ELSEVIER

Atherosclerosis 201 (2008) 138–147

ATHEROSCLEROSIS

www.elsevier.com/locate/atherosclerosis

The apolipoprotein(a) gene: Linkage disequilibria at three loci differs in African Americans and Caucasians

Jill Rubin [a], Han Jo Kim [a], Thomas A. Pearson [c], Steve Holleran [b], Lars Berglund [a,d,e,*], Rajasekhar Ramakrishnan [b]

[a] *Departments of Medicine, Columbia University, New York, NY, United States*
[b] *Departments of Pediatrics, Columbia University, New York, NY, United States*
[c] *Department of Community and Preventive Medicine, University of Rochester, Rochester, NY, United States*
[d] *Department of Medicine, University of California, Davis, Davis, CA, United States*
[e] *VA Northern California Health Care System, United States*

Received 26 July 2007; received in revised form 18 December 2007; accepted 15 January 2008
Available online 4 March 2008

Abstract

Lipoprotein(a) (Lp(a)) is an independent, genetically regulated cardiovascular risk factor. Lp(a) plasma levels are largely determined by the apolipoprotein(a) (apo(a)) component, and differ across ethnicity. Although a number of polymorphisms in the apo(a) gene have been identified, apo(a) genetic regulation is not fully understood. To study the relation between apo(a) gene variants, we constructed haplotypes and assessed linkage equilibrium in African Americans and Caucasians for three widely studied apo(a) gene polymorphisms (apo(a) size, +93 C/T and pentanucleotide repeat region (PNR)). Apo(a) size allele frequency distributions were different across ethnicity ($p < 0.01$). For African Americans, PNR frequencies were similar across apo(a) sizes, suggesting linkage equilibrium. For Caucasians, the PNR and the PNR–C/T haplotype frequencies differed for large and small apo(a), with the T and PNR 9 alleles associated with large apo(a) size ($p < 0.0002$); also, the PNR 9 allele was more common on a T allele, while PNR 8 was more common on a C allele. On a C allele background, small PNR alleles were more common and the PNR 10 allele less common among African Americans than Caucasians ($p < 0.001$). The ethnic difference in apo(a) size distribution remained controlling for C/T and PNR alleles ($p = 0.023$). In conclusion, allele and haplotype frequencies and the nature of the linkage disequilibrium differed between African Americans and Caucasians at three apo(a) gene polymorphisms.

Keywords: Genetics; African Americans; Genotyping; Polymorphism; Apo(a); PNR; Linkage disequilibrium; Haplotypes

1. Introduction

Apolipoprotein(a), apo(a), is the defining component of lipoprotein(a), Lp(a), an independent, genetically regulated risk factor for cardiovascular disease [1–6]. The apo(a) gene, a major predictor of Lp(a) plasma levels, has a limited species distribution and has been detected only in humans, Old World primates and in the hedgehog, although in the latter, apo(a) is molecularly distinct from the first two [1,7,8]. Multiple genetic variants have been described for the apo(a) gene, among which a size variation, due to a variable number of repeats coding for a so-called kringle region [kringle 4 (K4)] to a major extent impact plasma Lp(a) levels [9–11]. Further, Lp(a) levels vary across African American–Caucasian ethnicity, with levels generally higher among the former [12–17]. Although the apo(a) size polymorphism is an important predictor of Lp(a) levels with an inverse relationship between apo(a) size and Lp(a) levels, the latter vary considerably among individuals carrying apo(a) isoforms of the same size, implicating the presence of other predictors [15,18,19]. Beyond the apo(a) size variation, two other polymorphisms, an 1 kb upstream pentanucleotide repeat region (PNR) with

* Corresponding author at: Department of Medicine, University of California, Davis, UCD Medical Center, CTSC, 2921 Stockton Boulevard, Suite 1400, Sacramento, CA 95817, United States. Tel.: +1 916 703 9120; fax: +1 916 703 9124.
E-mail address: lars.berglund@ucdmc.ucdavis.edu (L. Berglund).

0021-9150/$ – see front matter © 2008 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.atherosclerosis.2008.01.002

5 to 12 TTTTA repeats in the 5′-flanking region of the gene, and a C/T polymorphism at +93 in the promotor region have been studied widely [20–25]. We have previously reported on the PNR, C/T, and apo(a) gene size polymorphisms as predictors of allele-specific apo(a) levels and dominance pattern in Caucasians and African Americans [13,26]. However, there is a void of information on haplotypes. Although a number of apo(a) polymorphisms have been described, information based on the three studied polymorphisms would provide insight into the relationship between these common gene variants, and provide a basis to deduce haplotype distributions in the two ethnic groups.

2. Methods

2.1. Study population

Subjects were recruited from a multiethnic patient population scheduled for diagnostic coronary arteriography either at Harlem Hospital Center in New York City or at the Mary Imogene Bassett Hospital in Cooperstown, NY. Briefly, a total of 648 patients, 401 men and 247 women, ethnically self-identified as 344 Caucasians, 232 African Americans, and 72 Other, were enrolled. Data on apo(a) allele sizes were available in 430 subjects (263 Caucasians, 167 African Americans). The +93 C/T and PNR polymorphisms were determined from DNA samples in 354 subjects (215 Caucasians and 139 African Americans), with complete data available in 264 subjects (160 Caucasian and 104 African American). The design of the Harlem–Bassett study, the recruitment procedure and the clinical characteristics of the subjects have been described previously [13,15,27]. The study was approved by the Institutional Review Boards at Harlem Hospital, the Mary Imogene Bassett Hospital, Columbia University College of Physicians and Surgeons, and University of California Davis. Informed consent was obtained from all participants. Details of the determination of the apo(a) size, C/T and PNR repeat polymorphisms are given elsewhere [13,26,28].

2.2. Allele frequency distributions

The complete apo(a) size genotype distributions for Caucasians and African Americans are given in the online Appendix tables. The homozygotes as well as subjects with size difference of 1 K4 repeat are shown in bold and underlined, respectively. For the C/T and PNR polymorphisms, allele frequencies were compared between Caucasians and African Americans by χ^2 analysis. For the gene size polymorphism, the large number of alleles necessitated a modification. First, an overall comparison was done by computing the cumulative frequency distributions for the two ethnic groups using the Kolmogorov–Smirnov test [29]. Secondly, allele sizes were grouped into four ranges reflecting the largest ethnic differences in the cumulative frequency distribution and proportions in the ranges compared using χ^2 analysis.

2.3. Haplotypes

To estimate the apo(a) size distribution on T and C alleles or on PNR alleles (or PNR distribution on the T allele), it was necessary to determine haplotypes. As haplotypes are ambiguous in subjects heterozygous at two or at all three loci we estimated haplotypes for double and triple heterozygotes as described in detail in the Appendix. Briefly, the primary data are the number of subjects with each multi-locus genotype. To estimate the haplotype frequencies from the primary data, we derived formulas for the probability of observing each multi-locus genotype from haplotype frequencies. Formally, there are two cases: (1) with homozygous genotypes at all loci and (2) with one or more heterozygous loci. For subjects homozygous at either or both of two loci (or at two or three of three loci), haplotypes were determined without the need for estimation. To perform haplotype estimation for subjects with heterozygous loci, we developed software, available online at http://www.biomath.info/poolfit. For incomplete data, haplotypes over unobserved loci were added up as described in detail in the Appendix. The complexity in enumerating the possible haplotype pairs for each multi-locus genotype, the weighting and the multiple subgroups made this estimation problem intractable with available commercial software packages, prompting us to modify POOLFIT, an existing program which we had developed for fitting lipoprotein tracer data by nonlinear weighted least squares, writing special code for haplotypes, available online at http://www.biomath.info/poolfit.

As the T allele frequency was very low in African Americans, C/T haplotypes were not calculated for this group. In order to avoid very low haplotype frequencies that would be poorly estimated, grouping of PNR or apo(a) size alleles, respectively, was performed. Estimated haplotype frequencies were used to resolve ambiguities regarding double or triple heterozygotes prior to testing for linkage disequilibrium.

2.4. Linkage disequilibrium

Linkage disequlibrium between the C/T and PNR loci in Caucasians was studied by comparing the PNR allele distributions on a C vs. T allele background. For subjects homozygous at either or both loci, haplotypes were known unambiguously and the distributions could be determined directly. For example, the 22 Caucasian subjects homozygous for the C allele with a PNR 8/9 genotype contributed 22 PNR 8 and 22 PNR 9 alleles on the C allele background. For double heterozygotes, the haplotype frequencies, estimated as described above, were used: if $f_{C8}, f_{C9}, f_{T8}, f_{T9}$ are the frequencies of the four possible haplotypes for subjects double heterozygous subjects with a C/T and a PNR 8/9 genotype, the fraction of C alleles in the double heterozygotes that carry