ORIGINAL ARTICLE

# Three single-nucleotide polymorphisms in *LPA* account for most of the increase in lipoprotein(a) level elevation in African Americans compared with European Americans

**J-P Chretien, J Coresh, Y Berthier-Schaad, W H L Kao, N E Fink, M J Klag, S M Marcovina, F Giaculli, M W Smith**

*J Med Genet* 2006;**43**:917–923. doi: 10.1136/jmg.2006.042119

See end of article for authors' affiliations

Correspondence to: J Coresh, Johns Hopkins University, 2024 E Monument St, Suite 2-600, Baltimore, MD 21205, USA; coresh@jhu.edu

Accepted 1 July 2006
**Published Online First 13 July 2006**

**Background:** The extent which universally common or population-specific alleles can explain between-population variations in phenotypes is unknown. The heritable coronary heart disease risk factor lipoprotein(a) (Lp(a)) level provides a useful case study of between-population variation, as the aetiology of twofold higher Lp(a) levels in African populations compared with non-African populations is unknown.
**Objective:** To evaluate the association between *LPA* sequence variations and Lp(a) in European Americans and African Americans and to determine the extent to which *LPA* sequence variations can account for between-population variations in Lp(a).
**Methods:** Serum Lp(a) and isoform measurements were examined in 534 European Americans and 249 African Americans from the Choices for Healthy Outcomes in Caring for End-Stage Renal Disease Study. In addition, 12 *LPA* variants were genotyped, including 8 previously reported *LPA* variants with a frequency of >2% in European Americans or African Americans, and four new variants.
**Results:** Isoform-adjusted Lp(a) level was 2.23-fold higher among African Americans. Three single-nucleotide polymorphisms (SNPs) were independently associated with Lp(a) level ($p<0.02$ in both populations). The Lp(a)-increasing SNP (G-21A, which increases promoter activity) was more common in African Americans, whereas the Lp(a)-lowering SNPs (T3888P and G+1/inKIV-8A, which inhibit Lp(a) assembly) were more common in European Americans, but all had a frequency of <20% in one or both populations. Together, they reduced the isoform-adjusted African American Lp(a) increase from 2.23 to 1.37-fold(a 60% reduction) and the between-population Lp(a) variance from 5.5% to 0.5%.
**Conclusions:** Multiple low-prevalence alleles in *LPA* can account for the large between-population difference in serum Lp(a) levels between European Americans and African Americans.

Most genetic association studies focus on within-population, rather than between-population, differences in disease susceptibility. Although between-population variation accounts for only 5–15% of human genetic diversity,[1] it is invoked increasingly to explain differences in disease risk and drug response across populations.[2 3] It is unknown whether such differences result from universally common alleles, which are hypothesised to contribute to within-population risk variation for common diseases,[4 5] or from alleles that are population specific or nearly so. The allelic spectrum of between-population variation has important implications, as identifying susceptibility alleles that are population specific may require strategies different from those identifying alleles that are common in all populations.

The coronary heart disease risk factor lipoprotein(a) (Lp(a)) level is twofold higher in African than in non-African populations,[6 7] and provides a useful case study of between-population variation. The cardiovascular pathogenicity of Lp(a) is established best in Caucasians[8 9] and probably involves the low-density lipoprotein-bound plasminogen homologue apolipoprotein(a) (apo(a)), may increase low-density lipoprotein delivery[10] and may inhibit plasminogen-mediated thrombolysis.[11] The apo(a) gene (*LPA*; MIM 152200) explains about 90% and 80% of Lp(a) level variance in European American[12] and African American[13] populations, respectively. A genomically unusual 5.6-kb variable tandem repeat in *LPA* encodes the KIV-2 units, whose number determines apo(a) isoform size and explains half of the *LPA* effect on Lp(a) level.[12 14] The inverse association between isoform size and Lp(a) level probably reflects impaired cellular secretion of larger isoforms, which has been observed in vitro.[15]

Isoform distributions are similar across populations, and do not explain higher African Lp(a) levels.[7] Eight *LPA* variants (all but one are single-nucleotide polymorphisms (SNPs)) have been associated with the isoform-adjusted Lp(a) level in Europeans[16–18 14 19 20] or Africans.[21 18] Some investigators have speculated that unidentified *trans*-acting[16 22] or environmental[23] factors may explain the between-population difference, as none of these have been shown to contribute substantially to higher African Lp(a) levels. . However, a substantial contribution of *LPA* to the between-population difference cannot be excluded and provides the simplest explanation.

Previous analyses have not considered multiple *LPA* variants simultaneously or comprehensively, and the extensive linkage disequilibrium across the gene[15 18 14 19] is expected to confound single-locus effect estimates. Also, few studies

**Abbreviations:** apo(a), apolipoprotein(a); Lp(a), lipoprotein(a); SNP, single nucleotide polymorphism; CHOICE, Choices for Healthy Outcomes in Caring for End-Stage Renal Disease; ESRD, end-stage renal disease

included both European and African or African American populations, precluding direct quantification of *LPA* variant contributions to the between-population difference. We report the results of a simultaneous analysis of multiple *LPA* variants, isoforms and Lp(a) level in a cohort of European Americans and African Americans.

## SUBJECTS AND METHODS

Subjects were drawn from the Choices for Healthy Outcomes in Caring for End-Stage Renal Disease (CHOICE) Cohort Study, a national, prospective study of patients receiving incident dialysis, initiated in 1995 to investigate treatment choices of dialysis modality and dose, and outcomes of dialysis care.[24] The CHOICE Study recruited 1041 patients receiving incident dialysis (767 haemodialysis and 274 peritoneal dialysis) aged >17 years from 81 not-for-profit US clinics between 1995 and 1998. Patients were enrolled a median of 45 days from initiation of chronic dialysis (98% within 4 months). This study consisted of 534 European American and 249 African American CHOICE participants who contributed blood samples to a specimen bank. Self-reported race was used.

Lp(a) and isoform measurements in this subpopulation were described previously.[25] The Lp(a) level was measured with a direct-binding double monoclonal antibody-based ELISA directed towards a non-repeating apo(a) epitope (so that measurement is insensitive to isoform size[7]) and was expressed as nmol/l of Lp(a) protein. Isoforms were characterised by high-resolution sodium dodecyl sulphate-agarose gel electrophoresis followed by immunoblotting and were expressed as number of KIV-2 units. When immunoblotting detected two isoforms, the more intense band defined the predominant isoform (which is the major contributor to plasma Lp(a)) and the other band defined the weaker isoform.

Twelve *LPA* loci were genotyped (fig 1A). Of the eight *LPA* variants associated with the isoform-adjusted Lp(a) level in any previous study, we included the five with European American or African American frequency of >2% (the remaining rare alleles may be causal,[18] but power would be limited in this study). We also genotyped two coding SNPs (LPA L3847V and LPA L3861V) that were not associated with Lp(a) level in one study of Europeans,[14] and identified four SNPs (L1 G-1712T, L1 C-1617A, L1 A1557G and L1 A-1230G) through denaturing high performance liquid chromatography screening[26] of a 5′ L1-based *LPA* enhancer[27] (L1 SNPs are numbered from the 3′ end of the enhancer-containing fragment, Genbank AF027597) and one silent substitution (LPA G4481G; rs3189802) by searching known *LPA* sequences in The Single Nucleotide Polymorphism database. L1 and *LPA* SNPs were genotyped using single-base extension[28] and an Applied Biosystems 3100 (Applied Biosystems, Foster City, CA, USA). The pentanucleotide short tandem repeat LPA –1400 repeat (rpt) was genotyped using fragment-length analysis with an Applied Biosystems 3700, combined with GeneScan and Genotyper software (Applied Biosystems).

Linear regression was used to estimate isoform and allelic effects on log(Lp(a)+1) level (the transformation induced approximate normality), as genotype effects were approximately additive. Effects were expressed as the exponentiated regression coefficient, or the ''Lp(a) ratio'', representing the additive effect of one allele. Predominant and weaker isoforms were included as categorical covariates in all analyses. As non-expressed (null) isoforms are not observable, the second isoform for subjects with a single band on agarose was imputed as identically sized to the observed isoform or as a null isoform using SNP-KIV-2 haplotype frequencies estimated with the 3LOCUS program, which allows for null alleles.

The frequency of the null isoform was estimated to be 20% in European Americans and 15% in African Americans, which is consistent with previous estimates on the basis of simultaneous assessments of isoforms and the isoform-encoding KIV-2 tandem repeat.[6] Subjects for whom imputation could not be performed with high (⩾90%) probability (122 European Americans and 58 African Americans) were excluded from analyses requiring isoform data. All results were qualitatively similar using more or less stringent assignment probabilities.

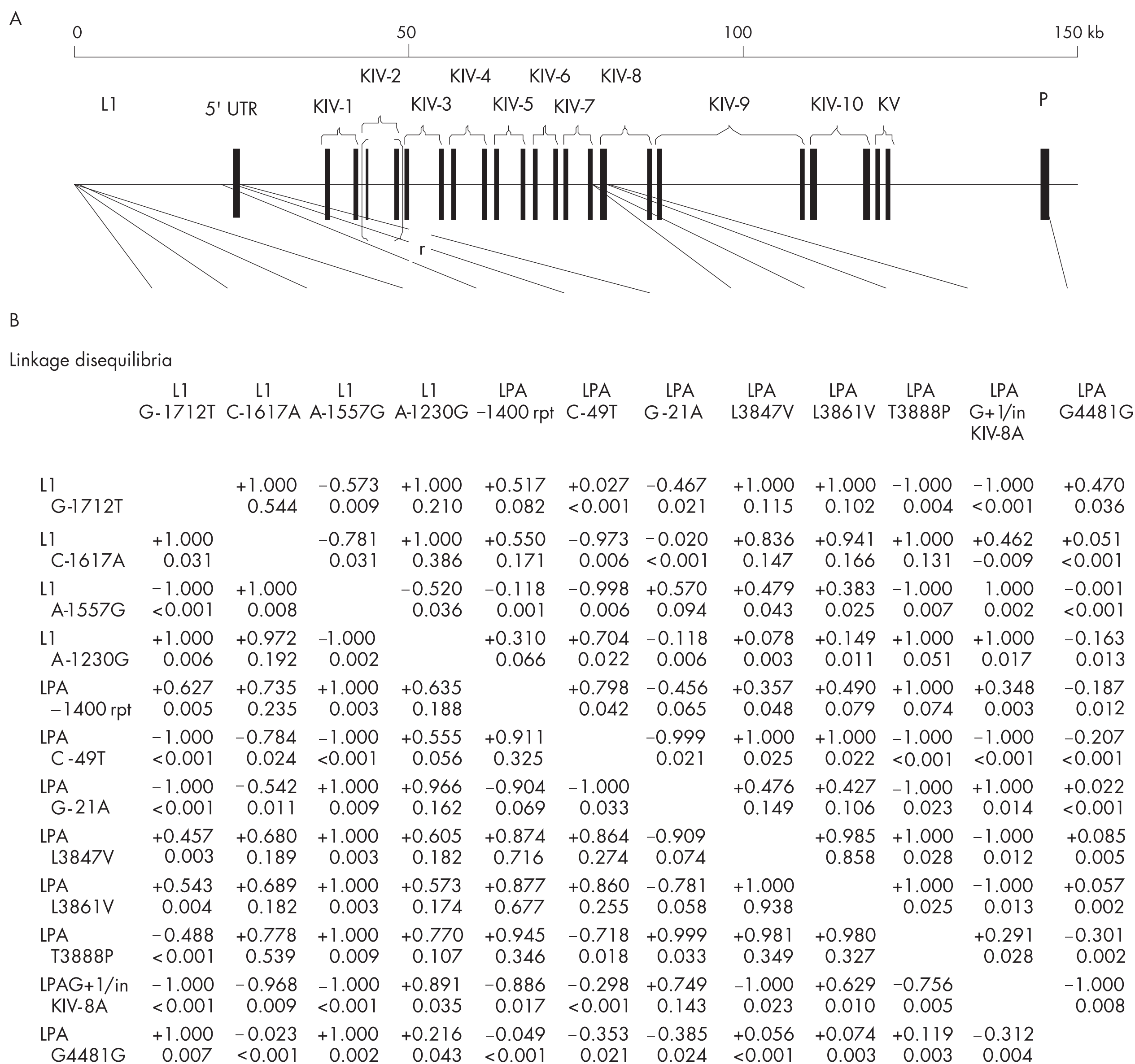
Pairwise linkage disequilibrium was estimated using Arlequin.[29]

## RESULTS

As in other populations undergoing dialysis,[30] the Lp(a) level was increased in CHOICE, but isoform distributions and associations with Lp(a) level were consistent with those in the general population.[31] Isoform distributions were similar in European Americans (median size of predominantly expressed isoform (interquartile range (IQR)) 25 (18–29) KIV-2 units) and African Americans (23 (20–27) units), and Lp(a) level was inversely associated with isoform size in both populations (fig 2A). An increase in the Lp(a) level was greatest among isoforms with 18–22 and 26–27 KIV-2 units in African Americans compared with European Americans (fig 2A). After direct adjustment to the isoform distribution in European Americans, the Lp(a) level remained 2.23-fold higher in African Americans ($p<0.001$).

We evaluated isoform-adjusted associations with the Lp(a) level initially locus-by-locus, simulating the previous studies. Table 1 gives the directions of association with the Lp(a) level in previous studies and allele frequencies in this study. All five variants previously associated with the Lp(a) level were significant in European Americans or African Americans. Only two of these (LPA T3888P and LPA G+1/inKIV-8A) were significant in both populations, whereas two (LPA –1400 rpt (10-repeat allele) and LPA C-49T) were significant only in European Americans and one (LPA G-21A) only in African Americans. Two SNPs (LPA L3847V and LPA L3861V) that previously were not associated with the Lp(a) level in Europeans were significant in European Americans, as was one novel SNP (L1 A-1230G). However, three regions of closely spaced (<5 kb) loci comprised all but one of the genotyped variants (fig 1A). There was strong linkage disequilibrium in these regions. Overall, three large blocks were identified in this region, consistent with HapMap data in the European population (fig 1B). These results suggested that the underlying haplotype structure could confound single-locus analysis.

To estimate independent associations with Lp(a) level, we next considered all loci simultaneously using population-stratified stepwise regression.[4] Results were identical using forward addition ($p<0.05$ for inclusion) or backward elimination ($p\geqslant0.05$ for exclusion) procedures (isoform effects were retained in all models). LPA G-21A, LPA T3888P and LPA G+1/inKIV-8A were independently associated with isoform-adjusted Lp(a) level in each population, whereas no other loci were significant in either (table 2). Directions of association were consistent across populations. The promoter variant LPA G-21A predicted Lp(a) increases of 60% in European Americans ($p = 0.003$) and 42% in African Americans ($p = 0.001$); the non-conservative LPA T3888P was associated with Lp(a) reductions of 34% in European Americans ($p = 0.018$) and 56% in African Americans ($p = 0.007$); and the splice-site LPA G+1/inKIV-8A predicted Lp(a) reductions of 71% in European




B

Linkage disequilibria

| | L1 G-1712T | L1 C-1617A | L1 A-1557G | L1 A-1230G | LPA −1400 rpt | LPA C-49T | LPA G-21A | LPA L3847V | LPA L3861V | LPA T3888P | LPA G+1/in KIV-8A | LPA G4481G |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| L1 G-1712T | | +1.000<br>0.544 | −0.573<br>0.009 | +1.000<br>0.210 | +0.517<br>0.082 | +0.027<br><0.001 | −0.467<br>0.021 | +1.000<br>0.115 | +1.000<br>0.102 | −1.000<br>0.004 | −1.000<br><0.001 | +0.470<br>0.036 |
| L1 C-1617A | +1.000<br>0.031 | | −0.781<br>0.031 | +1.000<br>0.386 | +0.550<br>0.171 | −0.973<br>0.006 | −0.020<br><0.001 | +0.836<br>0.147 | +0.941<br>0.166 | +1.000<br>0.131 | +0.462<br>−0.009 | +0.051<br><0.001 |
| L1 A-1557G | −1.000<br><0.001 | +1.000<br>0.008 | | −0.520<br>0.036 | −0.118<br>0.001 | −0.998<br>0.006 | +0.570<br>0.094 | +0.479<br>0.043 | +0.383<br>0.025 | −1.000<br>0.007 | 1.000<br>0.002 | −0.001<br><0.001 |
| L1 A-1230G | +1.000<br>0.006 | +0.972<br>0.192 | −1.000<br>0.002 | | +0.310<br>0.066 | +0.704<br>0.022 | −0.118<br>0.006 | +0.078<br>0.003 | +0.149<br>0.011 | +1.000<br>0.051 | +1.000<br>0.017 | −0.163<br>0.013 |
| LPA −1400 rpt | +0.627<br>0.005 | +0.735<br>0.235 | +1.000<br>0.003 | +0.635<br>0.188 | | +0.798<br>0.042 | −0.456<br>0.065 | +0.357<br>0.048 | +0.490<br>0.079 | +1.000<br>0.074 | +0.348<br>0.003 | −0.187<br>0.012 |
| LPA C-49T | −1.000<br><0.001 | −0.784<br>0.024 | −1.000<br><0.001 | +0.555<br>0.056 | +0.911<br>0.325 | | −0.999<br>0.021 | +1.000<br>0.025 | +1.000<br>0.022 | −1.000<br><0.001 | −1.000<br><0.001 | −0.207<br><0.001 |
| LPA G-21A | −1.000<br><0.001 | −0.542<br>0.011 | +1.000<br>0.009 | +0.966<br>0.162 | −0.904<br>0.069 | −1.000<br>0.033 | | +0.476<br>0.149 | +0.427<br>0.106 | −1.000<br>0.023 | +1.000<br>0.014 | +0.022<br><0.001 |
| LPA L3847V | +0.457<br>0.003 | +0.680<br>0.189 | +1.000<br>0.003 | +0.605<br>0.182 | +0.874<br>0.716 | +0.864<br>0.274 | −0.909<br>0.074 | | +0.985<br>0.858 | +1.000<br>0.028 | −1.000<br>0.012 | +0.085<br>0.005 |
| LPA L3861V | +0.543<br>0.004 | +0.689<br>0.182 | +1.000<br>0.003 | +0.573<br>0.174 | +0.877<br>0.677 | +0.860<br>0.255 | −0.781<br>0.058 | +1.000<br>0.938 | | +1.000<br>0.025 | −1.000<br>0.013 | +0.057<br>0.002 |
| LPA T3888P | −0.488<br><0.001 | +0.778<br>0.539 | +1.000<br>0.009 | +0.770<br>0.107 | +0.945<br>0.346 | −0.718<br>0.018 | +0.999<br>0.033 | +0.981<br>0.349 | +0.980<br>0.327 | | +0.291<br>0.028 | −0.301<br>0.002 |
| LPAG+1/in KIV-8A | −1.000<br><0.001 | −0.968<br>0.009 | −1.000<br><0.001 | +0.891<br>0.035 | −0.886<br>0.017 | −0.298<br><0.001 | +0.749<br>0.143 | −1.000<br>0.023 | +0.629<br>0.010 | −0.756<br>0.005 | | −1.000<br>0.008 |
| LPA G4481G | +1.000<br>0.007 | −0.023<br><0.001 | +1.000<br>0.002 | +0.216<br>0.043 | −0.049<br><0.001 | −0.353<br>0.021 | −0.385<br>0.024 | +0.056<br><0.001 | +0.074<br>0.003 | +0.119<br>0.003 | −0.312<br>0.004 | |

**Figure 1** Genotyped polymorphic sites in the *LPA* gene and linkage disequilibria. (A) Organisation of the *LPA* region and locations of genotyped variants. (B) Pairwise linkage disequilibria among genotyped variants in European Americans (lower triangle) and African Americans (upper triangle). Linkage disequilibrium was measured using both the signed normalised disequilibrium coefficient D′ taken between European American minor alleles (upper value for each locus pair) and $r^2$ (lower value) because of the known property that D′ attains extreme values for low-frequency alleles. Physically clustered sites are boxed. For LPA −1400 rpt, alleles were dichotomised as eight repeats versus all others. L1, enhancer-containing retrotransposon upstream of *LPA* (open box); 5′UTR, 5′ untranslated region of *LPA*; KIV-1 to 10, Kringle IV-encoding regions, types 1–10; KV, Kringle V-encoding region; P, protease (catalytically inactive)-encoding domain (closed boxes represent exons). The 5.6-kb tandem exon repeat that encodes KIV-2 occurs 8–43 times in the study population.

Americans ($p = 2.8\times10^{-8}$) and 88% in African Americans ($p = 1\times10^{-5}$).

As Lp(a)-increasing LPA G-21A was less common in European Americans than in African Americans (16.3% *v* 41.7%), and Lp(a)-lowering LPA T3888P and LPA G+1/inKIV-8A were more common in European Americans (14.3% *v* 2.4% and 4.7% *v* 0.7%), all three SNPs contributed to higher African American Lp(a). Strong population-specific association with isoform size limited contributions to isoform subsets. LPA G-21A was rare among 24–30 KIV-2 unit isoforms in European Americans (fig 2B) but was strongly associated with them in African Americans (fig 2C), generating part of the large between-population difference in Lp(a) level across 24–30 KIV-2 unit isoforms (fig 2A). LPA T3888P occurred almost exclusively on isoforms with 17–21 KIV-2 units in European Americans (fig 2B), but was rare on these and other isoforms in African Americans (fig 2C), partially accounting for differences in Lp(a) levels among isoforms with 18–21 KIV-2 units (fig 2A). The splice site LPA G+1/inKIV-8A was strongly associated with non-expressed isoforms in European Americans (fig 2B) and was rare in African Americans, occurring only on isoforms with 21 and 28 KIV-2 units (fig 2C).

To quantify the contributions of isoforms and these three SNPs to Lp(a) differences within and between populations, we estimated log-Lp(a) level associations in a series of models:

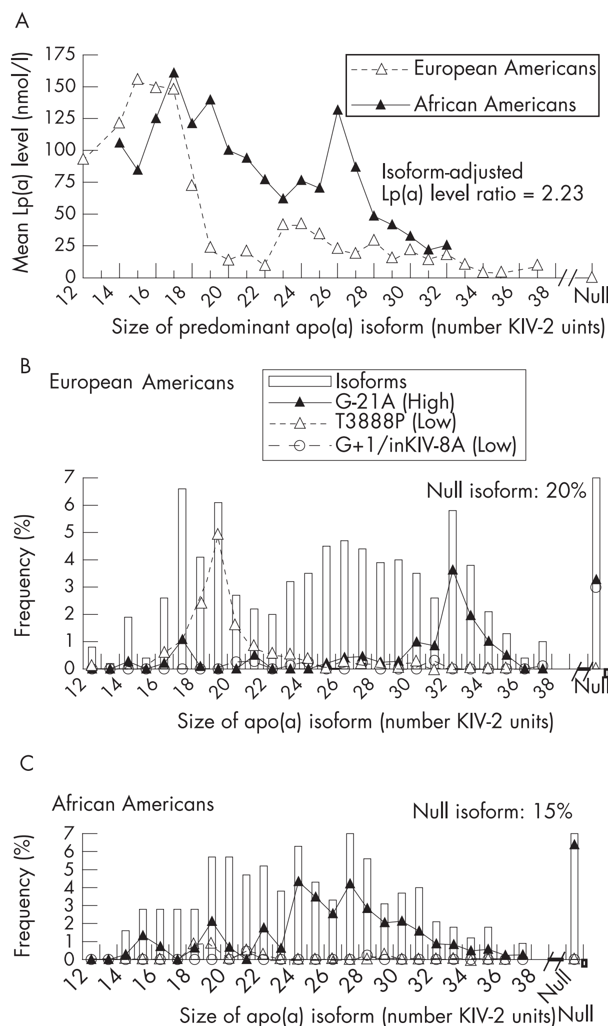



**Figure 2** Relationship between lipoprotein (Lp)(a) level and apolipoprotein (apo)(a) isoforms and between *LPA* single-nucleotide polymorphisms (SNPs) and apo(a) isoforms. (A) Mean Lp(a) level versus isoform size. Log-transformed Lp(a) was directly adjusted to the distribution of both predominant and weaker isoforms in European Americans, and the marginal adjusted mean for each stratum of predominant isoform was reverse transformed and plotted. The ''null'' category refers to non-expressed isoforms. Data for isoform strata with <3 subjects are excluded. (B) Distribution of LPA G-21A, LPA T3888P and LPA G+1/inKIV-8A on isoforms in European Americans. In multiple-locus analysis, LPA G-21A was associated with higher Lp(a) level, and LPA T3888P and LPA G+1/inKIV-8A with lower Lp(a) level in both populations. The range of isoform sizes corresponds to panel A (note that panel A is derived from individual-level measurements and panel B from allelic-level measurements). The bar for null isoforms is truncated and the frequency provided in the figure. (C) Distribution of LPA G-21A, LPA T3888P and LPA G+1/inKIV-8A on isoforms in African Americans.

firstly, with population affiliation (African American *v* European American) only; secondly, adding isoforms; and thirdly, adding SNPs (table 3). In model 2, isoforms explained 35.3% of the total Lp(a) variance and modestly reduced the population Lp(a) ratio (from 2.5 to 2.23-fold higher among African Americans, a 13% reduction) and between-population Lp(a) variance (from 8.7% to 5.3%). In model 3, the SNPs explained 7% of the total Lp(a) variance and reduced the isoform-adjusted population Lp(a) ratio by 60% (from 2.23 to 1.37-fold (95% CI 1.06 to 1.78) higher among African Americans) and nearly eliminated the remaining between-population Lp(a) variance (reduced from 5.3% to 0.5%).

## DISCUSSION

In this study of patients receiving incident dialysis, we observed an association between Lp(a) level and several *LPA* sequence variations (LPA G-21A, LPA T3888P and LPA G+1/inKIV-A). Moreover, these three polymorphisms together accounted for most of the unexplained Lp(a) increases in African Americans relative to European Americans. These results should not be affected by the increased levels of Lp(a) in patients receiving dialysis, as observed in both our study and previous studies,[24] since the relationship between genotypes and Lp(a) levels has not been reported to differ by the presence of end-stage renal disease (ESRD). Although the mechanism of the excess Lp(a) increase in ESRD is unknown, the observations of similar isoform distributions among patients with ESRD and controls,[24] and a substantial reduction of Lp(a) levels after renal transplantation[27] suggest that the Lp(a) increase observed in all patients receiving dialysis, compared with the general population, is mainly an effect of ESRD (or dialysis), not a cause. Additionally, marked differences in aortic-renal vein Lp(a) levels in patients undergoing coronary angiography[1] and the demonstration of apo(a) fragments in urine[23] support renal involvement in Lp(a) catabolism, providing a possible explanation for the increase.

The Lp(a)-level associations observed for LPA G-21A, LPA T3888P and LPA G+1/inKIV-A in multiple-locus analysis are consistent with previous functional data. Lp(a)-increasing

**Table 1** Single-locus associations with apolipoprotein(a) isoform-adjusted lipoprotein(a) level*

| Locus | Previous studies† | | This study | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Europeans or European Americans | Africans | European Americans (n = 534) | | | African Americans (n = 249) | | |
| | | | Frequency | Lp(a) ratio‡ | p Value | Frequency | Lp(a) ratio‡ | p Value |
| L1 G-1712T | – | — | 0.5 | 0.59 | 0.485 | 13.6 | 1.08 | 0.686 |
| L1 C-1617A | – | — | 16.3 | 0.80 | 0.177 | 21.3 | 0.89 | 0.473 |
| L1 A-1557G | – | — | 0.5 | 2.66 | 0.158 | 16.7 | 1.25 | 0.097 |
| L1 A-1230G | – | — | 49.2 | 0.76§ | 0.008 | 39.7 | 0.93 | 0.493 |
| LPA −1400 rpt: 6 | 10 rpt ↓ [25], [12] | NS[25] | 0.0 | NA | NA | 1.6 | 0.69 | 0.407 |
| 7 | 11 rpt ↓ [32]?NS[15] | | 0.1 | 1.99 | 0.613 | 14.3 | 1.35 | 0.105 |
| 8 | | | 68.9 | Reference | NA | 69.5 | Reference | NA |
| 9 | | | 14.8 | 0.78 | 0.107 | 6.5 | 0.69 | 0.158 |
| 10 | | | 14.7 | 0.61§ | 0.009 | 5.4 | 0.71 | 0.258 |
| 11 | | | 1.5 | 0.78 | 0.609 | 1.4 | 0.37 | 0.075 |
| 12 | | | 0.0 | NA | NA | 1.4 | 0.88 | 0.838 |
| LPA C-49T | NS[12 32 33 2] | ↓ [33] | 15.7 | 0.73§ | 0.029 | 3.5 | 1.56 | 0.112 |
| LPA G-21A | ↑ [12 32], NS[15] | — | 16.3 | 1.16 | 0.317 | 41.7 | 1.38§ | 0.004 |
| LPA L3847V | NS[12] | — | 32.8 | 0.78§ | 0.036 | 52.9 | 1.03 | 0.757 |
| LPA L3861V | NS[12] | — | 33.9 | 0.77§ | 0.023 | 55.8 | 1.05 | 0.590 |
| LPA T3888P | ↓ [12 34]NS[2] | NA[2] | 14.3 | 0.66§ | 0.029 | 2.9 | 0.30* | $6.381 \times 10^{-5}$ |
| LPA G+1/inKIV-8A | ↓ [34] | NA[34] | 4.7 | 0.40§ | $2.000 \times 10^{-5}$ | 0.7 | 0.09* | $2.539 \times 10^{-5}$ |
| LPA G4481G | – | – | 46.9 | 1.03 | 0.796 | 47.1 | 0.96 | 0.647 |

Lp(a), lipoprotein(a); NA, not applicable because the allele was not present in the given population; rpt, repeats.
*Associations and frequencies correspond to the European or European American minor allele (for single nucleotide polymorphisms) or to the stated allele (for LPA –1400 rpt).
†Arrows denote the direction of allelic association with apolipoprotein(a) isoform-adjusted Lp(a) level.[25 12 32 15 33 34 4 21]
‡The Lp(a) ratio is the exponentiated coefficient from linear regression of log(Lp(a) level+1) on apo(a) isoforms and the number of allele copies (0, 1 or 2).
§Associations with $p<0.05$.

**Table 2** Multiple-locus associations with apolipoprotein(a) isoform-adjusted lipoprotein(a) level*

| Locus | European Americans | | | African Americans | | |
|---|---|---|---|---|---|---|
| | Frequency† | Lp(a) ratio (95% CI) | p | Frequency† | Lp(a) ratio (95% CI) | p |
| LPA G-21A | 16.3 | 1.60 (1.18 to 2.18) | 0.003 | 41.7 | 1.42 (1.16 to 1.73) | 0.001 |
| LPA T3888P | 14.3 | 0.66 (0.46 to 0.93) | 0.018 | 2.9 | 0.44 (0.24 to 0.80) | 0.007 |
| LPA G+1/inKIV-8A | 4.7 | 0.29 (0.18 to 0.46) | 2.773E-8 | 0.7 | 0.12 (0.04 to 0.37) | 8.852E-6 |

Lp(a), lipoprotein(a).
*Associations are displayed for single nucleotide polymorphisms significant in stepwise regression analysis of all loci in either population.
†Frequencies are repeated from table 1.

**Table 3** Lipoprotein(a) level explained by population, isoforms, and single-nucleotide polymorphisms*

| Model | Predictor | Lp(a) variance explained† (%) | | | African American:European American Lp(a) ratio‡ |
|---|---|---|---|---|---|
| | | European Americans | African Americans | Total population | |
| 1 | Population | NA | NA | 8.7 | 2.50 |
| 2 | Population | NA | NA | 5.3 | 2.23 |
| | Isoforms | 44.0 | 61.6 | 35.3 | |
| 3 | Population | NA. | NA | 0.5 | 1.37 |
| | Isoforms | 44.0 | 64.2 | 37.4 | |
| | SNPs | 5.7 | 10.5 | 7.0 | |

Lp(a), lipoprotein(a); NA, not applicable because the allele was not present in the given population.
*Population affiliation is African American versus. European American. SNPs are LPA G-21A, LPA T3888P and LPA G+1/inKIV-8A.
†The Lp(a) variance explained is the partial $R^{12}$ after inclusion of the other predictors in the linear regression model of log(Lp(a) level +1). Models for the European American and African American columns are population stratified, whereas the model for the Total population column is not.
‡The African American:European American Lp(a) ratio is the exponentiated regression coefficient for population affiliation, representing the Lp(a) level ratio comparing African Americans with European Americans adjusted for the other predictors in the model.

G-21A raised *LPA* promoter activity by 90% in a chloramphenicol acetyltransferase assay in transfected HepG2 cells.[36] Lp(a)-lowering T3888P occurs in apo(a) KIV-8, which is required for lysine-mediated apo(a)–apolipoprotein B-100 interaction during Lp(a) assembly,[37] and reduced maximal attainable binding of Lp(a) to lysine by 75% in vitro.[38] Lp(a)-lowering G+1/inKIV-A markedly impairs Lp(a) assembly.[17] Although LPA C-49T reduced luciferase expression by about 50% in transfected HepG2 cells[36, 39] through early translation initiation and termination,[39] we did not observe a significant independent association with Lp(a) level in either population.

The Lp(a) reduction of 27% in single-locus analysis of European Americans was not explained by association with LPA G-21A, LPA T3888P or LPA G+1/inKIV-8A and approached significance in multiple-locus analysis ($p = 0.092$). However, association with 65% higher Lp(a) in multiple-locus analysis of African Americans ($p = 0.062$) suggests that any effect on Lp(a) level is modest, or can be modified by genetic or environmental factors.

LPA G-21A, LPA T3888P and LPA G+1/inKIV-8A together accounted for most of the unexplained Lp(a) increases in African Americans relative to European Americans. Incorporation of these variants into ongoing epidemiological studies in the general population should be undertaken to better elucidate the interactions of *LPA* sequence, apo(a) isoforms and Lp(a) levels in the development of coronary disease. It is notable that each of these SNPs had a frequency of $<20\%$ in one or both populations and two (LPA T3888P and LPA G+1/inKIV-8A) had a frequencies of $<5\%$ in African Americans. Thus, alleles that are not universally common may explain large between-population differences.

Whether the allelic spectrum of between-population variation for other risk factors and diseases is similar to that for Lp(a) is unknown. In some cases, specific candidate variants (eg, those consistently associated with the phenotype in one population) may be available, and could be assessed simultaneously in multiple populations to quantify their contributions to between-population variations. The situation may be more challenging when potentially causal variants are not available, as current mapping resources are not be ideally suited for identifying variants that explain between-population differences. For example, The Single Nucleotide Polymorphism database has low sensitivity for alleles that are not universally common.[2] Expanding publicly available genetic analysis resources to include population-specific variation could improve prospects for explaining between-population differences.

## ACKNOWLEDGEMENTS

We thank SJ O'Brien for discussion throughout this project; M Levasseur, S Shrestha, M Subleski, and A Truelove for genotyping assistance; Y Liu, the CHOICE staff and Dialysis Clinics Incorporated for the clinical data; JC Long for the 3LOCUS program; and the CHOICE participants.

### Authors' affiliations

**J-P Chretien, J Coresh, Y Berthier-Schaad, W H L Kao, N E Fink, M J Klag,** Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA
**S M Marcovina, F Giaculli,** Department of Medicine, Northwest Lipid Research Laboratories, University of Washington, Seattle, Washington, USA
**M W Smith,** Laboratory of Genomic Diversity and Basic Research Program, Science Applications International Corp-Frederick, National Cancer Institute at Frederick, Frederick, Maryland, USA

This study was supported by grants R01-HS-08365 (AHRQ), R01-HL-62985 (NHLBI) and R01-DK-07024 (NIDDK), K24-DK-02856 (NIDDK; Klag), 01-40197N (AHA Established Investigator—JC), K01-DK067207 (NIDDK; WHLK), National Center for Research Resources (NIH) GCRC grant M01-RR00052. This study was funded in part with federal funds from the National Cancer Institute, National Institutes of Health, under Contract Number NO1-CO-12400. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products or organisations imply endorsement by the US government.

Competing interests: None declared.

## REFERENCES

1 **Rosenberg NA**, Prtichard JK, Weber JL, *et al.* Genetic structure of human populations. *Science* 2002;**298**:2381–5.
2 **Carlson CS**, Eberle MA, Rieder MJ, *et al.* Additional SNPs and linkage-disequilibrium analyses are necessary for whole-genome association studies in humans. *Nat Genet* 2003;**33**:518–21.
3 **Risch N**, Burchard E, Ziv E, *et al.* Categorization of humans in biomedical research: genes, race and disease. *Genome Biol* 2002;**3**:comment2007.
4 **Cordell HJ**, Clayton DG. A unified stepwise regression procedure for evaluating the relative effects of polymorphisms within a gene using case/control or family data: application to HLA in type 1 diabetes. *Am J Hum Genet* 2002;**70**:124–41.
5 **Reich DE**, Lander ES. On the allelic spectrum of human disease. *Trends Genet* 2001;**17**:502–10.
6 **Kraft HG**, Lingenhel A, Pang RW, *et al.* Frequency distributions of apolipoprotein(a) kringle IV repeat alleles and their effects on lipoprotein(a) levels in Caucasian, Asian, and African populations: the distribution of null alleles is non-random. *Eur J Hum Genet* 1996;**4**:74–87.
7 **Marcovina SM**, Albers JJ, Wijsman E, *et al.* Differences in Lp[a] concentrations and apo[a] polymorphs between black and white Americans. *J Lipid Res* 1996;**37**:2569–85.
8 **Danesh J**, Collins R, Peto R. Lipoprotein(a) and coronary heart disease. Meta-analysis of prospective studies. *Circulation* 2000;**102**:1082–5.
9 **Tsimikas S**, Brilakis ES, Miller ER, *et al.* Oxidized phospholipids, Lp(a) lipoprotein, and coronary artery disease. *N Engl J Med* 2005;**353**:46–57.
10 **Scanu AM**. Lipoprotein(a). Link between structure and pathology. *Ann Epidemiol* 1992;**2**:407–12.
11 **Edelberg JM**, Gonzalez-Gronow M, Pizzo SV. Lipoprotein(a) inhibition of plasminogen activation by tissue-type plasminogen activator. *Thromb Res* 1990;**57**:155–62.
12 **Boerwinkle E**, Leffert CC, Lin J, *et al.* Apolipoprotein(a) gene accounts for greater than 90% of the variation in plasma lipoprotein(a) concentrations. *J Clin Invest* 1992;**90**:52–60.
13 **Mooser V**, Scheer D, Marcovina SM, *et al.* The Apo(a) gene is the major determinant of variation in plasma Lp(a) levels in African Americans. *Am J Hum Genet* 1997;**61**:402–17.
14 **Prins J**, Leus FR, Bouma BN, *et al.* The identification of polymorphisms in the coding region of the apolipoprotein (a) gene—association with earlier identified polymorphic sites and influence on the lipoprotein (a) concentration. *Thromb Haemost* 1999;**82**:1709–17.
15 **Brunner C**, Lobentanz E-M, Petho-Schramm A, *et al.* The number of identical kringle IV repeats in apolipoprotein(a) affects its processing and secretion by HepG2 cells. *J Biol Chem* 1996;**271**:32403–10.
16 **Mooser V**, Mancini FP, Bopp S, *et al.* Sequence polymorphisms in the apo(a) gene associated with specific levels of Lp(a) in plasma. *Hum Mol Genet* 1995;**4**:173–81.
17 **Ogorelkova M**, Gruber A, Utermann G. Molecular basis of congenital lp(a) deficiency: a frequent apo(a) 'null' mutation in caucasians. *Hum Mol Genet* 1999;**8**:2087–96.
18 **Ogorelkova M**, Kraft HG, Ehnholm C, *et al.* Single nucleotide polymorphisms in exons of the apo(a) kringles IV types 6 to 10 domain affect Lp(a) plasma concentrations and have different patterns in Africans and Caucasians. *Hum Mol Genet* 2001;**10**:815–24.
19 **Puckey LH**, Lawn RM, Knight BL. Polymorphisms in the apolipoprotein(a) gene and their relationship to allele size and plasma lipoprotein(a) concentration. *Hum Mol Genet* 1997;**6**:1099–107.
20 **Trommsdorff M**, Kochl S, Lingenhel A, *et al.* A pentanucleotide repeat polymorphism in the 5′ control region of the apolipoprotein(a) gene is associated with lipoprotein(a) plasma concentrations in Caucasians. *J Clin Invest* 1995;**96**:150–7.
21 **Kraft HG**, Windegger M, Menzel HJ, *et al.* Significant impact of the +93 C/T polymorphism in the apolipoprotein(a) gene on Lp(a) concentrations in Africans but not in Caucasians: confounding effect of linkage disequilibrium. *Hum Mol Genet* 1998;**7**:257–64.
22 **Scholz M**, Kraft HG, Lingenhel A, *et al.* Genetic control of lipoprotein(a) concentrations is different in Africans and Caucasians. *Eur J Hum Genet* 1999;**7**:169–78.
23 **Rotimi CN**, *et al.* Serum distribution of lipoprotein(a) in African Americans and Nigerians: potential evidence for a genotype-environmental effect. *Genet Epidemiol* 1997;**14**:157–68.
24 **Powe NR**, Klag MJ, Sadler JH, *et al.* Choices for healthy outcomes in caring for end stage renal disease. *Semin Dial* 1996;**9**:9–11.
25 **Astor BC**, Eustace JA, Klag MJ, *et al.* Race-specific association of lipoprotein(a) with vascular access interventions in hemodialysis patients: the CHOICE Study. *Kidney Int* 2002;**61**:1115–23.

26 **Underhill PA**, Jin L, Lin AA, *et al*. Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. *Genome Res* 1997;**7**:996–1005.
27 **Yang Z**, Boffelli D, Boonmark N, *et al*. Apolipoprotein(a) gene enhancer resides within a LINE element. *J Biol Chem* 1998;**273**:891–7.
28 **Lindblad-Toh K**, Winchester E, Daly MJ, *et al*. Large-scale discovery and genotyping of single-nucleotide polymorphisms in the mouse. *Nat Genet* 2000;**24**:381–6.
29 Arlequin Version 2. A software for population genetics data analysis, Genetics and Biometry Laboratory, Switzerland: University of Geneva, 2000.
30 **Kronenberg F**, Konig P, Neyer U, *et al*. Multicenter study of lipoprotein(a) and apolipoprotein(a) phenotypes in patients with end-stage renal disease treated by hemodialysis or continuous ambulatory peritoneal dialysis. *J Am Soc Nephrol* 1995;**6**:110–20.
31 **Longenecker JC**, Klag MJ, Marcovina SM, *et al*. High lipoprotein(a) levels and small apolipoprotein(a) size prospectively predict cardiovascular events in dialysis patients. *J Am Soc Nephrol* 2005;**16**:1794–802.
32 **Brazier L**, Tiret L, Luc G, *et al*. Sequence polymorphisms in the apolipoprotein(a) gene and their association with lipoprotein(a) levels and myocardial infarction. The ECTIM Study. *Atherosclerosis* 1999;**144**:323–33.
33 **Burchard EG**, Ziv E, Coyle N, *et al*. The importance of race and ethnic background in biomedical research and clinical practice. *N Engl J Med* 2003;**348**:1170–5.
34 **Chakravarti A**. Population genetics—making sense out of sequence. *Nat Genet* 1999;**21**(Suppl 1):56–60.
35 **Suzuki K**, Kuriyama M, Saito T, *et al*. Plasma lipoprotein(a) levels and expression of the apolipoprotein(a) gene are dependent on the nucleotide polymorphisms in its 5′-flanking region. *J Clin Invest* 1997;**99**:1361–6.
36 **Gabel BR**, Koschinsky ML. Sequences within apolipoprotein(a) kringle IV types 6–8 bind directly to low-density lipoprotein and mediate noncovalent association of apolipoprotein(a) with apolipoprotein B-100. *Biochemistry* 1998;**37**:7892–8.
37 **Prins J**, Leus FR, van der Hoek YY, *et al*. The identification and significance of a Thr→Pro polymorphism in kringle IV type 8 of apolipoprotein(a). *Thromb Haemost* 1997;**77**:949–54.
38 **Zysow BR**, Lindahl GE, Wade DP, *et al*. C/T polymorphism in the 5′ untranslated region of the apolipoprotein(a) gene introduces an upstream ATG and reduces in vitro translation. *Arterioscler Thromb Vasc Biol* 1995;**15**:58–64.

# ECHO

## Intestinal permeability in families with Crohn's disease

Please visit the *Journal of Medical Genetics* website [www.jmedgenet.com] for a link to the full text of this article.

A genetically impaired intestinal barrier function has long been suspected to be a predisposing factor for Crohn's disease (CD). Recently, mutations of the capsase recruitment domain family, member 15 (*CARD15*), have been identified and associated with CD.

To test the association of *CARD15* with an impaired intestinal barrier, the authors studied 128 patients with quiescent CD, 129 first degree relatives, 66 non-related household members, and 96 healthy controls. There were three main findings. Healthy first degree relatives of patients with CD showed increased permeability in contrast with unrelated household members and controls. Secondly, the prevalence of the *CARD15* 3020insC mutation was similar in first degree relatives and CD patients and higher compared with controls. Thirdly, in healthy first degree relatives, high mucosal permeability and the presence of a *CARD15* 3020insC mutation were significantly associated.

The data indicate a genetic rather than an environmental basis for the intestinal barrier dysfunction in CD. The authors speculate that the *CARD15* 3020insC mutation could be one genetic factor involved in impairment of intestinal barrier function. However, it is obvious that this is not the only factor.

The association between the *CARD15* gene mutation and intestinal hyperpermeability in healthy first degree CD relatives may be one step towards the identification of other target genes involved in similar processes. Considering the high intestinal bacterial load in CD patients compared with controls, early barrier dysfunction gains special significance with respect to the pathogenesis of CD. However, longitudinal studies in CD families are necessary to investigate which additional factors may lead to the outbreak of the disease.

▲ Buhner S, *et al. Gut* 2006;**55**:342–7.