Circ. Cardiovasc. Genet.* **3**, 39–46 (2010).
- 10 Marcovina, S. M., Albers, J. J., Wijsman, E., Zhang, Z., Chapman, N. H. & Kennedy, H. Differences in Lp(a) concentrations and apo(a) polymorphs between black and white Americans. *J. Lipid Res.* **37**, 2569–2585 (1996).
- 11 Deo, R. C., Wilson, J. G., Xing, C., Lawson, K., Kao, W. H., Reich, D. *et al*. Single-nucleotide polymorphisms in LPA explain most of the ancestry-specific variation in Lp(a) levels in African Americans. *PLoS ONE* **6**, e14581 (2011).
- 12 Payne, T. J., Wyatt, S. B., Mosley, T. H., Dubbert, P. M., Guiterrez-Mohammed, M. L., Calvin, R. L. *et al*. Sociocultural methods in the Jackson Heart Study: conceptual and descriptive overview. *Ethn. Dis.* **15**, S6–S8–48 (2005).
- 13 Wilson, J. G., Rotimi, C. N., Ekunwe, L., Royal, C. D., Crump, M. E., Wyatt, S. B. *et al*. Study design for genetic analysis in the Jackson Heart Study. *Ethn. Dis.* **15**, S6–S30–37 (2005).
- 14 Carpenter, M. A., Crow, R., Steffes, M., Rock, W., Heilbraun, J., Evans, G. *et al*. Laboratory, reading center, and coordinating center data management methods in the Jackson Heart Study. *Am J Med Sci* **328**, 131–144 (2004).
- 15 Deo, R. C., Reich, D., Tandon, A., Akyzbekova, E., Patterson, N., Waliszewska, A. *et al*. Genetic differences between the determinants of lipid profile phenotypes in African and European Americans: the Jackson Heart Study. *PLoS Genet.* **5**, e1000342 (2009).
- 16 Asselbergs, F. W., Guo, Y., van Iperen, E. P., Sivapalaratnam, S., Tragante, V., Lanktree, M. B. *et al*. Large-scale gene-centric meta-analysis across 32 studies identifies multiple lipid loci. *Am. J. Hum. Genet.* **91**, 823–838 (2012).
- 17 Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D. *et al*. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
- 18 Li, Y., Willer, C. J., Ding, J., Scheet, P. & Abecasis, G. R. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet. Epidemiol.* **34**, 816–834 (2010).
- 19 Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
- 20 Patterson, N., Hattangadi, N., Lane, B., Lohmueller, K. E., Hafler, D. A., Oksenberg, J. R. *et al*. Methods for high-density admixture mapping of disease genes. *Am. J. Hum. Genet.* **74**, 979–1000 (2004).
- 21 Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A. & Reich, D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
- 22 Valdar, W., Sabourin, J., Nobel, A. & Holmes, C. C. Reprioritizing genetic associations in hit regions using LASSO-based resample model averaging. *Genet. Epidemiol.* **36**, 451–462 (2012).
- 23 Scheet, P. & Stephens, M. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *Am. J. Hum. Genet.* **78**, 629–644 (2006).
- 24 Abecasis, G. R. & Cookson, W. O. GOLD—graphical overview of linkage disequilibrium. *Bioinformatics* **16**, 182–183 (2000).
- 25 Barrett, J. C., Fry, B., Maller, J. & Daly, M. J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263–265 (2005).
- 26 Schaid, D. J. Power and sample size for testing associations of haplotypes with complex traits. *Ann. Hum. Genet.* **70**, 116–130 (2006).
- 27 Madsen, B. E. & Browning, S. R. A groupwise association test for rare mutations using a weighted sum statistic. *PLoS Genet.* **5**, e1000384 (2009).
- 28 Lee, S., Wu, M. C. & Lin, X. Optimal tests for rare variant effects in sequencing association studies. *Biostatistics (Oxford, England)* **13**, 762–775 (2012).
- 29 Wu, M. C., Lee, S., Cai, T., Li, Y., Boehnke, M. & Lin, X. Rare-variant association testing for sequencing data with the sequence kernel association test. *Am. J. Hum. Genet.* **89**, 82–93 (2011).
- 30 Adzhubei, I., Jordan, D. M. & Sunyaev, S. R. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr. Protoc. Hum. Genet.* Chapter **7**, Unit **7.20**; doi: 10.1002/0471142905.hg0720s76 (2013).
- 31 Ng, P. C. & Henikoff, S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res.* **31**, 3812–3814 (2003).
- 32 Steyrer, E., Durovic, S., Frank, S., Giessauf, W., Burger, A., Dieplinger, H. *et al*. The role of lecithin: cholesterol acyltransferase for lipoprotein (a) assembly. Structural integrity of low density lipoproteins is a prerequisite for Lp(a) formation in human plasma. *J. Clin. Invest.* **94**, 2330–2340 (1994).
- 33 Qin, H., Morris, N., Kang, S. J., Li, M., Tayo, B., Lyon, H. *et al*. Interrogating local population structure for fine mapping in genome-wide association studies. *Bioinformatics* **26**, 2961–2968 (2010).
- 34 Wang, X., Zhu, X., Qin, H., Cooper, R. S., Ewens, W. J., Li, C. *et al*. Adjustment for local ancestry in genetic association analysis of admixed populations. *Bioinformatics* **27**, 670–677 (2011).
- 35 Heng, C. K., Saha, N. & Tay, J. S. Lack of association of apolipoprotein E polymorphism with plasma Lp(a) levels in the Chinese. *Clin. Genet.* **48**, 113–119 (1995).
- 36 Muros, M. & Rodriguez-Ferrer, C. Apolipoprotein E polymorphism influence on lipids, apolipoproteins and Lp(a) in a Spanish population underexpressing apo E4. *Atherosclerosis* **121**, 13–21 (1996).
- 37 Muls, E., Kempen, K., Vansant, G., Cobbaert, C. & Saris, W. The effects of weight loss and apolipoprotein E polymorphism on serum lipids, apolipoproteins A-I and B, and lipoprotein(a). *Int. J. Obes. Relat. Metab. Disord.* **17**, 711–716 (1993).
- 38 Schaefer, E. J., Lamon-Fava, S., Johnson, S., Ordovas, J. M., Schaefer, M. M., Castelli, W. P. *et al*. Effects of gender and menopausal status on the association of apolipoprotein E phenotype with plasma lipoprotein levels. Results from the Framingham Offspring Study. *Arterioscler. Thromb.* **14**, 1105–1113 (1994).
- 39 Bailleul, S., Couderc, R., Rossignol, C., Fermanian, J., Boutouchent, F., Farnier, M. A. *et al*. Lipoprotein(a) in childhood: relation with other atherosclerosis risk factors and family history of atherosclerosis. *Clin. Chem.* **41**, 241–245 (1995).
- 40 de Knijff, P., Kaptein, A., Boomsma, D., Princen, H. M., Frants, R. R. & Havekes, L. M. Apolipoprotein E polymorphism affects plasma levels of lipoprotein(a). *Atherosclerosis* **90**, 169–174 (1991).
- 41 Routi, T., Ronnema, T., Salo, P., Seppanen, R., Marniemi, J., Viikari, J. *et al*. Effects of prospective, randomized cholesterol-lowering dietary intervention and apolipoprotein E phenotype on serum lipoprotein(a) concentrations of infants aged 7–24 mo. *Am. J. Clin. Nutr.* **63**, 386–391 (1996).
- 42 Klausen, I. C., Gerdes, L. U., Hansen, P. S., Lemming, L., Gerdes, C. & Faergeman, O. Effects of apoE gene polymorphism on Lp(a) concentrations depend on the size of apo(a): a study of 466 white men. *J. Mol. Med.* **74**, 685–690 (1996).
- 43 Horita, K., Eto, M., Saito, M., Nakata, H., Iwashima, Y., Ito, H. *et al*. Effects of apolipoprotein E polymorphism on plasma lipoprotein(a) levels. *Artery* **20**, 324–336 (1993).
- 44 Frikke-Schmidt, R., Nordestgaard, B. G., Agerholm-Larsen, B., Schnohr, P. & Tybjaerg-Hansen, A. Context-dependent and invariant associations between lipids, lipoproteins, and apolipoproteins and apolipoprotein E genotype. *J. Lipid Res.* **41**, 1812–1822 (2000).
- 45 Lindahl, G., Mailly, F., Humphries, S. & Seed, M. Apolipoprotein E phenotype and lipoprotein(a) in familial hypercholesterolaemia: implication for lipoprotein(a) metabolism. *Clin. Invest.* **72**, 631–638 (1994).
- 46 Anuurad, E., Lu, G., Rubin, J., Pearson, T. A. & Berglund, L. ApoE genotype affects allele-specific apo(a) levels for large apo(a) sizes in African Americans: the Harlem-Basset Study. *J. Lipid Res.* **48**, 693–698 (2007).

Supplementary Information accompanies the paper on Journal of Human Genetics website (<http://www.nature.com/jhg>)