A Common \textit{LPA} Null Allele Associates With Lower Lipoprotein(a) Levels and Coronary Artery Disease Risk

Theodosios Kyriakou, Udo Seedorf, Anuj Goel, Jemma C. Hopewell, Robert Clarke, Hugh Watkins, Martin Farrall, on behalf of the PROCARDIS Consortium

Objective—Increased levels of lipoprotein(a) are a highly heritable risk factor for coronary artery disease (CAD). The genetic determinants of lipoprotein(a) levels are mainly because of genetic variation in the apolipoprotein(a) gene (\textit{LPA}). We have tested the association of a relatively common null allele of \textit{LPA} with lipoprotein(a) levels and CAD risk in a large case–control cohort. We have also examined how null allele genotyping complements apolipoprotein(a) isoform typing to refine the relationship between \textit{LPA} isoform size and circulating lipoprotein(a) levels.

Approach and Results—The \textit{LPA} null allele (rs41272114) was genotyped in the PROCARDIS case–control cohort (4073 CAD cases and 4225 controls). Lipoprotein(a) levels were measured in 909 CAD cases and 922 controls; apolipoprotein(a) isoform size was estimated using sodium dodecyl sulfate–agarose gel electrophoresis and a high-throughput quantitative polymerase chain reaction–based method. Null carriers are common (null allele frequency, 3%) and have significantly lower circulating lipoprotein(a) levels (\(P=2.1\times10^{-10}\)) and reduced CAD risk (odds ratio, 0.79 [0.66–0.97]; \(P=0.023\)) compared with noncarriers. An additive allelic model of apolipoprotein(a) isoform size, refined by null allele genotype and quantitative polymerase chain reaction values, showed a sigmoid relationship with lipoprotein(a) levels, with baseline levels for longer isoform alleles and progressively higher levels of lipoprotein(a) for shorter isoform alleles.

Conclusions—The \textit{LPA} null allele (rs41272114) is associated with decreased circulating lipoprotein(a) levels and decreased CAD risk. Incorporating rs41272114 refined apolipoprotein(a) isoform size typing obtained by immunoblotting and quantitative polymerase chain reaction. A joint genomic and isoform analysis revealed details of the relationship between apolipoprotein(a) isoform size and circulating lipoprotein(a) level consistent with a threshold effect on lipoprotein secretion. (\textit{Arterioscler Thromb Vasc Biol.} \textbf{2014;}34:00–00.)

Key Words: coronary artery disease $\cdot$ lipoprotein(a)

Lipoprotein(a) is a complex lipoprotein particle found in the circulation of higher primates. It comprises a low-density lipoprotein-like particle with the addition of an apolipoprotein(a) molecule covalently bonded to apolipoprotein B.\textsuperscript{1} In humans, increased lipoprotein(a) levels in the blood have been associated with increased risk of cardiovascular disease.\textsuperscript{2,5}

Lipoprotein(a) levels vary substantially among individuals within a population, are uncorrelated with classic risk factors, and remain relatively constant throughout an individual’s lifespan.\textsuperscript{6} A large proportion of lipoprotein(a) level variation is inherited and is mainly attributed to genetic variation within the \textit{LPA} gene that encodes apolipoprotein(a), the characteristic protein moiety of the particle.\textsuperscript{7} Genetic variation in the \textit{LPA} gene is defined by a repetitive genomic region harboring 2 exons encoding one of the kringle domains (kringle IV [KIV]-2) in apolipoprotein(a), resulting in a highly polymorphic protein that includes multiple copies of the kringle domain ranging in number from 3 to >40.\textsuperscript{8} This variation has been proposed to directly affect circulating lipoprotein(a) levels by influencing the processing and secretion of apolipoprotein(a) isoforms from producing hepatocytes.\textsuperscript{9,10} The copy number of KIV-2 repeats is inversely related to levels with larger molecules, hindering efficient secretion and resulting in lower circulating lipoprotein(a) levels.\textsuperscript{5}

Depending on the population studied, a significant proportion of the genetic variation attributed to the \textit{LPA} gene is independent of the copy number variation (CNV).\textsuperscript{11} This is consistent with the observation that isoforms of the same size differ widely in concentration, suggesting that variation in the number of kringle domains is not the only important \textit{LPA} contributor to lipoprotein(a) quantitative variation. There are reports of single nucleotide polymorphisms (SNPs) associated with lipoprotein(a) levels, which either serve as proxy
SNP rs41272114 is a functional polymorphism that results in alternative splicing and a prematurely terminated protein molecule that lacks the domain that is normally covalently linked to apolipoprotein B and is unable to form a mature lipoprotein(a) particle. The functional effect of this relative common SNP, commonly referred to as a null allele, on circulating lipoprotein(a) levels has been found to be significant. However, this SNP has not been included in any of the widely used commercial high-throughput SNP arrays and thus has not been examined in the large-scale genetic epidemiological studies that have highlighted the role of lipoprotein(a) as an independent risk factor for coronary artery disease, ischemic stroke, and other atherosclerotic diseases.

The main aim of the present study was to evaluate the effect of the LPA null allele on circulating lipoprotein(a) levels and CAD risk in PROCARDIS, a European CAD case-control cohort. A second aim was to refine the relationship between apolipoprotein(a)/LPA isoform size/CNV copy number and lipoprotein(a) levels by combining null allele genotyping with kringle isoform typing based on immunoblotting and genomic methods.

Results

LPA Null Allele in the PROCARDIS Study: Effect on Circulating Lipoprotein(a) Levels and Association With CAD Risk.

In the PROCARDIS cohort, minor allele frequency of the null rs41272114 SNP was 0.03, and genotype frequencies were within Hardy–Weinberg equilibrium ($\chi^2=0.58; 1 \text{ df}; P=0.44$). Lipoprotein(a) levels were measured in 2000 individuals with a mean concentration of 21.6 mg/dL (SD, 24.6; median, 11.5 mg/dL; interquartile range, 6.03–27.7 mg/dL). After excluding individuals with missing genotypes, data were available for 1831 individuals (909 cases and 922 controls). The distribution of lipoprotein(a) levels in LPA null carriers and noncarriers is shown in Figure 1. There was a significant association between circulating lipoprotein(a) levels and LPA null carrier status ($P=2.1\times10^{-10}$) with a 5.25 mg/dL (95% confidence interval [CI], 3.75–6.75) reduction in median lipoprotein(a) levels in carriers. This association was also evident ($P=4.9\times10^{-10}$) in a joint analysis with 2 SNPs rs3798220 and rs10455872 that have been previously shown to be strongly associated with lipoprotein(a) levels; the LPA null allele was in linkage equilibrium ($r^2<0.002$) with both these SNPs. We then tested the variant for association with CAD risk in 8211 individuals (4022 cases and 4189 controls) and observed a significant protective effect with $\approx 20\%$ reduction in CAD risk (odds ratio, 0.79; 95% CI, 0.66–0.97; $P=0.023$) in carriers.

Correlation of Kringle Isoform Typing Using qPCR and Immunoblotting

Figure 2 shows the relationship between proteomic (isoform size) and genomic (relative quantity [RQ]) measurements of kringle copy number. Carriers and noncarriers of the LPA null allele show positive correlations of similar magnitude ($\rho=0.37, 95\% \text{ CI}, 0.19–0.54; \rho=0.44, 95\% \text{ CI}, 0.40–0.48$, respectively). The correlation between isoform size as measured by the 2 assay types when the null allele is not taken into account although positive is imperfect because of the inherent inaccuracies of the assay methods. This was not surprising and consistent with previous reports in smaller cohorts.
Regression Analysis of Lipoprotein(a) Levels

Figure 3 summarizes a regression analysis of median lipoprotein(a) levels on RQ levels, which shows a nonlinear association pattern with a pseudo-$R^2$ value of 0.08. Samples with RQ values >5th vigintile (≥225th percentile point) have similar lipoprotein(a) levels; samples with smaller RQ values show increasing lipoprotein(a) levels. The LPA null variant was a significant independent predictor of lipoprotein(a) levels ($P=2.7\times10^{-6}$), with carriers showing a 6.15 mg/dL (95% CI, 3.59–8.71) reduction in levels.

Figure 4 summarizes a regression analysis of median lipoprotein(a) levels on kringle isoform size; kringle alleles were assumed to have an additive effect on lipoprotein(a) levels (ie, a codominant inheritance model), and the model included information on the null LPA allele and RQ values. Lipoprotein(a) levels show a nonlinear pattern of association with kringle alleles with a pseudo-$R^2$ value of 0.29; alleles with ≥23 repeats had similar lipoprotein(a) levels (≈6 mg/dL per allele), whereas shorter alleles show progressively increasing lipoprotein(a) levels that plateau at ≈42 mg/dL. Covariation by RQ values was a significant independent predictor of lipoprotein(a) levels ($P=0.0002$), but RQ values have relatively little influence on the association pattern in comparison with the isoform data (Figure I in the online-only Data Supplement). The wide CIs for the quantitative polymerase chain reaction (qPCR) data (Figure 3) demonstrate the relative imprecision of this method compared with the isoform immunoblotting method (Figure 4). A supplementary regression analysis that disregards the null LPA allele shows a similar sigmoid pattern but with broader CIs (Figure II in the online-only Data Supplement).

Discussion

Genome-wide and gene-centric association studies have led to the identification of 48 genetic loci associated with CAD risk. The LPA locus on chromosome 6q shows the strongest genetic association with CAD discovered to date.20,21 In this study, we have confirmed the effect of a common LPA null allele on circulating lipoprotein(a) levels and shown for the first time a clear association with decreased CAD risk.

We and others have previously mapped quantitative trait loci for lipoprotein(a) and estimated that the vast majority (74 to >90%) of the total variation in circulating levels was specific to the LPA locus.7,22,23 Several studies sought to identify genetic variation in LPA that explains variation in lipoprotein(a) levels and further define the association with CAD risk beyond the contribution of the well-established copy number variation in the KIV-2 domains. This resulted mainly in identification of SNPs that act as proxy markers of the KIV-2 CNV3,24 or have no proven functional role.12,14,25

The rs41272114 single nucleotide polymorphism resulting in an LPA null allele was previously reported by Ogorelkova et al.23 The minor allele of this SNP causes disruption of a donor splice site and results in a truncated allelic form of apolipoprotein(a), including only kringle domains 1 to 7 (KIV 1–7). Because of the presumed functional importance of specific kringle domains, mutations and truncations related to these are expected to lead to loss of function. Expression of the alternative spliced transcript from the null allele in HepG2 cells showed that the truncated apolipoprotein(a) molecule is expressed and secreted, but because it lacks the KIV-9 domain it is not possible to covalently link to apolipoprotein B-100 and form a stable lipoprotein(a) particle.13 This truncated apolipoprotein(a) form was found to be extensively degraded in the plasma of carriers and was not detectable by immunoblotting.13 A null allele effect was clearly evident in the samples of the PROCARDIS study where we found a clear association between circulating levels and carrier status of rs41272114 with a strong effect on lipoprotein(a) levels per null allele, resulting in a significant reduction of median levels. The well-established association of circulating levels and CAD risk prompted us to investigate the effect of the LPA null allele on CAD risk in the entire PROCARDIS collection. As expected, reduced levels of lipoprotein(a) because of the null allele had a significant protective effect against risk of disease. This finding highlights the contribution and importance of the null allele as the only known functional polymorphism other than...
the CNV in modulating lipoprotein(a) levels and CAD disease risk. It is important to note that the rs41272114 SNP resulting in the LPA null allele is not included in any of the available genome-wide, gene-centric, or exome genotyping arrays. Furthermore, these arrays do not include proxy SNPs in strong linkage disequilibrium (r2<0.6), and rs41272114 is not present in the HapMap2 or HapMap3 imputation training sets that have been widely used to drive genome-wide association studies. The findings presented here emphasize the importance of including the null allele in genotyping strategies to be used in future genetic and epidemiological studies (Table).

It was apparent that null allele carrier status could also supplement and refine genotyping data of lipoprotein(a) isoform size because of the KIV-2 CNV. Agarose gel separation followed by immunoblotting and qPCR are 2 methods routinely used in the past for genotyping the LPA CNV in studies including large numbers of samples.

Lipoprotein(a) krimge isoform typing using immunoblotting is an expensive and low-throughput method that has limited accuracy. It has, however, the advantage that it can provide protein isoform size information encoded separately by each LPA allele (provided they are secreted at detectable levels). Because of the extensive apolipoprotein(a) gene size polymorphism, homozygotes are expected to be unusual in populations in Hardy–Weinberg equilibrium; a heterozygosity index of 94% was reported among Europeans.26 In the PROCARDIS study, a double band was observed in only a third of the samples assayed using immunoblotting. Null allele carrier status accounted for 12% of the single band immunoblots, with the others probably explained by nonsecreted large isoforms or low assay sensitivity.

A qPCR method using genomic DNA as template has been recently used to overcome the inherent restrictions of immunoblotting and estimate the number of krimge repeats in samples from large epidemiological cohorts.3,16 However, the high-throughput merits of speed and low cost of the qPCR method are tempered by the inability to distinguish the relative contributions of each allele because it can be used only for detection of the sum of the numbers of krimge domains. Furthermore, this method cannot identify alleles that do not produce a detectable apolipoprotein(a) molecule because of the steric impediment in secretion from producing cells.

Apolipoprotein(a) isoform information derived from immunoblotting and qPCR was supplemented by null allele carrier status to perform a regression analysis of circulating lipoprotein(a) levels in the PROCARDIS samples. Isoform alleles showed a sigmoid pattern of association with lipoprotein(a) levels, suggesting that there is a threshold effect such that relatively short alleles with ≤21 KIV-2 repeats are readily secreted by hepatocytes. This in vivo result is consistent with and provides detail to the inferences drawn from an in vitro study contrasting the expression of 11 and 22 KIV-2 repeat containing plasmids in HepG2 cells.9

In conclusion, the LPA null allele has a consistent and substantial effect in reducing circulating lipoprotein(a) levels and reducing CAD risk. Genotyping of the null allele enhances the data obtained by either immunoblotting or qPCR for measuring krimge KIV-2 copy number and significantly improves the characterization of LPA-lipoprotein(a) for quantitative genetic analyses.

Acknowledgments

We gratefully acknowledge technical assistance from Bertram Tambyrahaj at Leibniz-Institut für Arterioskleroseforschung an der Universität Münster, Münster, Germany.

Sources of Funding

The PROCARDIS study was supported by the European Community Sixth Framework Program (LSHM-CT-2007-037273), AstraZeneca, the British Heart Foundation, the Oxford British Heart Foundation Centre of Research Excellence, the Wellcome Trust (075491/Z/04), the Swedish Research Council, the Knut and Alice Wallenberg Foundation, the Swedish Heart–Lung Foundation, the Torsten and Ragnar Söderberg Foundation, the Strategie Cardiovascular Program of Karolinska Institutet and Stockholm County Council, and the Foundation for Strategic Research and the Stockholm County Council (560283).

Disclosures

None.

References


**Significance**

Concentration of circulating lipoprotein(a) is an independent heritable factor associated with risk of coronary artery disease. The rs41272114 single nucleotide polymorphism that results in a null allele is not included in any of the widely used commercial high-throughput single nucleotide polymorphism arrays and has not been previously examined in large-scale genetic epidemiological studies. Here, we show the association of this functional single nucleotide polymorphisms with reduced lipoprotein(a) levels and risk of coronary artery disease and highlight its importance for future genetic and epidemiological studies.
A Common LPA Null Allele Associates With Lower Lipoprotein(a) Levels and Coronary Artery Disease Risk
Theodosios Kyriakou, Udo Seedorf, Anuj Goel, Jemma C. Hopewell, Robert Clarke, Hugh Watkins and Martin Farrall

Arterioscler Thromb Vasc Biol. published online June 12, 2014; Arteriosclerosis, Thrombosis, and Vascular Biology is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2014 American Heart Association, Inc. All rights reserved.
Print ISSN: 1079-5642. Online ISSN: 1524-4636

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://atvb.ahajournals.org/content/early/2014/06/12/ATVBAHA.114.303462

Data Supplement (unedited) at:
http://atvb.ahajournals.org/content/suppl/2014/06/12/ATVBAHA.114.303462.DC1

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

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

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

Study population

CAD cases were recruited through the Precocious Coronary Artery Disease (PROCARDIS) study from four European countries (United Kingdom, Italy, Sweden, Germany). Ascertainment criteria for PROCARDIS cases were myocardial infarction or symptomatic acute coronary syndrome before age 66 years. Controls were recruited from the same country as the cases and had no personal or sibling history of CAD before age 66 years. The protocol was approved by the ethics committee at each participating centre, and all individuals provided written informed consent.

LPA SNP genotyping

Genotyping of rs41272114 was performed (KBioscience, Hoddesdon Herts, United Kingdom) using the proprietary competitive allele-specific polymerase chain reaction single-nucleotide polymorphism genotyping system with fluorescence resonance energy transfer quencher cassette oligos (KASPar).[1] The SNP is an A/G variation found on chr6:161,006,077 on human genome build hg19 with a minor allele (A) frequency of 0.03 in European populations (http://browser.1000genomes.org). The rs3798220 and rs10455872 SNPs were genotyped with the use of the HumanCVD BeadChip (Illumina)13 with the Infinium II assay as described previously[2, 3].

KIV-2 copy number variation typing

Isoelectric focusing followed by immunoblotting: Sodium dodecyl sulfate-agarose gel electrophoresis was used to fractionate the reduced plasma proteins according to their size before immunoblotting with an apo(a)-specific antibody. Apo(a) size was determined relative to human apo(a) isoform standards (Immuno AG, Austria) and a serum pool with pre-determined apo(a) isoforms and band size was evaluated by the AIDA software. An
extensive description of the method used has been reported elsewhere [4]. The accuracy of this method is ±2 repeats of the true isoform length.

qPCR: A multiplexed qPCR with genomic DNA as template was carried out using TaqMan® probes for LPA KIV-2 and RNaseP, an endogenous single-copy control gene with the Gene-expression qPCR master mix (Applied Biosystems®), in a real-time PCR system (Applied Biosystems®) following manufacturer's guidelines. The Taqman probe against LPA anneals specifically on exon 4 and has been described previously [5]. Threshold cycle values for each reaction were corrected for reaction efficiency. A relative quantity (RQ) for the sum of KIV-2 domain copies was calculated as a ratio of the efficiency corrected values of the LPA exon 4 reaction over the RNaseP reaction.

*Lp(a) quantification*

Lp(a) levels were measured using a latex-enhanced immunoturbidimetric assay (Randox Laboratories, UK) based on a highly specific polyclonal rabbit anti-apo(a) antibody on an ADVIA 1800 autoanalyser (Siemens AG, Germany) that complied with the International Federation of Clinical Chemistry recommendations[6]. The Lp(a) assay has been validated against the ELISA reference method and uses five calibrators derived from WHO reference material SRM2B. All assays were performed in EDTA plasma samples that had been stored (for up to 10 years) in liquid nitrogen or in -80C freezers prior to analysis.

*Statistical Methods*

After dropping individuals where the genotype calling algorithm failed to call genotypes of the LPA null SNP, association analysis of CAD was carried out in 4,022 cases and 4,189 controls. Additive (on the logit-scale) genetic effects were modelled by defining genotypes with dosage 0, 1 and 2 corresponding to null SNP genotypes GG, GA and AA. Country-of-origin was included as a categorical main-effect to allow for differences in allele frequencies across the populations. The familial relatedness in the PROCARDIS study was taken into account using a robust (sandwich) estimator of the variance. The analysis was performed
using STATA™ v10.1. Lp(a) kringle isoform typing using immunoblotting and qPCR and circulating Lp(a) levels were available for 1,860 samples.

Quantile regression models were used to estimate median Lp(a) levels conditional on LPA null carrier status, kringle isoforms and RQ values using the bsqreg Stata 10.1 procedure. RQ measurements were grouped as ranked measurements in twenty equal frequency bins (vigintiles) to estimate median levels for individual (categorical) RQ groups. Sources of covariation from gender, country-of-origin and CAD status were included in all three models. The null mutation was included as a covariate in the RQ model and in the specification of indicator variables for kringle alleles; RQ was included as a covariate in the kringle model. Standard errors of the regression coefficients were estimated by bootstrapping with 1,000 replicates. Median regression is a computational intensive method that fits linear models that are robust to the usual normality and homoscedasticity of residuals assumptions; they avoid any loss of information that can follow the analysis of a transformed phenotype (e.g. logarithmic-transformation of Lp(a) levels) in a conventional (least-squares) linear regression analysis.

Figure S-I: **Median regression analysis of Lp(a) levels on kringle isoform alleles.**

The regression model includes information on the null *LPA* allele but not RQ values; this reduced model contrasts with the complete model presented in Figure 4 that includes RQ values as an independent predictor of Lp(a) levels. Median Lp(a) levels are indicated by square markers, upper and lower 95% confidence limits are indicated by whiskers.
Figure S-II: **Median regression analysis of Lp(a) levels on kringle isoform alleles.** The regression model includes information on RQ values but not the null LPA allele; this reduced model contrasts with the complete model presented in Figure 4 that incorporates information on the null allele. Median Lp(a) levels are indicated by square markers, upper and lower 95% confidence limits are indicated by whiskers.