Human Molecular Genetics, 1995, Vol. 4, No. 2 173–181

Sequence polymorphisms in the apo(a) gene associated with specific levels of Lp(a) in plasma

Vincent Mooser, Francesco P.Mancini, Sabina Bopp[1], Attila Pethö-Schramm[1], Rudy Guerra[2], Eric Boerwinkle[3], Hans-Joachim Müller[1] and Helen H.Hobbs*

Departments of Internal Medicine and Molecular Genetics, University of Texas Southwestern Medical Center, Dallas, TX 75235, USA, [1]Department of Molecular Biology, Boehringer Mannheim GmbH, Sandhoferstrasse 116, D-68305 Mannheim, Germany, [2]Department of Statistical Science, Southern Methodist University, Dallas, TX 75080 and [3]Center for Demographic and Population Genetics, University of Texas Health Science Center at Houston, Houston, TX 77225, USA

Received September 13, 1994; Revised and Accepted November 16, 1994

Most of the interindividual variations in plasma levels of lipoprotein(a) [Lp(a)] can be attributed to sequence differences linked to the apolipoprotein(a) [apo(a)] locus. Plasma levels of Lp(a) tend to be inversely related to the number of kringle 4 (K4)-encoding sequences in the apo(a) gene, but there are several exceptions to this general trend. Other aspects of the apo(a) gene, in addition to the number of K4 repeats, affect plasma levels of Lp(a). To identify sequences in the apo(a) gene that contribute to plasma Lp(a) levels, we characterized the relationship between a length polymorphism [$(TTTTA)_n$] located 1.3 kb 5′ of the first exon of the apo(a) gene, the number of K4 repeats in the gene, and the plasma levels of Lp(a). There was significant linkage disequilibrium between the number of TTTTA repeats and the number of K4 repeats. All of the apo(a) alleles with 11 TTTTA repeats contained fewer than 24 K4 repeats and were paradoxically associated with low plasma Lp(a) levels (≤3 mg/dl). To determine whether this association was due to the effect of the 11 TTTTA copies on apo(a) gene transcription, we measured the ability of fragments containing 11 or eight TTTTA repeats to promote transcription when introduced into cultured human hepatocarcinoma cells. No difference was found in the transcriptional activity of the two fragments. The TTTTA repeat constitutes the first sequence polymorphism at the apo(a) locus, other than the number of K4 repeats, which is associated with plasma concentrations of Lp(a).

INTRODUCTION

Lipoprotein(a) [Lp(a)] is a cholesterol ester-rich particle that has two components, a low density lipoprotein (LDL) particle and a large glycoprotein, apolipoprotein(a) [apo(a)]. A distinguishing feature of Lp(a) is the highly polymorphic nature of the apo(a) component. Apo(a) isoforms range in size from 250 to 900 kDa due to a length polymorphism in the apo(a) gene (1–3). The apo(a) gene contains between 12 and 51 copies of a DNA sequence encoding a tandemly repeated cysteine-rich motif called kringle 4 (K4), which is found in one copy in the closely related plasminogen gene (2,4). More than 97% of Caucasians are heterozygous for apo(a) alleles containing different numbers of K4 repeats (4). The high degree of length polymorphism resembles that of variable number of tandem repeats (VNTRs) but differs from these simple sequences in two important respects (5). First, the repeat unit in the apo(a) gene is much larger (5.5 kb) and second, the repeat unit includes coding sequences.

Plasma concentrations of Lp(a) vary over a much wider range than any other class of lipoproteins—from 0.1 mg/dl to more than 100 mg/dl—but they remain relatively constant over time within any given individual (6). The plasma level of Lp(a) is highly heritable and approximately 90% of the variability in plasma Lp(a) concentration is attributable to sequences at or closely linked to the apo(a) locus (7). This relationship is partly due to the effect of K4 repeat numbers on plasma levels of Lp(a). There is an inverse relationship between the number of K4 repeats in the apo(a) gene and plasma concentration of Lp(a) (1). However, there are obvious exceptions to this general finding. Individuals with apo(a) alleles of the same size may have very different plasma levels of Lp(a) (1,8,9). Family studies have shown that these differences are also linked to the apo(a) locus. Therefore, there must be sequence differences linked to the apo(a) locus, other than the number of K4 repeats, that contribute importantly to

*To whom correspondence should be addressed

interindividual differences in the plasma concentration of Lp(a) (7,8).

In vivo turnover studies suggest that variations in plasma Lp(a) levels associated with apo(a) alleles of the same length are due to differences in rates of apo(a) production (10). Differences in rates of production may result from alterations in apo(a) gene transcription, translation, mRNA or protein stability. Azrolan and colleagues examined the size and concentration of hepatic apo(a) mRNA in the cynmologus monkey and related these variables to the plasma concentrations of Lp(a) (11). The plasma Lp(a) levels tended to correlate directly with the abundance of hepatic apo(a) mRNA but there was no significant relationship between the size of the apo(a) mRNA transcript and the steady-state level of the mRNA. Thus, at least in the cynmologus monkeys, factors independent of the numbers of K4 repeat in the apo(a) gene contribute to interindividual differences in both apo(a) mRNA abundance and plasma Lp(a) concentration.

Previously, a sequence in the 5′ flanking region of the apo(a) gene was shown to influence expression of chimeric apo(a) promoter-luciferase reporter gene constructs in transfected HepG2 cells (12). A cloned 1.5 kb fragment from the 5′ flanking region was isolated from genomic DNA of one individual with a high and another with a low plasma level of Lp(a). When the DNA sequences of the two cloned fragments were compared, two differences were found. There was a single base pair substitution (G to A) 914 nucleotides upstream of the ATG, and a difference in the number of pentanucleotide repeats (TTTTA) located 1.4 kb 5′ of the ATG codon. The apo(a) allele from the individual with high plasma Lp(a) level had eight copies of the TTTTA tandem repeat, whereas the allele associated with low plasma Lp(a) levels had nine tandem repeats. When these two 1.5 kb fragments were placed upstream of a luciferase reporter gene and expressed in cultured HepG2 hepatocarcinoma cells, the construct containing eight copies of the TTTTA repeat had a 5-fold higher transcriptional activity than that with nine copies of the repeat.

To determine whether the size variation in the TTTTA repeat contribute to the interindividual variations in plasma Lp(a) levels, we analyzed this repeat in 67 Caucasian families using a PCR-based method and determined the relationship between the number of TTTTA and K4 repeats and the plasma concentration of Lp(a).

RESULTS

The length polymorphism identified by Wade *et al.* is located 1.3 kb 5′ of exon 1 of the apo(a) gene at the 3′ end of an *Alu* sequence (Fig. 1A; 12). To analyze the length of this polymorphism, oppositely oriented oligonucleotides complementary to sequences flanking the repeat were used together with PCR to amplify the repeat from total genomic DNA of 214 Caucasians, 44 African-Americans and 31 Chinese individuals, all unrelated. A total of eight fragments of different sizes were identified in the sample. The size of the fragments differed by multiples of five nucleotides and ranged from 93 to 128 bp (Fig. 1B). Six fragments generated from three heterozygous individuals were sequenced in their entirety and the only difference was the number of TTTTA repeats, which ranged from five to 11 (data not shown).

An identical analysis was performed using genomic DNA from a rabbit, baboon and gorilla (Fig. 1C). The gorilla and

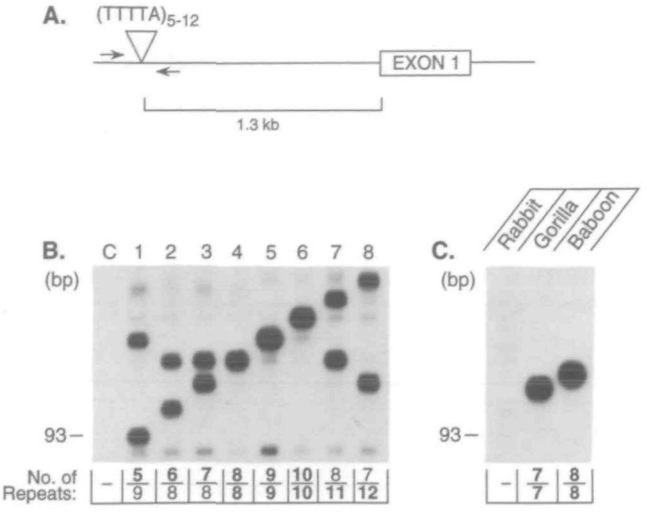

Figure 1. (**A**) Schematic map of the 5′ region of the apo(a) gene and location of the TTTTA length polymorphism. (**B**) Two oppositely oriented 25 base primers flanking the length polymorphism site were used to amplify the intervening sequences from genomic DNA in the presence of [α-^{32}P]dCTP. The amplified radiolabeled fragments were size-fractionated on a denaturing 6% polyacrylamide gel and eight fragments that differed in length by five nucleotides were identified. There were also a number of light bands visible in all the lanes. The source of their amplification products was not determined. (**C**) Genomic DNA from rabbit, gorilla and baboon were similarly analyzed.

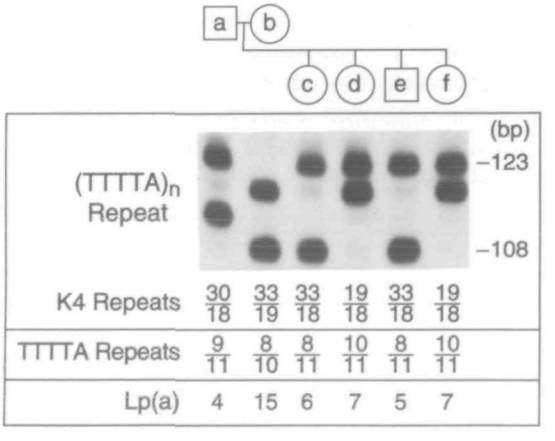

Figure 2. Segregation of the TTTTA polymorphism in a Caucasian nuclear family. Radiolabeled fragments containing the TTTTA length polymorphic sequence were amplified using PCR from genomic DNA, and the amplification products were analyzed by gel electrophoresis exactly as described in Figure 1.

baboon were homozygous for apo(a) alleles containing seven and eight TTTTA repeats, respectively. No amplification product was identified using rabbit genomic DNA. The amplified fragment from the gorilla was sequenced and found to be identical to the human sequence, except for the number of TTTTA repeats (data not shown).

The TTTTA polymorphism segregated in a stable codominant Mendelian fashion in 67 unrelated nuclear Caucasian families. The result of the analysis of a family in which all four parental alleles generated fragments of different length is shown in Figure 2.

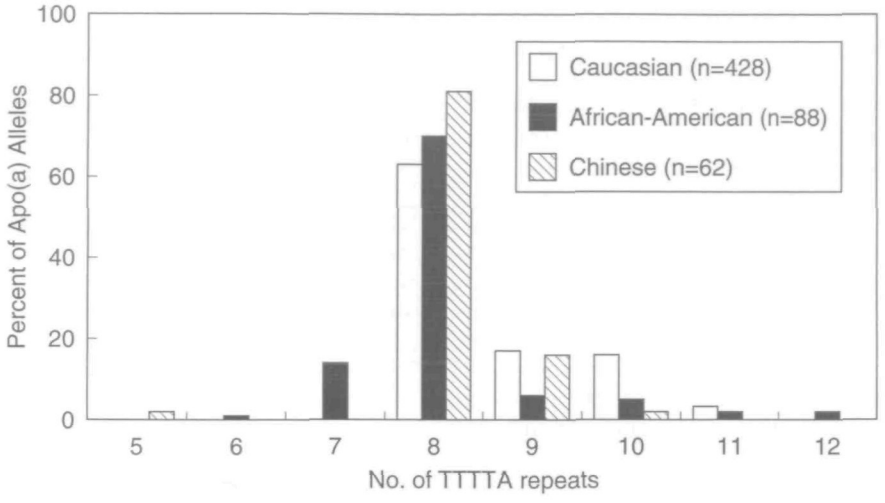

Figure 3. Distribution of the TTTTA alleles in 214 unrelated Caucasian, 44 African-American and 31 Chinese individuals. The TTTTA length polymorphism was analyzed exactly as described in Figure 1. On the *x* axis, the number of tandem copies of the TTTTA repeat is given.

The size distribution of apo(a) alleles in three ethnic groups is shown in Figure 3. In a Caucasian sample of 214 unrelated individuals, the apo(a) allele with $(TTTTA)_8$ was most common with a frequency of 64%. Alleles with 9, 10 and 11 TTTTA repeats accounted for 17, 16 and 3% of the Caucasian sample. The TTTTA genotypic frequencies in the Caucasian population did not differ significantly from those expected in a Hardy–Weinberg equilibrium (chi square = 2.50, d.f. = 6, *P*-value = 0.87) and the heterozygosity index was 51%. In contrast, 97% of the 31 unrelated Chinese analyzed had eight or nine copies of the TTTTA repeat and the heterozygosity index in this group was 35%. Among African-Americans, the most frequent allele was $(TTTTA)_8$. Three alleles—$(TTTTA)_6$, $(TTTTA)_7$ and $(TTTTA)_{12}$—were only observed in the sample of African-Americans. The heterozygosity index in the African-Americans was 48%.

To investigate for linkage disequilibrium between the TTTTA length polymorphism and the number of K4 repeats, we analyzed the distribution of apo(a) haplotypes with respect to their number of TTTTA and K4 repeats (Fig. 4). The apo(a) haplotypes were constructed by analyzing the cosegregation of the two length polymorphisms in 67 unrelated Caucasian families. The value of the standardized measure of linkage disequilibrium between the TTTTA repeat locus and the K4 repeat locus was 0.475 in this sample (13). However, the degree of association did not appear homogeneous among alleles. Apo(a) alleles with eight TTTTA repeats were associated with alleles containing a broad range of K4 repeats (15–40). Most of the apo(a) alleles with nine TTTTA repeats were associated with a moderate number (25–34) of K4 repeats, whereas alleles with 10 TTTTA repeats (n = 38) segregated predominately with few (<24) K4 repeats. The most striking observation, however, was that all of the apo(a) alleles with 11 TTTTA repeats (n = 6) were associated with a low number (<24) of K4 repeats.

During the course of these studies, a biallelic single-strand conformation polymorphism (SSCP-5) was identified in intron 1 of the apo(a) gene, as shown in Figure 5. This polymorphism was examined in the 67 unrelated Caucasian nuclear families

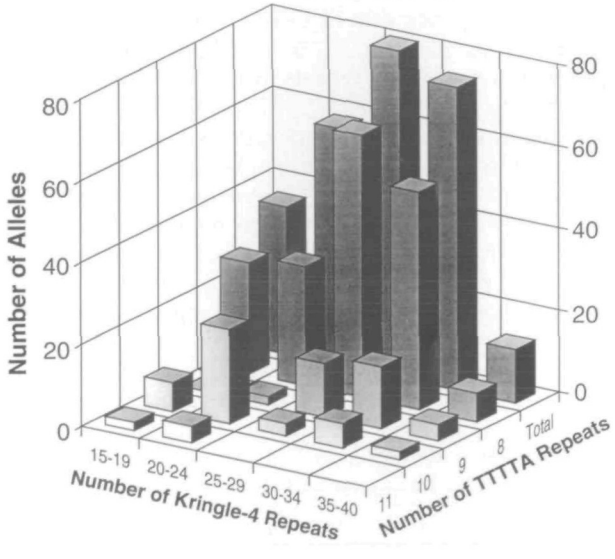

Figure 4. Distribution of the apo(a) alleles (n = 244) according to the number of K4 repeats and the number of TTTTA repeats. The number of TTTTA repeats was determined as described in Figure 1. The number of K4 repeats was assessed in 43 families by pulsed-field gel electrophoresis of *Kpn*I or *Hpa*I digested genomic DNA and genomic blotting using an apo(a) specific probe, and by immunoblotting of plasma Lp(a) as previously described (3). The relationship between the number of TTTTA and K4 repeats was deduced from the cosegregation of the two length polymorphisms in 67 unrelated Caucasian nuclear families, as described in Materials and Methods.

by subjecting a 275 bp PCR-fragment to denaturation and then size-fractionation on a nondenaturing gel. Haplotypes were constructed by analyzing the cosegregation of SSCP-5 and the TTTTA polymorphism in families. All of the apo(a) alleles with eight (n = 168) or nine (n = 44) repeats were associated with the most common SSCP-5 allele (A) and all of the alleles containing 11 (n = 6) repeats were associated with the less common SSCP-5 allele (B). Both alleles of SSCP-5 were found in association with $(TTTTA)_{10}$. Approximately one-

quarter of the apo(a) alleles with 10 TTTTA repeats cosegregated with the A pattern (n = 12) and the remaining three-quarters with the B pattern (n = 37).

To determine whether apo(a) alleles with specific numbers of TTTTA repeats were associated with different plasma Lp(a) levels, the plasma Lp(a) level associated with each apo(a) allele was estimated as described in Materials and Methods. The relationship between the number of K4 repeats in the apo(a) gene and the plasma Lp(a) level is depicted in Figure 6. As expected from previous studies, an inverse relationship between apo(a) size and Lp(a) levels was apparent (1,3). However, several exceptions to the general pattern were observed. In particular, 32 of the 68 alleles (47%) with <24 K4 repeats were associated with a plasma level of Lp(a) less than 10 mg/dl. In contrast, only two of 134 alleles with ≥24 K4 repeats were associated with a plasma level of Lp(a) greater than 20 mg/dl and these two alleles had 25 and 26 K4 repeats, respectively.

Next, the relationship between the size of the apo(a) alleles, the SSCP pattern at SSCP-5, and the plasma level of Lp(a) were analyzed separately for each TTTTA allele (Fig. 7). The apo(a) alleles associated with patterns A and B at SSCP-5 are denoted by closed and opened circles, respectively. The size distribution of alleles with $(TTTTA)_8$ was very similar to that of the entire sample (Fig. 6). As noted previously, the apo(a) alleles with nine TTTTA repeats tended to be associated with a moderate to high number (25–37) of K4 repeats. All of the apo(a) alleles with eight and nine repeats were associated with pattern A at SSCP-5. The distribution of plasma Lp(a) levels did not differ between the $(TTTTA)_8$ and $(TTTTA)_9$ groups, if the alleles were matched for the number of K4 repeats. SSCP-5 separated the apo(a) alleles with 10 repeats (n = 38) into two groups; those with pattern A (n = 9) were associated with a moderate number (26–35) of K4 repeats and those with pattern B (n = 29) all had a smaller number (19–23) of K4 repeats. Of the 29 apo(a) alleles with 10 TTTTA repeats and fewer than 24 K4 repeats, 24 were associated with plasma levels of Lp(a) less than or equal to 11 mg/dl, and five with an Lp(a) level >40 mg/dl. All of the apo(a) alleles with 11 copies of the TTTTA repeat had SSCP-5 pattern B, contained fewer than 24 K4 repeats and were associated with low plasma levels of Lp(a) (≤3 mg/dl).

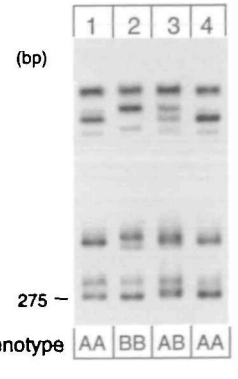

Figure 5. SSCP analysis of a 275 bp PCR-amplified fragment from intron 1 of the apo(a) gene in four unrelated individuals (SSCP-5). In this biallelic polymorphism the two alleles are designated A and B.

The apo(a) alleles containing 11 TTTTA repeats all cosegregated in families with low plasma levels of Lp(a). Immunoblot analysis of an equal amount of plasma from each individual of a nuclear family in which one parent has the $(TTTTA)_{11}$ allele is shown in Figure 8. Segregation analysis of the TTTTA and K4 length polymorphism revealed that the father has one allele with eight TTTTA and 19 K4 repeats (lane 1). This allele is associated with a large amount of apo(a) in the plasma as shown by the prominent band on the immunoblot (lanes 1,4,5). His other apo(a) allele contained 32 K4 repeats, as determined by pulsed-field gel analysis of the apo(a) gene (data not shown), but was not associated with a detectable protein product (lanes 1 and 3). The mother has an apo(a) allele with 11 TTTTA repeats and 18 K4 repeats which is associated with small amounts of apo(a) (lane 2). The mother's other apo(a) allele contains 24 K4 repeats and this allele generates most of the apo(a) circulating in her plasma. Thus, though both the maternal and paternal apo(a) alleles have a similar number of K4 repeats (i.e. 18 and 19), the two alleles are associated with very different plasma levels of Lp(a), as evidenced by the differences in the intensity of the apo(a) signals by immunoblotting.

A luciferase reporter gene assay was employed to compare the relative transcriptional activity of a 1.5 kb fragment from the 5′ flanking region of the apo(a) gene that included either eight or 11 TTTTA repeats. This region of the apo(a) gene was amplified from genomic DNA from the second child in the family depicted in Figure 8 (lane 4). This individual has one allele that is associated with a high plasma level of Lp(a) and that contains eight TTTTA and 19 K4 repeats that he inherited from his father; his other allele, which he inherited from his mother, is of similar length (i.e. has 18 K4 repeats), has 11 TTTTA repeats, and is associated with a low plasma concentration of Lp(a). These fragments were subcloned into a plasmid 5′ of a luciferase gene. A recombinant pGL3E plasmid that contained an apo(a) 5′ region fragment from an unrelated individual (with eight TTTTA repeats), and a promoterless plasmid served as positive and negative controls, respectively. Each plasmid was transfected on five separate occasions into cultured HepG2 cells and the expressed luciferase enzyme activity was quantitated. Figure 9 summarizes the relative transcriptional activity of the fragment, after normalization for transfection efficiency using β-galactosidase as a control. There were no significant differences in the luciferase activities produced by the fragments containing eight or 11 TTTTA repeats (Fig. 9). Similar experiments were performed with identical luciferase constructs that did not contain the SV40 enhancer element and there were lower levels of activity but no significant differences between the two constructs containing eight and 11 repeats.

DISCUSSION

In this paper we analyzed the relationship between the plasma concentration of Lp(a) and two length polymorphisms in the apo(a) gene—a pentanucleotide repeat $[(TTTTA)_n]$ in the 5′ flanking region and a 5.5 kb K4-encoding tandem repeat in the mid-region of the gene. The pentanucleotide repeat is located at the 3′ end of an *Alu* sequence, which is a frequent site of length polymorphisms (14,15). There was linkage disequilibrium between this polymorphism and the number of

K4 repeats in the apo(a) gene. Specific apo(a) allele haplotypes were associated with only a limited range of plasma Lp(a) concentrations. The most dramatic example was the apo(a) alleles with 11 copies of the TTTTA repeat; all of these alleles were associated with apo(a) alleles containing relatively few K4 repeats (18–23) and producing low plasma levels of Lp(a) (≤3 mg/dl). This was an unexpected observation since small apo(a) alleles tend to be associated with high plasma levels of Lp(a). Only 28% of the apo(a) alleles with less than 24 K4 repeats were associated with plasma Lp(a) levels ≤3 mg/dl, yet all of the apo(a) alleles with 11 repeats had such low levels. This is the first sequence variation at the apo(a) locus, other than the number of K4 repeats, that is associated with specific plasma levels of Lp(a).

The identification of the sequences in the apo(a) gene that are responsible for the 1000-fold interindividual variation in plasma levels of Lp(a) has been complicated by the fact that the apo(a) gene is highly polymorphic both in length and in sequence. Since the number of K4 repeats in the apo(a) gene strongly influences plasma Lp(a) levels, comparisons must be performed between apo(a) alleles with a similar number of repeats. The large size and the repetitive nature of the apo(a)

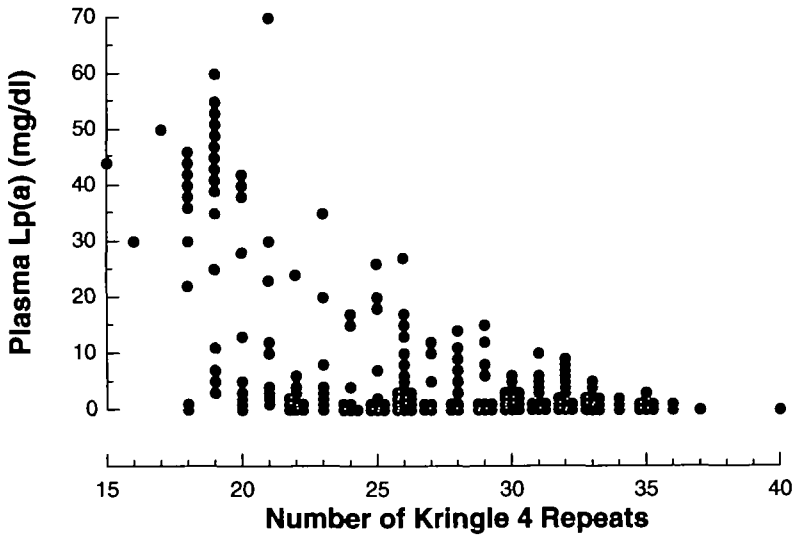

Figure 6. Relationship between number of K4 repeats in the apo(a) gene and plasma levels of Lp(a) in 202 alleles from unrelated Caucasians. The plasma level of Lp(a) associated with each allele was estimated as described in Materials and Methods.

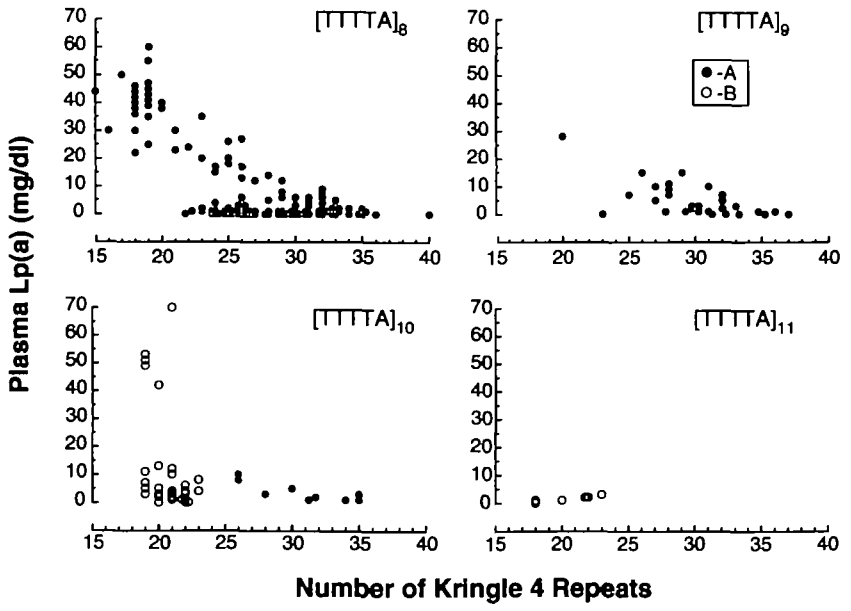

Figure 7. Relationship between number of K4 repeats in the apo(a) gene, SSCP-5, and the plasma levels of Lp(a) for each of the (TTTTA) alleles in a sample of 202 apo(a) alleles from unrelated Caucasians. ●: SSCP-5, pattern A. ○: SSCP-5, pattern B.

gene confounds the molecular characterization of individual apo(a) alleles (16). The apo(a) gene is highly polymorphic in sequence, thus the significance of specific sequence differences identified between apo(a) alleles is difficult to assess (8). Finally, the apo(a) gene is expressed almost exclusively in the liver, which renders it experimentally difficult to study the effects of specific sequence variations on apo(a) biosynthesis and metabolism (17). Our finding of linkage between apo(a) allele haplotypes and plasma levels of Lp(a) provides a strategy by which apo(a) alleles can be classified based on their DNA haplotypes and plasma Lp(a) levels. Those alleles with similar (but nonidentical) haplotypes that are associated with different plasma levels of Lp(a) can be selected for more detailed molecular characterization.

The mechanism responsible for the low plasma levels of Lp(a) associated with 11 TTTTA repeats is not known. Tissue culture studies suggested that the low levels of plasma Lp(a) are not attributable to a direct effect of the repeat on apo(a) gene transcription. However, we cannot definitively exclude the possibility that 11 copies of the repeat affects the transcription rate of the apo(a) gene *in vivo* since all of the sequences required for high level of apo(a) gene transcription have not yet been identified. Recently, exon 1, which encodes the 5′ untranslated region of the apo(a) gene, was shown to play an essential role in apo(a) gene transcription (18). We screened this sequence for variations using the SSCP technique as described (8). Two previously reported SSCP polymorphisms were found within this region, namely a G to A polymorphism at position −22 and a C to T polymorphism at position −49. There was no association between either of these sequence variations and the plasma level of Lp(a) (data not shown). Consequently, it is unlikely that the sequence responsible for the low levels of Lp(a) associated with alleles with $(TTTTA)_{11}$ alleles resides in exon 1. Therefore, it is likely that $(TTTTA)_{11}$ is in linkage disequilibrium with the actual sequence responsible for the low plasma level of Lp(a) associated with this allele. These sequences may effect the stability or transport of either the mRNA transcript or the apo(a) glycoprotein. Given the cysteine-rich nature of the K4 repeats, missense mutations within this region might be expected to interfere with the proper folding of the apo(a) protein during its biosynthesis and result in less efficient export of apo(a) out of the cell. Since apo(a) mRNA and glycoprotein are only expressed in the liver, it is not possible to examine the apo(a) mRNA transcript or the newly synthesized apo(a) glycoprotein generated from the apo(a) alleles with $(TTTTA)_{11}$.

To demonstrate that a sequence variation at a particular site in the apo(a) gene may be associated with different plasma Lp(a) levels, sequence comparisons must be made between apo(a) alleles with similar numbers of K4 repeats. As an example, this same TTTTA length polymorphism was previously implicated in contributing to variation in plasma Lp(a) levels by affecting apo(a) gene transcription (12). A fragment containing eight TTTTA repeats was transcriptionally more active in tissue culture when compared with the same fragment with nine repeats. In our sample, apo(a) alleles with eight TTTTA repeats tended to be associated with higher plasma levels of Lp(a) than alleles with nine TTTTA repeats (10.7 ± 15.9 mg/dl; n = 127 vs 4.5 ± 6.3 mg/dl; n = 31, mean ± standard deviation), but this difference could be attributed entirely to the differences in the distribution of the numbers of K4 repeats between the two groups. There was no significant difference between the two groups when matched for similar

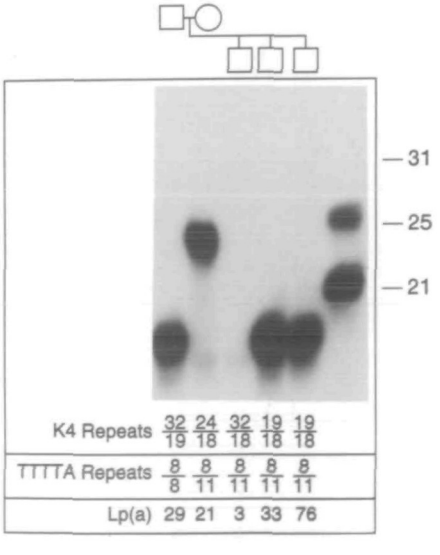

Figure 8. Immunoblot analysis of the apo(a) isoforms in a Caucasian family. A 3 µl aliquot of plasma from each family member (lanes 1–5) and a control individual subject (lane 6) was subjected to reduction and size-fractionated on a 6% agarose gel. The proteins were transferred to a nitrocellulose filter and blotted with IgG-1A[2], a mouse monoclonal antibody that had been conjugated with horseradish peroxidase. The number of TTTTA and K4 repeats were determined as described in Figure 4.

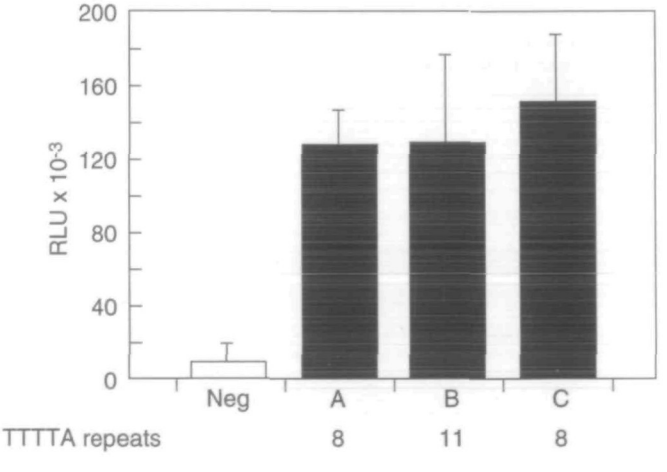

Figure 9. Comparison of the transcriptional activity of a 1.5 kb fragment from the 5′ flanking region of the apo(a) gene containing $(TTTTA)_8$ or $(TTTTA)_{11}$. Allelic 1.5 kb fragments (−1499 to 12) from the 5′ flanking region of the apo(a) gene were amplified from genomic DNA and subcloned upstream of the luciferase gene in pGL3E as described in Materials and Methods. Plasmids were selected which contained either eight (A) or 11 (B) TTTTA repeats and each was transfected five times into HepG2 cells. An apo(a) promoter fragment from an unrelated individual (C) and a promoterless luciferase plasmid (Neg) were employed as positive and negative controls. A β-galactosidase plasmid was co-transfected to normalize for variable transfection efficiencies. Sixty-four hours after transfection, cytoplasmic extracts were prepared as described in Materials and Methods, and the luciferase and β-galactosidase activities were measured. The bars represent the mean values (plus standard deviation) of five independent transfections.

numbers of K4 repeats (3.2 ± 5.6 mg/dl; n = 88 vs 3.8 ± 4.6 mg/dl; n = 29 for apo(a) alleles with 25–37 K4 repeats). Therefore, it can be concluded that the experiments in tissue culture do not explain the interindividual variations in plasma Lp(a) levels in the general Caucasian population.

In Caucasians, the observed relationships between the numbers of TTTTA and K4 repeats, and plasma levels of Lp(a) suggest a possible intraspecific phylogeny for the apo(a) gene. The apo(a) alleles containing eight TTTTA repeats are associated with the entire range of K4 repeat numbers (15–40), and plasma Lp(a) levels (<0.1–67 mg/dl), a result consistent with this allele [$(TTTTA)_8$] being the ancestral apo(a) allele. Mutations that established the other apo(a) alleles [$(TTTTA)_{9-11}$] are likely to have occurred relatively recently since these alleles are associated with specific subsets of K4 repeats. For example, the apo(a) alleles with nine TTTTA repeats almost always had a moderate number of K4 repeats (94% have 25–37 K4 repeats). Thus, it is likely that an apo(a) allele with eight TTTTA repeats and a moderate number of K4 repeats acquired an additional TTTTA repeat to generate the $(TTTTA)_9$ allele. Subsequent recombinational events involving the K4-encoding sequences probably account for the range in the number of K4 repeats associated with $(TTTTA)_9$. Based on the analysis of another polymorphism, SSCP-5, the apo(a) alleles with 10 TTTTA repeats consist of two populations. Approximately 25% (nine of 38) of the alleles with 10 repeats are associated with a moderate number of K4 repeats. These alleles are likely to have been derived from an allele with nine repeats. In support of this scenario, all of the apo(a) alleles with 10 TTTTA and more than 25 K4 repeats share the same allele at SSCP-5 as do the alleles with $(TTTTA)_8$ and $(TTTTA)_9$. All of the $(TTTTA)_{10}$ alleles with fewer than 25 K4 repeats have the B allele at SSCP-5, as do all the $(TTTTA)_{11}$ alleles. The $(TTTTA)_{10}$ and $(TTTTA)_{11}$ alleles comprise 44% of all the apo(a) alleles with fewer than 24 repeats. From this analysis, it would be predicted that the apo(a) alleles with 10 or 11 TTTTA repeats with <24 K4 repeats and low plasma Lp(a) levels have a similar sequence variation.

The relationship between the number of TTTTA repeats and the plasma levels of Lp(a) described in the present study was based on the analysis of a sample of 67 Caucasian families. These results cannot be extrapolated to other ethnic groups. Indeed, both the distribution of Lp(a) levels and the number of K4 repeats have been shown to differ between ethnic groups (1,19–21). Our data show that this difference in ethnic distribution is also the case for the TTTTA polymorphism. For example, apo(a) alleles with seven TTTTA repeats comprised 14% of the alleles in the African-American sample, but were not detected in either the Chinese or the Caucasian sample.

The most important finding in this paper is the fact that there is an association between sequence polymorphism in the apo(a) gene and plasma levels of Lp(a). This finding is of practical interest since it provides a strategy by which alleles at this highly polymorphic locus can be classified and selected for detailed molecular analysis to determine sequence differences that contribute importantly to plasma levels of Lp(a).

MATERIAL AND METHODS

Venous blood was collected from 67 unrelated Caucasian nuclear families, 80 unrelated Caucasian individuals, 44 unrelated African-American, and 31 unrelated Orientals born in mainland China, who all live within 50 miles of Dallas, Texas. Plasma was isolated, aliquoted and maintained at −70°C. The plasma Lp(a) levels were measured either at GeneScreen as previously described (3), or by using a sandwich ELISA assay developed by Drs Santica Marcovina and John Albers (University of Washington, Seattle, WA) which uses a mouse anti-apo(a) monoclonal antibody (Ab-40) as a capture antibody and a rabbit polyclonal antibody against human apoB-100 as a detecting antibody. In 43 of the 67 families, the number of K4 repeats in the apo(a) gene was determined using pulsed-field gel electrophoresis and genomic blotting as previously described (2,3). In all individuals from the 67 families, the size of the apo(a) isoforms were determined by immunoblotting using a mouse monoclonal antibody, 1gG-1A[2], which recognizes the common K4 repeat of apo(a) (kindly provided by Dr Utermann, Innsbruck, Austria), exactly as previously described (3).

Detection of the TTTTA polymorphism

Genomic DNA was extracted from leukocytes isolated from whole blood and used as a template for amplification employing PCR and two oppositely oriented 25-base primers (FM-1: 5′-ATTTGCGGAAAGATTGATACTATGC-3′ and FM-2: 5′-CCAAAATCA CGTCAGTGCACTTCAA-3′) which flank the length polymorphism (22). The primers were synthesized using a 394 DNA synthesizer (Applied Biosystems, Inc., Foster City, CA, USA). Amplification was performed in a total volume of 10 µl containing 100 ng of genomic DNA, 15 pmol of each primer, 0.5 nmol of dNTP, 1.7 pmol [α-^{32}P]dCTP (3000 Ci/mmol), 1.25 U of *Taq* polymerase (Pharmacia, Piscataway, NJ, USA) in the buffer provided by the manufacturer. Genomic DNA was denatured for 5 min at 95°C and then submitted to 30 cycles of amplification with denaturation for 1 min at 96°C, annealing for 1 min at 55°C, and extension for 2 min at 70°C. Two µl of the amplification products were diluted in 30 µl of formamide dye [98% (v/v) formamide, 0.01% (wt/vol) xylene cyanol, 0.01% (wt/vol) bromophenol blue, 10 mM Na_2 EDTA, pH 8.0] and denatured by boiling for 5 min. A 3 µl aliquot was loaded on to a 6% denaturing polyacrylamide gel, and electrophoresed for 90 min at 55 W. Gels were dried and exposed to Reflection NEF-495 films (Dupont) at room temperature for 14 h.

Sequencing of PCR-amplified fragments

Genomic DNA from three individuals heterozygous for the TTTTA polymorphism, as well as from a gorilla, were amplified by PCR using oligos FM-1 and FM-2 in a volume of 50 µl containing 250 ng of genomic DNA, 75 pmol of each primer, 2.5 nmol of dNTP and 2.5 U *Taq* polymerase in the buffer described above. The amplified fragments were size-fractionated on a 6% nondenaturing polyacrylamide gel. The individual bands of different size were cut of the gels, the DNA purified and the fragments sequenced using a ^{32}P-end-labeled oligonucleotide (FM-1) exactly as described previously (23).

Estimation of plasma levels of Lp(a) associated with each apo(a) allele

The plasma concentration of Lp(a) associated with each independent apo(a) allele in 43 unrelated Caucasian families was estimated by comparing the results of the immunoblot analysis of the apo(a) isoforms to the pulsed-field gel (PFGE) analysis. Our immunoblotting technique detects all apo(a) isoforms associated with plasma Lp(a) concentrations of at least 0.1 mg/dl (19). In individuals heterozygous for a 'null' allele (i.e. an allele which can be demonstrated by pulsed-field gel electrophoresis, but undetectable by immunoblotting), all of the circulating plasma Lp(a) was attributed to the other allele.

The contribution of each of the four parental alleles to the plasma Lp(a) levels could be unambiguously estimated in 34 of the 43 families. In those families in which several individuals were heterozygous for a null allele and shared the same expressing allele, the Lp(a) level attributable to the expressing apo(a) alleles was estimated by taking the mean of the plasma Lp(a) levels of these individuals. (The r value between plasma Lp(a) levels in these individuals was 0.88.) In 14 families, only three of the four parental alleles could be included. One such family is shown in Figure 8. In this family, the father, who is heterozygous for apo(a) alleles containing 19 and 32 K4 repeats, had only the protein product of *apo(a)K4*$_{19}$ visible by immunoblot analysis of his plasma; consequently, an Lp(a) level of 0 mg/dl was attributed to *apo(a)K4*$_{32}$ and a plasma level of 29 to *apo(a)K4*$_{19}$. *Apo(a)K4*$_{32}$ was donated to the first child of this family; thus, all of the plasma Lp(a) in this child (3 mg/dl) was attributed to the maternal allele which contained 18 K4 repeats. Finally, the second maternal allele, which contained 24 K4 repeats, was not included in the analysis since this allele was only present in the mother and she had another expressing allele. In five families, all four parental alleles had detectable plasma Lp(a) by immunoblotting, so the amount of Lp(a) associated with each allele could not be determined; these families were excluded from the analysis. In 24 families, PFGE analysis was not performed

and the number of K4 repeats was deduced from the size of the apo(a) isoforms on immunoblotting. Therefore, no null alleles were included from this second set of families since the number of K4 repeats associated with the null allele was not ascertained. An unambiguous plasma Lp(a) level could be attributed to 54 of the possible 96 apo(a) alleles in this group. Taken together, a total of 202 apo(a) alleles were included in the analysis relating apo(a) genotypes to plasma levels of Lp(a). The percentage of null alleles in the entire sample (57 of 202, 28%) was similar to the 29% observed previously in an unrelated Caucasian sample (21).

Analysis of SSCP polymorphism in intron 1 of the apo(a) gene

A 275 bp fragment initiating 148 bp downstream the ATG site in of the apo(a) gene was PCR amplified using the oppositely oriented primers VM8 (5'-TGATTCATGAAATTCCCAGTTC-3') and MV4 (5'-AGCATGGCACCTG-AACAGA GATGAT-3'). Amplification was performed in a total volume of 10 µl containing 100 ng of genomic DNA, 15 pmol of each primer, 0.5 nmol of dNTP, 1.7 pmol of [α-^{32}P]dCTP (3000 Ci/mmol) and 1.25 U of *Taq* polymerase. The sample was subjected to denaturation for 3 min at 94°C, and then 35 cycles of denaturation for 40 s at 94°C, annealing for 30 s at 59°C, and extension for 2 min at 72°C. One µl of amplification product was diluted in 30 µl of dye and boiled for 5 min. A sample of 2.5 µl of the mix was loaded on a 6% nondenaturing polyacrylamide gel containing 10% (v/v) glycerol and subjected to electrophoresis for 14 h at 300 V at room temperature as previously described (8). Gels were dried and exposed to Reflection NEF-495 film for 4–14 h.

Statistical methods

Standard chi-square statistics comparing observed frequencies with expected frequencies under a null hypothesis of random association were calculated to test for Hardy–Weinberg equilibrium at the TTTTA locus, and linkage disequilibrium between the TTTTA polymorphism and the K4 polymorphism. The test statistic for Hardy–Weinberg equilibrium was calculated for the Caucasian sample of 134 unrelated individuals from 67 unrelated families. To summarize the degree of association between the TTTTA repeat locus and the K4 repeat locus, Hedrick's standardized D [Equation 14 in Hedrick (13)] was calculated. The statistical significance of D was assessed by the permutation method of Chakraborty *et al.* (24). The linkage disequilibrium test statistic was calculated for the 202 haplotypes used to estimate the plasma levels of Lp(a) associated with each of the apo(a) alleles in the sample. Haplotypes involving 15, 16, or 17 K4 repeats were pooled within each TTTTA length class, and similarly, haplotypes involving 36–40 K4 repeats were pooled prior to computing the linkage disequilibrium statistics (25).

Transfection and luciferase reporter gene assays

A 1.5 kb fragment from the 5' region of the apo(a) gene was amplified by PCR using the following primers: #71, 5'-GTCAAGATCTACCACTCTTGC-TTTACTTCATG-3' and MV1, 5'-GGACTGGCCAGCAGTGCCCAGAAA-GTGT-3'. The 5' primer contains 10 extra nucleotides to introduce a *Bgl*II restriction site at the 5' end of the PCR product. Amplification was carried out in a total volume of 50 µl containing 250 ng of genomic DNA, 75 pmol of each primer, 2.5 nmol of dNTP, 2.5 U of *Taq* polymerase in the buffer provided by the manufacturer. After denaturation of the samples for 2 min at 88°C and 3 min at 94°C, the samples were submitted to 35 cycles of amplification with denaturation for 40 s at 94°C, annealing for 30 s at 59°C, and extension for 2 min at 72°C. These cycles were followed by a final extension step at 72°C for 10 min. After phenol–chloroform extraction and ethanol precipitation, the insert was digested with *Bgl*II and *Msc*I. The fragments containing nucleotides −1499 to −12 of the apo(a) gene were cloned between the *Bgl*II and *Sma*I restriction sites of pGL3E, a modified pGL2 Enhancer plasmid (Promega, Madison, WI) which had the multiple cloning site at the 5' end of the luciferase gene substituted by a synthetic polylinker comprising *Bgl*II and *Sma*I cloning sites. Recombinant plasmids containing either $(TTTTA)_8$ or $(TTTTA)_{11}$ were identified by PCR using FM-1 and FM-2 and purified by Qiagen columns (Studio City, CA).

HepG2 cells were obtained from American Type Culture Collection and grown up in Dulbecco's minimal essential medium with Earle's salts, supplemented with 10% fetal calf serum, 2 mM L-glutamine, 1% (w/v) nonessential amino acids and 1% (w/v) sodium pyruvate. Medium and supplements were from Boehringer Mannheim. Lipofection of HepG2 cells was performed as described (26) employing a mixture of 5 µg pGL3E (the apo(a)-promoter/luciferase plasmid) and 1 µg pCMV-β (Clontech, Palo Alto, CA), a plasmid which contains the β-galactosidase gene of *Escherichia coli* in an eukaryotic expression vector. Sixty-four hours after transfection, cytoplasmic extracts were prepared as described (26). Ten µl of these extracts were employed to determine luciferase activities (27). A 10 s integral of the relative light units (RLU) was read in a LB950-Luminometer (Berthold, Bud, Wildbad, Germany). Twenty µl of cytoplasmic extracts were diluted with 30 µl water to measure β-galactosidase activities. Following a 5 min preincubation at 37°C, the reaction was initiated by adding 50 µl prewarmed substrate solution (5 mM *o*-nitrophenol-β-D-galactoside in 100 mM sodium phosphate buffer, 0.1 mM $MgSO_4$, 40 mM β-mercaptoethanol, pH 7.3). After a 10–20 min incubation at 37°C, the yellow reaction products were quantitated by absorbance reading at 405 nm. Luciferase activities were normalized for β-galactosidase activities to correct for variation in the transfection efficiency.

ACKNOWLEDGMENTS

We wish to thank Kathy Schueler and Tommy Hyatt for their excellent technical assistance and Jonathan Cohen, Michael Brown and David Russell for helpful discussions and review of the manuscript. We also want to thank Drs Santica Marcovinia and John Albers for measuring the plasma Lp(a) levels. This work was supported by NIH HL-47619, HL-20948, the Perot Family Fund, the Swiss Academy for Medical Sciences, the Ciba-Geigy Foundation and the Holderbank Foundation. Helen H.Hobbs is an Established Investigator of the American Heart Association.

ABBREVIATIONS

Apolipoprotein(a), apo(a); kringle, K; lipoprotein(a), Lp(a); low density lipoprotein, LDL; polymerase chain reaction, PCR; pulsed-field gel electrophoresis, PFGE; single-strand conformation polymorphism, SSCP

REFERENCES

1. Utermann, G. (1989) The mysteries of lipoprotein(a). *Science*, **246**, 904–910.
2. McLean, J. W., Tomlinson, J. E., Kuang, W. J., Eaton, D. L., Chen, E. Y., Fless, G. M., Scanu, A. M. and Lawn, R. M. (1987) cDNA sequence of human apolipoprotein(a) is homologous to plasminogen. *Nature*, **330**, 132–137.
3. Lackner, C., Boerwinkle, E., Leffert, C. C., Rahmig, T. and Hobbs, H. H. (1991) Molecular basis of apolipoprotein(a) isoform size heterogeneity as revealed by pulsed-field gel electrophoresis. *J. Clin. Invest.*, **87**, 2077–2086.
4. Lackner, C., Cohen, J. C. and Hobbs, H. H. (1993) Molecular definition of the extreme size polymorphism in apolipoprotein(a). *Hum. Mol. Genet.*, **2**, 933–940.
5. Jeffreys, A. J., Wilson, V. and Thein, S. L. (1985) Hypervariable 'minisatellite' regions in human DNA. *Nature*, **314**, 67–73.
6. Albers, J. J., Marcovina, S. M. and Lodge, M. S. (1990) The unique lipoprotein(a): properties and immunochemical measurement. *Clin. Chem.*, **36**, 2019–2026.
7. Boerwinkle, E., Leffert, C. C., Lin, J., Lackner, C., Chiesa, G. and Hobbs, H. H. (1992) Apolipoprotein(a) gene accounts for greater than 90% of the variation in plasma lipoprotein(a) concentrations. *J. Clin. Invest.*, **90**, 52–60.
8. Cohen, J. C., Chiesa, G. and Hobbs, H. H. (1993) Sequence polymorphisms in the apolipoprotein(a) gene. Evidence for dissociation between apolipoprotein(a) size and plasma lipoprotein(a) levels. *J. Clin. Invest.*, **91**, 1630–1636.
9. Perombelon, Y. F. N., Soutar, A. K. and Knight, B. L. (1994) Variation in lipoprotein(a) concentration associated with different apolipoprotein(a) alleles. *J. Clin. Invest.*, **93**, 1481–1492.
10. Rader, D. J., Cain, W., Ikewaki, K., Talley, G., Zech, L. A., Usher, D. and Brewer, H. B. (1994) The inverse association of plasma lipoprotein(a) concentrations with apolipoprotein(a) isoform size is not due to differences in Lp(a) catabolism but to differences in production rate. *J. Clin. Invest.*, **93**, 2758–2763.
11. Azrolan, N., Gavish, D. and Breslow, J. L. (1991) Plasma lipoprotein(a) concentration is controlled by apolipoprotein(a) (Apo(a)) protein size and the abundance of hepatic apo(a) mRNA in a cynomolgus monkey model. *J. Biol. Chem.*, **266**, 13866–13872.
12. Wade, D. P., Clarke, J. G., Lindahl, G. E., Liu, A. C., Zysow, B. R., Meer, K., Shwartz, K. S. and Lawn, R. M. (1993) 5' control regions of the apolipoprotein(a) gene and members of the related plasminogen gene family. *Proc. Natl Acad. Sci. USA.*, **90**, 1369–1373.
13. Hedrick, P. W. (1987) Gametic disequilibrum measures: Proceed with caution. *Genetics*, **117**, 331–341.

14. Economou, E. P., Bergen, A. Q., Warren, A. C. and Antonarakis, S. E. (1991) The polydeoxyadenylate tract of *Alu* repetitive elements is polymorphic in the human genome. *Proc. Natl Acad. Sci. USA*, **87**, 2951–2954.
15. Zuliani, G. and Hobbs, H. H. (1990) A high frequency of length polymorphisms in repeated sequences adjacent to *Alu* sequences. *Am. J. Hum. Genet.*, **46**, 963–969.
16. Magnaghi, P., Citterio, E., Malgaretti, N., Acquati F., Ottolenghi, S. and Taramelli, R. (1994) Molecular characterization of the human apo(a)-plasminogen gene family clustered on the telomeric region of chromosome 6 (6q26–27). *Hum. Mol. Genet.*, **3**, 437–442.
17. Kraft, H. G., Mensel, H. J., Hoppichler, F., Vogel, W. and Utermann, G. (1989) Changes of genetic apolipoprotein phenotypes caused by liver transplantation. Complications for apolipoprotein synthesis. *J. Clin. Invest.*, **83**, 137–142.
18. Wade, D. P., Lindahl, G. E. and Lawn, R. M. (1994) Apolipoprotein(a) gene transcription is regulated by liver-enriched *trans*-acting factor hepatocyte nuclear factor 1α. *J. Biol. Chem.*, **269**, 19757–19765.
19. Helmhold, M., Bigge, J., Muche, R., Mainoo, J., Thiery, J., Seidel, D. and Armstrong, V. W. (1991) Contribution of the apo(a) phenotype to plasma Lp(a) concentrations shows considerable ethnic variation. *J. Lipid Res.*, **32**, 1919–1928.
20. Sandholzer, C., Hallman, D. M., Saha, N., Sigurdsson, G., Lackner, C., Csàszàr, A., Boerwinkle, E. and Utermann, G. (1991) Effects of the apolipoprotein(a) size polymorphism on the lipoprotein(a) concentration in 7 ethnic groups. *Hum. Genet.*, **86**, 607–614.
21. Gaw, A., Boerwinkle, E., Cohen, J. C. and Hobbs, H. H. (1994) Comparative analysis of the *apo(a)* gene, apo(a) glycoprotein and plasma concentrations of Lp(a) in three ethnic groups—evidence for no common 'null' allele at the *Apo(a)* locus. *J. Clin. Invest.*, **93**, 2526–2534.
22. Saiki, R. K., Scharg, S., Faloona, F., Mullis, K. B., Horn, G. T., Erlich, H. A. and Arneheim, N. (1985) Enzymatic amplification of β-globin genomic sequences and restriction site 1 analysis for diagnosis of sickle cell anemia. *Science*, **230**, 1350–1354.
23. Cohen, J. C., Cali, J. J., Jelinek, D. F., Mehrabian, M., Sparkes, R. S., Lusis, A. J., Russell, D. W. and Hobbs, H. H. (1992) Cloning of the human cholesterol 7α-hydroxylase gene (CYP7) and localization to chromosome 8q11–q12. *Genomics*, **14**, 153–161.
24. Chakraborty, R., Zhong, Y., DeAndrade, M., Clemens, P. R., Fenwick, R. G. and Caskey, C. T. (1994) Linkage disequilibria among $(CA)^n$ polymorphisms in the human dystrophin gene and their implications in carrier detection and prenatal diagnosis in Duchenne and Becker muscular dystrophies. *Genomics*, **21**, 567–570.
25. Weir, B. S. and Cockerham, C. C. (1978) Testing hypotheses about linkage disequilibrium with multiple alleles. *Genetics*, **88**,633–642.
26. Brunner, C., Kraft, H. G., Utermann, G. and Muller, H. J. (1993) Cys^{4057} of apolipoprotein(a) is essential for lipoprotein(a) assembly. *Proc. Natl Acad. Sci. USA*, **90**, 11643–11647.
27. DeWet, J. R., Wood, K. V., DeLuca, M., Helinski, D. R. and Subramani, S. (1987) Firefly luciferase gene: structure and expression in mammalian cells. *Mol. Cell. Biol.*, **7**, 725–737.