

A Pentanucleotide Repeat Polymorphism in the 5' Control Region of the Apolipoprotein(a) Gene Is Associated with Lipoprotein(a) Plasma Concentrations in Caucasians

Marion Trommsdorff, Silvano Köchl, Arno Lingenhel, Florian Kronenberg, Rhena Delport,* Hyward Vermaak,* Lone Lemming,† Ib Christian Klausen,‡ Ole Faergeman,‡ Gerd Utermann, and Hans-Georg Kraft

Institute of Medical Biology and Human Genetics, Innsbruck, Austria; *Institute of Chemical Pathology, University of Pretoria, South Africa; and †Department of Medicine and Cardiology, Aarhus Amtssygehus University Hospital, 8000 Aarhus, Denmark

Abstract

The enormous interindividual variation in the plasma concentrations of the atherogenic lipoprotein(a) [Lp(a)] is almost entirely controlled by the apo(a) locus on chromosome 6q26–q27. A variable number of transcribed kringle4 repeats (K4-VNTR) in the gene explains a large fraction of this variation, whereas the rest is presently unexplained. We here have analyzed the effect of the K4-VNTR and of a pentanucleotide repeat polymorphism (TTTTA)_n ($n = 6-11$) in the 5' control region of the apo(a) gene on plasma Lp(a) levels in unrelated healthy Tyroleans ($n = 130$), Danes ($n = 154$), and Black South Africans ($n = 112$). The K4-VNTR had a significant effect on plasma Lp(a) levels in Caucasians and explained 41 and 45% of the variation in Lp(a) plasma concentration in Tyroleans and Danes, respectively.

Both, the pentanucleotide repeat (PNR) allele frequencies and their effects on Lp(a) concentrations were heterogeneous among populations. A significant negative correlation between the number of pentanucleotide repeats and the plasma Lp(a) concentration was observed in Tyroleans and Danes. The effect of the 5' PNRP on plasma Lp(a) concentrations was independent from the K4-VNTR and explained from 10 to 14% of the variation in Lp(a) concentrations in Caucasians. No significant effect of the PNRP was present in Black Africans. This suggests allelic association between PNR alleles and sequences affecting Lp(a) levels in Caucasians. Thus, in Caucasians but not in Blacks, concentrations of the atherogenic Lp(a) particle are strongly associated with two repeat polymorphisms in the apo(a) gene. (*J. Clin. Invest.* 1995. 96:150–157.) Key words: lipoprotein(a) • atherosclerosis • kringle4 • population genetics • quantitative trait

Introduction

The analysis of quantitative and complex traits in humans is still a major challenge. The plasma concentration of lipoprotein(a)

[Lp(a)]¹ is a remarkable example for a quantitative genetic trait in humans (1). Concentrations vary over 1000-fold among subjects in the population, but are very stable in healthy individuals over time. The distribution of Lp(a) concentrations is heterogeneous across populations (1–3). Lp(a) consists of a low density lipoprotein, (LDL) particle and a high molecular weight glycoprotein designated apolipoprotein(a), [apo(a)] (1). The latter exhibits a genetic size polymorphism with over 30 alleles, which can be demonstrated at the DNA, RNA, and protein level (4–8). This allelic variation is caused by a variable number of tandemly repeated (VNTR) 5.6-kb modules, called kringles (6, 7), which are homologous to kringle4 from the plasminogen gene (9).

Numerous studies have shown that elevated Lp(a) plasma levels and small apo(a) isoforms are associated with premature coronary heart disease and stroke and represent an inherited risk for atherothrombotic disease (10, 11). Several in vitro functions have been assigned to apo(a) and Lp(a). These include the function as an "interloper" into the fibrinolytic system (12) and binding to fibrin (13). Furthermore, apo(a) inhibits the activation of transforming growth factor β in serum and the aortic wall of mice expressing human apo(a) (14). This may explain the potential of Lp(a) to enhance proliferation of human vascular smooth muscle cells (15), a process which is crucial in the formation of atherosclerotic plaques (16). The involvement in the pathogenesis of atherosclerosis qualifies apo(a) as a complex disease gene.

Sib pair linkage studies have demonstrated that Lp(a) plasma levels in healthy subjects are almost entirely controlled by variation at the apo(a) gene locus on chromosome 6q26–27 (7, 17). In Caucasian populations, ~ 50% of this variation is explained by the apo(a) gene size polymorphism, which is inversely correlated to Lp(a) concentration (7, 17, 18). A major effect of the apo(a) K4-VNTR has also been observed in Asian populations (3, 4). Surprisingly, the impact of the K4 repeat polymorphism on Lp(a) plasma levels is small in African Blacks (2, 3), though the size polymorphism is extensive and mean Lp(a) levels are higher in Blacks than in other ethnic groups (1, 2, 19). With the exception of the K4 repeat polymorphism, no other variation in the apo(a) gene with an effect on Lp(a) plasma levels has yet been identified in any population.

In an attempt to characterize the sequences, which control apo(a) transcription, the 5' region of the apo(a) gene was identified, cloned, and sequenced (20–22). Based on sequences from two subjects, Wade et al. (21) postulated that variation

Address correspondence to Prof. Dr. Gerd Utermann, Institute of Medical Biology and Human Genetics, Schöpfstr. 41, 6020 Innsbruck, Austria. Phone: 512-507-3451; FAX: 512-507-2861.

Received for publication 20 January 1995 and accepted in revised form 4 April 1995.

J. Clin. Invest.

© The American Society for Clinical Investigation, Inc.

0021-9738/95/07/0150/08 \$2.00

Volume 96, July 1995, 150–157

1. Abbreviations used in this paper: d.f., degree of freedom; Lp(a), lipoprotein(a); PFGE, pulse field gel electrophoresis; PNR, pentanucleotide repeat; PNRP, pentanucleotide repeat polymorphism; VNTR, variable number of tandem repeats.

in the length of a pentanucleotide repeat (TTTTA)_n (starting 1,373 bp upstream of the signal sequence) and/or a base substitution (914 bp upstream of the ATG) might be responsible for different promoter activities and thereby Lp(a) plasma levels. This has prompted extensive population genetic studies of apo(a) promoter variation in our laboratory. A specific PCR assay was developed to amplify the region comprising the pentanucleotide repeats (PNR). This was used to analyze whether a common polymorphism exists in the population, to determine the numbers of repeats and the frequencies of PNR alleles, and to study their potential impact on Lp(a) plasma concentrations in different populations.

Methods

Population samples. EDTA blood was obtained from healthy, unrelated individuals from three different populations. Samples from 130 unrelated Tyrolean blood donors were collected by the Blood Transfusion Service of the University of Innsbruck (73 women and 57 men; mean age: 39.9±11.2 yr). Blood samples were further obtained from 154 unrelated, ethnically homogeneous Danish men, who were randomly selected from the National Register of all men born in 1948 (*n* = 1954). All of them were 40 yr old when analyzed. 112 blood samples were further collected from unrelated, apparently healthy Black South African male policemen and from students attending the Medunsa University, South Africa (mean age: 42.7±20.0 yr).

DNA preparation. Lymphocytes were prepared from 10 ml EDTA blood by Ficoll-Paque centrifugation. The cell number was determined using a Coulter counter (Coulter Electronics, Dunstable, United Kingdom) and adjusted to 2.10⁷ cells/ml. Equal volumes of cell suspensions and 1% LMP agarose were mixed. The mixture was rapidly poured into plug-forming molds. The plugs were then digested with 2 mg/ml Protease K (Boehringer Mannheim, Mannheim, Germany) at 55°C for 48 h. After digestion the plugs were washed in TE buffer, then in TE with PMSF (0.04 mg/ml) and finally again in TE. They were stored in 0.5 M EDTA, pH 8, at 4°C or used immediately. One plug contains ~ 7 μg of genomic DNA.

Analysis of the pentanucleotide repeat polymorphism (PNRP). The 5' control region, starting 1,373 bp before the signal codon and containing the pentanucleotide repeat, was amplified by PCR (20). Before PCR the plugs were washed twice for half an hour in TE buffer, mixed with 100 μl H₂O bidest and melted for 5 min at 65°C. The set of primers (MWG-Biotech, Ebersberg, Germany) used were the following: Pro 1A 5'GAATTCATTTCGGGAAAGATTG 3' and Pro 1B 5'CTTCAACCGGGGTGAGAGTCTC 3'. 50 pmol of each primer was mixed with 5 μl of 10× incubation buffer (Boehringer Mannheim), 5 μl of 1 mM dNTP mix (Boehringer Mannheim), 5–8 μl of the melted plugs (0.2–0.3 μg DNA), and 0.3 μl of *Taq* Polymerase (Boehringer Mannheim). The final reaction volume was 50 μl. The reaction solution was overlaid with 100 μl of mineral oil. The PCR reaction was performed in a Biometra Thermocycler (Bio Rad Laboratories, Hercules, CA) under the following conditions: first cycle: 94°C for 1.30 min; the following 35 cycles: denaturation at 94°C for 40 s, annealing at 59°C for 30 s, extension at 72°C for 40 s; final extension at 72°C for 5 min. 10 μl of the PCR product was first tested on a 2.5% agarose gel and then loaded on a 10% polyacrylamide gel. Electrophoretic conditions were 4 h at 400 V, 150 mA, 30 W. The gels were stained with ethidium bromide and photographed under ultraviolet light.

Sequencing of PCR-amplified PNRP fragments. DNA from four individuals homozygous for the (TTTTA)_n repeat (i.e., with 8, 9, 10 pentanucleotide repeats) was amplified by PCR using primers PRO 1A and PRO 1B. Primer PRO 1A was biotinylated. 20 μl of the PCR product was treated with Dynabeads M-280 Streptavidin (Dynal, Oslo, Norway) for single strand separation. Each strand was sequenced using Sequenase version 2.0 (United States Biochemical Corp., Cleveland, OH) (23).

Analysis of apo(a) KpnI fragments. Plugs were digested with 2

× 40 U KpnI enzyme (Boehringer Mannheim) following the instructions of the supplier. They were then size separated in the CHEF mapper system (Bio Rad Laboratories) using its autoalgorithm to determine the best parameters. Electrophoresis was performed in 1% LE agarose (SeaKem; FMC, Rockland, ME) in 0.5% TAE buffer at 14°C. After electrophoresis, the gel was stained with ethidium bromide. DNA was nicked with 60 mJ ultraviolet light in the Gene Linker (Bio Rad Laboratories) and transferred to Hybond-N membrane (Amersham International, Amersham, United Kingdom) using the alkaline buffer method. To detect the KpnI alleles, the membrane was hybridized with a digoxigenin-labeled 340-bp PstI fragment derived from the apo(a) cDNA construct clone Lambda a41 (9) as already described (7).

Determination of Lp(a) concentrations. Plasma Lp(a) levels were measured using an enzyme-linked immunosorbent sandwich assay (ELISA) (24). The capture antibody was a polyclonal rabbit anti-human Lp(a) antibody and the detection antibody was our mouse monoclonal anti-human apo(a) antibody, IgG-1A2.

Statistical methods. Allele frequencies were estimated by gene counting. χ^2 statistics was used to test for Hardy-Weinberg equilibrium and to compare the distribution of KpnI alleles within phenotypes of the PNRP. To avoid a too high number of empty cells the phenotypes of the PNRP were combined before χ^2 analysis according to their smaller allele number (i.e., in the Tyrolean samples phenotypes with at least one allele 7, phenotypes with at least one allele 8, phenotypes with at least one allele 9, and phenotypes with at least one allele 10 or 11 were combined) into 4 and 5 groups in the Caucasians and the African Blacks, respectively. Comparisons of PNR allele frequencies in the different samples were also carried out using a Monte-Carlo simulation method (25). To stratify PNR phenotypic groups for the confounding effects of the KpnI size polymorphism, we have estimated the average Lp(a) concentration for binned KpnI alleles. The average apo(a) KpnI allele associated Lp(a) concentrations were calculated for each population. For this analysis we used data from 224 Tyroleans, 154 Danes, and 239 Blacks, where both, KpnI alleles and Lp(a) concentrations, had been determined. An additive model for the effect of apo(a) alleles on plasma Lp(a) levels was assumed (7). Because of the large number of apo(a) size alleles, these were binned before analysis (26). Each bin contains the KpnI fragments, which correspond to the originally described protein isoforms (4). Accordingly, the bins will be designated F, B, S1, S2, S3, S4. These correspond to KpnI fragments with 11–13, 14–16, 17–19, 20–22, 23–25, and 26–52 kringle4 repeats, respectively (7). Mean Lp(a) concentrations were estimated for the binned alleles as follows. The mean Lp(a) concentration of S4/S4 homozygotes was 9.5 mg/dl. The S4 allele associated mean Lp(a) concentration was therefore 4.75 mg/dl. The Lp(a) concentration associated with KpnI alleles with 23–25 KIV repeats (S3 isoform group) was calculated next from the combined groups of S3/S3 homozygotes and S3/S4 heterozygotes (the number of S3/S3 homozygotes was too low for separate analysis). The S3 specific mean concentration was calculated from the following equation: $G3 = ([S3] + [S3])p + ([S3] + [S4])(1 - p)$, therefore $[S3] = (G3 - (1 - p)[S4]) / (p + 1)$. *G3* is the mean Lp(a) concentration of the group of S3/S3 and S4/S3 phenotypes. $[S3]$ and $[S4]$ are the S3 and S4 specific Lp(a) concentrations. *p* is the frequency of S3/S3 homozygotes within this group. $1 - p$ is the frequency of S3/S4 heterozygotes within this group. The apo(a) allele-associated Lp(a) concentrations for the S1, S2, B, and F alleles were calculated accordingly. The effect of the K4 repeat polymorphism on Lp(a) levels was determined as outlined (2, 7).

To estimate the effect of the PNRP on Lp(a) plasma levels, two methods were used. First, expected Lp(a) concentrations were calculated for each pentanucleotide repeat phenotype from the frequencies of KpnI alleles within a phenotype and from the KpnI allele-associated Lp(a) concentrations. The impact of the pentanucleotide repeat alleles on the variation of Lp(a) concentration was estimated from the difference between measured and expected Lp(a) concentrations [Δ Lp(a)] by *R*² statistics of the analysis of variance (27). The distribution of the Δ Lp(a) concentration was found to be normal in all three populations.

Second, numbers of KpnI and pentanucleotide repeat alleles were tested as independent variables in a multivariate regression analysis. All

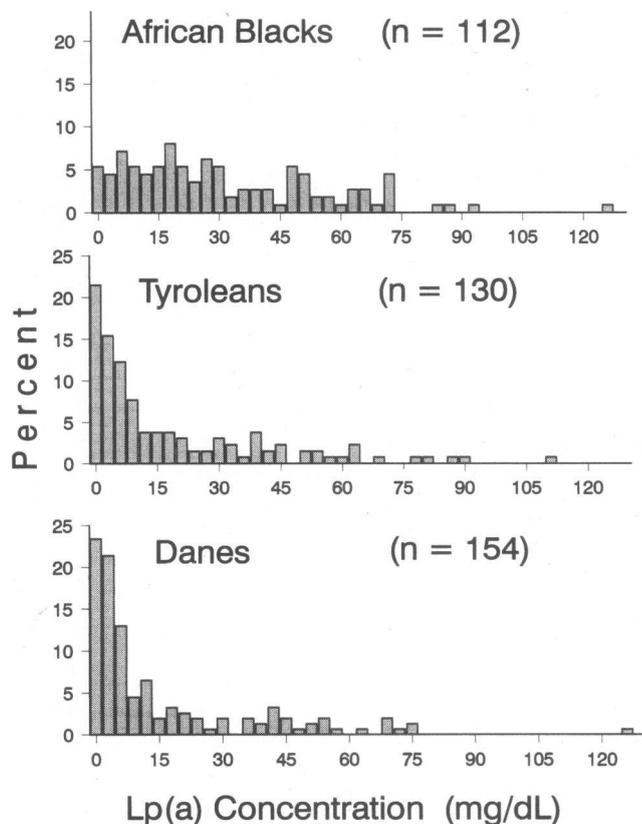


Figure 1. Histograms demonstrating the frequency distributions of Lp(a) plasma concentrations in the three study populations (Blacks, Tyroleans, and Danes).

calculations were performed with the SPSS for Windows release 5.0 package (SPSS Inc., Chicago, IL).

Results

Lp(a) plasma levels and KpnI alleles. To assess the impact of the PNRP on Lp(a) levels, it was necessary to first determine the frequencies and effects of the K4 repeat polymorphism in the same population samples. This enabled us to stratify for K4 repeat variation, which is known to have a marked effect on Lp(a) concentrations (7, 17). In our analysis KpnI alleles, apo(a) isoform phenotypes, and Lp(a) plasma levels were determined in unrelated subjects from three ethnically defined populations: 130 Tyroleans, 154 Danes, and 112 Black South Africans. A detailed description of the relation between apo(a)

K4 repeats, apo(a) isoforms, and Lp(a) concentrations in the studied populations will be given elsewhere (Lingenhel, A., M. Trommsdorff, R. Delpont, H. Verwaak, H. G. Kraft, and G. Utermann, manuscript in preparation). The distribution of Lp(a) concentrations in each population is shown in Fig. 1 and the mean and median Lp(a) levels are given in Table I. The range of Lp(a) concentrations was similar in all three populations. The distributions were skewed toward lower concentrations in all four populations but less so in the African than in the Caucasian groups. Skewness was highest in the Danes (2.06) and lowest in Blacks (0.99) (Table I). Average Lp(a) levels were significantly higher in Blacks than in Caucasians. Lp(a) concentrations in the South African Blacks were lower than in the Blacks from the Sudan (2) and from Zaire (28) previously studied by us and others. No correlation of Lp(a) concentration with age or sex was noted in any of the populations. Therefore unadjusted Lp(a) concentrations were used in further analysis.

Apo(a) K4 repeat phenotypes (KpnI fragments) were determined in all study subjects by pulse field gel electrophoresis (PFGE)/Southern blotting of KpnI-digested genomic DNA (6, 7). The distribution of KpnI alleles was broad in all populations (Fig. 2). There were a total of 38 alleles represented by the pooled population sample. The smallest apo(a) allele had 11 and the largest had 52 K4 repeats. In each population alleles of repeat length of ~30 K4 repeats were most common.

The number of K4 repeats in the apo(a) gene correlated inversely with Lp(a) plasma concentrations (data not shown). Using the R^2 from the analysis of variance 35.7–45.6% of the variation in Lp(a) concentrations was explained by the K4 repeat polymorphism in the three populations. This effect was not significant in Blacks. Thus, as in a previous study where we used apo(a) isoform size as a measure, the impact of the K4 repeat number was smaller in Blacks than in other populations.

Frequencies of pentanucleotide repeat alleles. The region starting at -1,373 bp upstream of the AUG codon and comprising the pentanucleotide repeats (20) was amplified by PCR using primers PRO 1A and PRO 1B (see Methods) from genomic DNA of all subjects from the three populations. PCR products from at least one subject in each population were sequenced and found to be identical except for the pentanucleotide repeat number. In the pooled sample we detected 6 different alleles containing 6, 7, 8, 9, 10, or 11 pentanucleotide repeats with PCR product lengths ranging from 86 to 111 bp (Fig. 3). The most common allele had 8 (TTTTA) repeats. Alleles with 9 or 10 repeats were also common, whereas the alleles with 7 or 11 repeats were rare (Table II). An unimodal pattern of allelic distributions was present in all populations (Fig. 4). Allele frequency distributions in the Caucasians were highly skewed (Fig. 4).

Table I. Plasma Lp(a) Concentrations (mg/dl) in the Three Populations Studied

Population	Lp(a) concentration		Median	Range		Skewness	Percentage explained*
	Mean	SD		Minimum	Maximum		
Tyroleans	19.2	22.6	9.25	0.4	111	1.679	41.1 [†]
Danes	16.6	21.3	7.11	1.0	128	2.064	45.6 [†]
Blacks	32.7	24.3	27.0	0.3	126	0.990	35.7 ⁿ

* The percentage of the total variance of Lp(a) concentration explained by the K4 polymorphism was estimated from the R^2 of the analysis of variance of log transformed Lp(a) values. [†] $P < 0.001$. n, not significant.

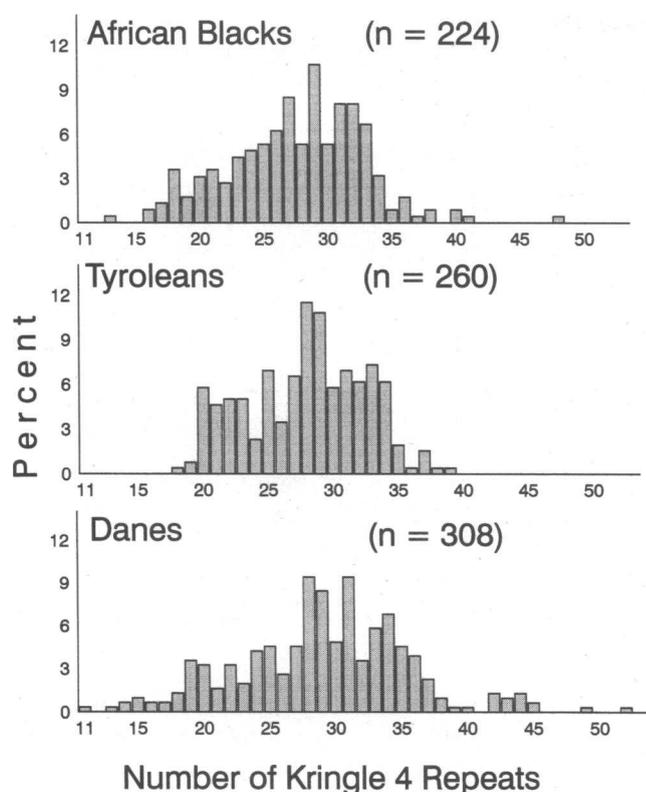


Figure 2. Histograms demonstrating the frequency distributions of apo(a) KpnI alleles (K4-VNTR alleles) in the three population samples studied. The number of K4 repeats was determined using PFGE and genomic blotting, as outlined in reference 7.

The allele frequency distribution was not significantly different between the Caucasian samples (i.e., the Tyroleans and the Danes) ($\chi^2 = 3.8$; degree of freedom [d.f.] = 3; $P = 0.2$). The allele frequency distribution was significantly different between the Tyroleans and the Blacks ($\chi^2 = 30.07$; d.f. = 4; $P < 0.0001$) and between the Danes and the Blacks ($\chi^2 = 37.32$; d.f. = 4; $P < 0.0001$). In Africans smaller alleles were more frequent, and an allele with only six (TTTTA) repeats was present only in Africans (Table II and III).

In all samples the PNR allele frequencies were in Hardy-Weinberg equilibrium (Table III). This was confirmed by a Monte-Carlo simulation test. The observed heterozygosity ranged from 43% in Danes to 50% in Tyroleans. Thus, in all

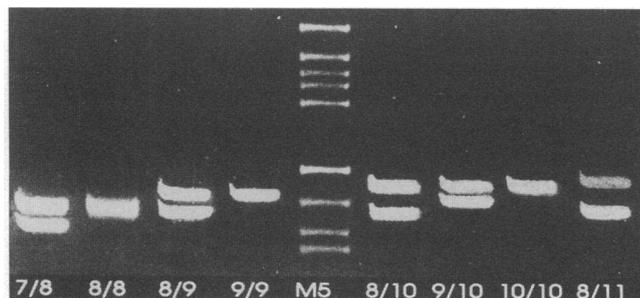


Figure 3. Ethidium bromide-stained polyacrylamide gel demonstrating the pentanucleotide repeat polymorphism. The 5' apo(a) region (nt -1442 to nt -1343 [20]) was amplified by PCR and subjected to polyacrylamide gel electrophoresis. PNR phenotypes are indicated below. The middle lane contains a marker.

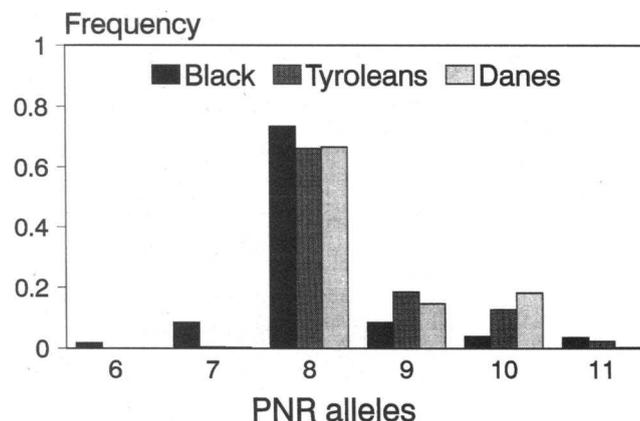


Figure 4. Histogram demonstrating the frequency distributions of PNR alleles in the three study populations.

three ethnic groups, a common PNRP exists in the apo(a) promoter region.

Effect of the PNRP on Lp(a) concentrations in Caucasians. First we analyzed the effect of the PNRP on plasma Lp(a) concentrations in the sample of healthy, unrelated Tyroleans. We categorized the individuals according to their PNR phenotype and compared the mean Lp(a) concentrations between these groups. Although the mean Lp(a) values were markedly different in the PNR phenotypes (i.e., 27.1 mg/dl for the 8/10 phenotype and 5.0 mg/dl for the 9/10 phenotype) no systematic correlation between the number of (TTTTA) repeats and the mean Lp(a) levels was noted (data not shown). This analysis did not consider possible differences in K4 repeat allele frequencies between PNR phenotypes, especially for small cells and the strong effect of the K4 repeat polymorphism on Lp(a) levels. Therefore, in a next step the effect of the KpnI alleles on Lp(a) concentrations was taken into account. Two types of analyses were performed. First, we calculated an expected Lp(a) concentration for each PNR type from the frequencies of the KpnI alleles within the PNR phenotypic groups and from the mean apo(a) KpnI allele-associated Lp(a) concentration (see Methods). We then calculated the difference (Δconc) between the measured and the expected mean Lp(a) concentrations for each PNR phenotype (Table IV). This value reflects the effect of the PNRP on Lp(a) concentrations. A significant negative correlation between the PNRP and the Δconc was noticed ($R = -0.323$; $P < 0.001$). This association is graphically illustrated in Fig. 5. The $\Delta\text{Lp(a)}$ concentration within a

Table II. Frequencies of PNR Alleles at the 5' Control Region of the apo(a) Gene

PNR allele	Caucasians			Blacks (n = 112)
	Tyroleans (n = 130)	Danes (n = 154)	Pooled (n = 284)	
6	0.000	0.000	0.000	0.018
7	0.004	0.003	0.003	0.085
8	0.661	0.666	0.664	0.734
9	0.185	0.146	0.164	0.085
10	0.127	0.182	0.156	0.040
11	0.023	0.003	0.012	0.036

Table III. Observed and Expected Frequencies of PNR Phenotypes in the Three Studied Populations*

PNR type	Blacks [‡]		Tyroleans [§]		Danes	
	Observed	Expected	Observed	Expected	Observed	Expected
6/6	0	0.036				
6/7	0	0.340				
6/8	3	2.930				
6/9	0	0.340				
6/10	0	0.160				
6/11	1	0.144				
7/7	0	0.800	0	0.002	0	0.001
7/8	14	13.850	1	0.642	0	0.615
7/9	2	1.600	0	0.181	0	0.135
7/10	1	0.755	0	0.152	1	0.168
7/11	1	0.680	0	0.019	0	0.002
8/8	60	59.800	47	42.050	73	68.300
8/9	14	10.160	19	23.750	26	29.900
8/10	7	6.520	13	19.890	33	37.330
8/11	5	5.870	4	2.567	0	0.615
9/9	0	0.800	5	3.350	7	3.280
9/10	1	0.755	8	5.618	5	8.180
9/11	2	0.680	0	0.725	0	0.135
10/10	0	0.178	0	2.353	8	5.100
10/11	0	0.320	0	0.588	1	0.168
11/11	0	0.144	0	0.039	0	0.001

* To test for Hardy-Weinberg equilibrium, phenotypes had to be combined in order to avoid too many cells (see Methods). [‡] $\Sigma\chi^2 = 1.067$, d.f. = 4, $P = 0.8$; [§] $\Sigma\chi^2 = 0.885$, d.f. = 3, $P = 0.8$; ^{||} $\Sigma\chi^2 = 2.786$, d.f. = 3, $P = 0.5$.

series of types with increasing repeat numbers decreases from +7.2 mg/dl (phenotype 8/8) to -14.6 mg/dl (phenotype 8/11).

Second, we performed a multivariate linear regression analysis considering K4 repeat variation and the PNR. This also demonstrated significant independent effects of the K4-VNTR and of the number of pentanucleotides on Lp(a) concentrations ($P = 0.0001$).

Both types of analyses therefore indicate that the PNR polymorphism has an effect on Lp(a) levels which is independent from the apo(a) size polymorphism. The magnitude of this effect was estimated in the Tyrolean sample using the R^2 statistics from the analysis of variance. Lp(a) concentrations were log transformed before analysis to normalize for their non-Gaussian distribution (29). The PNR alleles explain 14.3% of the variation of Lp(a) plasma levels. Together, the KpnI and the PNR alleles explain 55.4% of the variation of Lp(a) concentration in the Tyrolean population.

To confirm our results in an independent sample from the same geographic area we analyzed KpnI and PNR alleles and Lp(a) levels in subjects with coronary heart disease ($n = 100$) from the same Austrian area (Tyrol) by the same statistical procedures. The significant inverse correlation of the pentanucleotide repeat allele size with Lp(a) concentration was confirmed in this population sample, where the PNR explained 8% of the variation in Lp(a) levels (data not shown).

The PNRP was analyzed further in a second Caucasian population sample, namely Danes, applying the same approaches as described above. Pentanucleotide repeat number correlated also negatively with plasma Lp(a) concentrations in the Danes

Table IV. Impact of the PNRP on Plasma Lp(a) Concentrations (mg/dl) in Three Populations Expressed as Deviation (Δ) from the Expected Lp(a) Level* within a PNR Phenotype

PNR phenotype	Blacks		Tyroleans		Danes	
	Δ Lp(a)	n	Δ Lp(a)	n	Δ Lp(a)	n
6/8	-16.6	3	—	—	—	—
6/11	-7.2	1	—	—	—	—
7/7	—	—	—	—	—	—
7/8	4.3	14	-29.00	1	—	—
7/10	-39.4	1	—	—	-2.45	1
7/11	-12.9	1	—	—	—	—
8/8	5.4	60	7.20	58	3.31	73
8/9	-5.8	14	0.97	27	4.65	26
8/10	4.2	7	3.10	22	-5.52	33
8/11	2.1	5	-18.20	6	—	—
9/9	—	—	-9.85	6	-7.81	7
9/10	20.9	1	-8.30	7	-4.38	6
9/11	-3.2	2	—	—	—	—
10/10	—	—	-14.6	3	-14.7	7
10/11	—	—	—	—	4.7	1
Percentage explained [‡]	6.8		14.3		10.2	
P	0.285		0.046		0.040	

* Δ Lp(a): Lp(a) concentration measured - Lp(a) concentration calculated using the allele-specific Lp(a) concentration as described in Methods. [‡] The percentage explained was calculated using the R^2 statistics from the analysis of variance (see Methods).

(Table IV and Fig. 5). Multivariate regression analysis considering K4 repeats and PNR confirmed the association ($P = 0.0017$). In the Danish population the PNR polymorphism explains ~ 10% of the variation in plasma Lp(a) concentrations.

The relation between apo(a) PNRs and plasma Lp(a) became even clearer when the sum of the PNRs in an individual rather than the PNR phenotype was taken as a measure. Fig. 6 illustrates the pooled data from the two Caucasian samples. It is obvious that an almost linear negative correlation exists between the number of PNRs and plasma Lp(a) concentrations.

Effect of the PNRP on Lp(a) plasma levels in South African Blacks. We next extended our analysis to a non-Caucasian ethnic group: a population sample of South African Blacks. All analyses were performed as described for Caucasians. No significant effect of the PNRP on Lp(a) levels was detected in this sample of Black Africans. The graphic representations in Figs. 5 and 6 demonstrate that no systematic relation between repeat number and Lp(a) concentrations exists in this population.

Stability of the PNRP during meiosis. We analyzed the PNRP in 13 nuclear families to determine the meiotic stability of these repeats. Within 64 meioses we did not find a change in the number of repeats. The pentanucleotide repeat in the apo(a) promoter region thus seems to belong to a group of relatively stable repeats in the human genome. This is consistent with the unimodal distribution of allele frequencies in the three different ethnic groups, which all have allele 8 as the most frequent.

Discussion

PNR allele frequencies. Here we demonstrate the existence of a common PNRP in the 5' region of the apo(a) gene. This

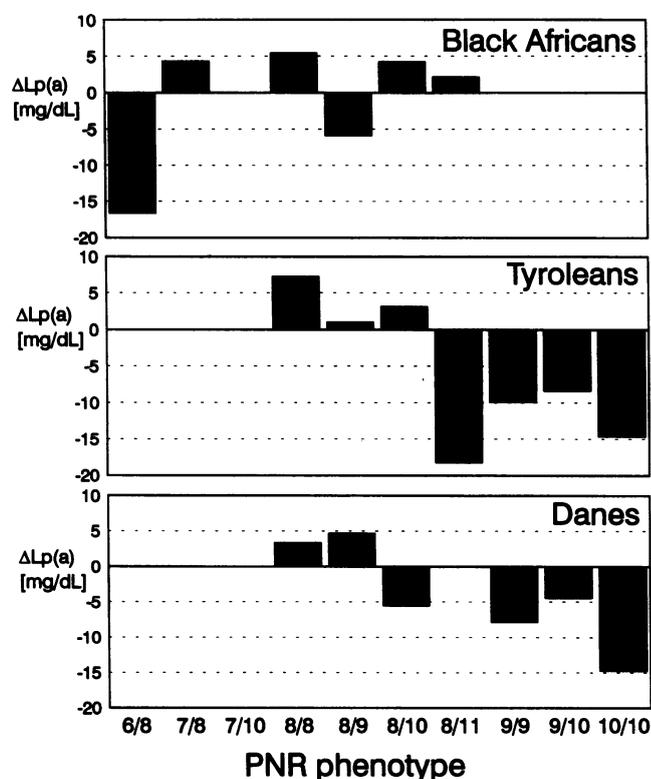


Figure 5. Histograms illustrating the impact of the pentanucleotide repeat polymorphism on Lp(a) plasma concentrations in the three study populations. Δ Lp(a) denotes the difference between measured and expected Lp(a) concentrations. Only phenotypes with $n > 3$ are shown (for details see text). Note that in the Caucasian samples PNRs with higher repeat numbers ($n = 9, 10, 11$) are associated with Lp(a) concentrations below expectation.

polymorphism probably evolved by duplication or by slippage mutation during DNA replication, as suggested for other repetitive sequences in the human genome (30–32). Allele frequencies at this locus are significantly different between Tyroleans and Blacks and Danes and Blacks ($\chi^2 = 30.07$; d.f. = 4; $P < 0.0001$ and $\chi^2 = 37.32$; d.f. = 4; $P < 0.0001$, respectively). In the Black population sample, alleles with lower repeat numbers were more frequent and we detected an allele with only six repeats not present in the Caucasian samples. As the African population is believed to be older than the Caucasian populations (33), this distribution difference may reflect the combined effect of mutation, migration, and genetic drift.

Tandem arrays of simple repeated sequences are widespread among eukaryotes. The specific (TTTTA) $_n$ repeat is also present in other human genes, e.g., in the β -globin gene, where 4, 5, or 6 repeats are found about 1,400 bp upstream of the signal sequence. This variation has no effect on β -globin promoter activity (31, 34). The PNRP in the β -globin gene seems to be a mutation hotspot and undergoes frequent duplication/deletion events. This is in contrast to the PNRP in the apo(a) gene, which is stable through meiosis, as shown by analysis of 64 meioses.

Effect of the PNRP on Lp(a) plasma levels. When analyzing the relation of the PNRP to Lp(a) levels, we faced the problem of detecting a possible effect of the PNRP in view of the known strong effect of the K4 repeat polymorphism (7, 17). Random deviations of K4 repeat allele frequencies in PNR types might

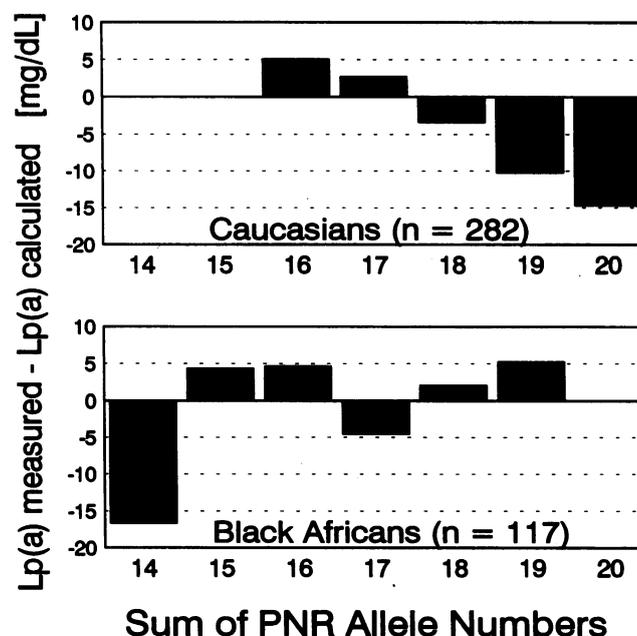


Figure 6. Histogram illustrating the relation between the sum of 5' apo(a) PNRs and plasma Lp(a) concentration in the pooled Caucasian samples (Tyroleans and Danes) and in the South African Blacks.

result in spurious effects or might mask true effects. Further allelic associations between K4 repeats and PNR alleles might also suggest an impact of PNR alleles which in fact only reflects the known effect of the K4 repeat. Theoretically, several approaches are possible to overcome this problem. PNRP groups may be matched for KpnI allele frequencies or phenotypes. An even more rigorous approach would be to analyze the PNRP and its effect on Lp(a) levels in subjects with identical KpnI alleles. This would require analyzing very large samples for the KpnI polymorphism by PFGE and Southern blotting. Due to the large number of apo(a) KpnI phenotypes, this approach is hardly possible.

We have chosen to perform a mathematical matching for KpnI types in our analysis of the data. We calculated expected Lp(a) concentrations for each PNR phenotypic class from the KpnI allele frequencies in the respective cells and from the mean Lp(a) concentrations associated with binned KpnI alleles (see Methods). The expected value should be identical with the one observed if the PNRP had no effect on the trait. Any significant deviation from the expected value reflects the effect of the factor under analysis, e.g., the PNRP. When we analyzed randomly selected subgroups from the Tyrolean sample for deviation of observed from expected values (data not shown) or applied the same procedure to estimate the effect of a Met/Thr polymorphism in kringle37 on Lp(a) levels (35), no significant deviations were found.

In contrast, observed Lp(a) concentrations in the PNRP groups did deviate significantly from expectation in two of the ethnic groups analyzed (Table IV). Importantly this deviation was not random but ordered, e.g., there was a clear inverse correlation of the pentanucleotide repeat number with Lp(a) levels in both samples (Figs. 5 and 6). In both populations (Tyroleans and Danes) in which the impact of the PNRP had been detected by the original analyses, regression analysis also showed a significant independent effect. Again no association of the PNRP with Lp(a) levels was seen in Blacks.

Association studies are prone to several sources of bias. Therefore stringent criteria have to be fulfilled before an association may be considered real. The most stringent criterion is the reproducibility of results (36). This criterion is fulfilled here. We consider it strong evidence against a chance result, that the same relation was obtained in two different ethnic groups (Tyroleans and Danes) and in two independent samples from the same ethnic group (Tyroleans from Austria).

We are aware that the significance level in each population singly is not high with a possibility for a false positive association of just below 1:20. For all three samples (Danes and the two independent Tyrolean samples) combined, the probability of a false positive result decreases to < 1:8,000. A further argument for a true association is the identical trend of the association with PNR length in the two Caucasian populations (and in the two Tyrolean samples, data not shown). In both Danes and Austrians the highest repeat lengths (10, 11) were associated with the lower Lp(a) concentrations. False positive results are unlikely to produce the same type of association in different samples. Whereas association studies in general have to be met with some scepticism, we believe that in this special situation even a relatively low significance is acceptable. Because the apo(a) gene locus determines > 90% of the variation in Lp(a) plasma levels, there is a high a priori probability that a sequence variation in the gene will be associated with Lp(a) levels, i.e., any sequence in the apo(a) gene is a candidate sequence. Together, these arguments make us believe that the observed association is real.

As pointed out by Rosenberg and Kidd (37) and Kidd (36), a real association will be found under three conditions. First, the marker studied is the etiologically relevant variable. This is not likely to be the case for the PNRP studied here. First we are not aware of a situation where a short repeat sequence in a gene affects transcriptional activity with the only exception of the trinucleotide repeat expansion in the fragile X mental retardation syndrome (FRA-X). It has been shown that the extensive expansion of trinucleotide repeats 5' of the FMR-1 gene results in a decrease in FMR-1 mRNA levels (38, 39). However, no such relation has been reported for the normal range of repeats of the FMR-1 locus or other loci. The second and stronger argument is the absence of an association in Blacks (Figs. 5 and 6). If the observed association reflects a causal relationship, it is expected to be present in all ethnic groups, which is not the case. Furthermore, we and others did not find an effect of PNR number on transcriptional activity in a reporter gene assay when individual 1.4 kb 5' apo(a) sequences containing the PNR were expressed in HepG2 cells (40) (Bopp, S., S. Köchl, H. J. Müller, and G. Utermann, unpublished observations).

A second cause underlying a real association is that the sample is a stratified mixture derived from two or more populations, which have different allele frequencies of the marker and of the trait. This situation can be rejected for the Caucasian populations studied here. Both Caucasian samples are ethnically homogeneous and have little or no recent admixture.

The third possibility is that linkage exists between the marker sequence and a sequence with etiologically relevant variation and that there exists linkage disequilibrium between the two sites. This situation most likely applies here and is compatible with the finding that the association was present in Caucasians but not in Blacks. For an association which reflects a linkage disequilibrium between the marker under study and a causal mutation, it is not expected that it is present in different ethnic groups, especially if these have diverged long ago in

human evolution. Therefore it is not surprising that the association which was present in two Caucasian populations (Danes and Austrians) was not found in Blacks. However, it is surprising to find evidence for allelic association at the apo(a) gene locus at all. The apo(a) gene contains a variable but large number of identical 5.6-kb repeats coding for so called kringle4 structures. These are flanked upstream by a single K4 with a different sequence, a large 14-kb intron, the signal sequence and the 5' promoter region containing the PNRP. Downstream the variable K4 repeat region is flanked by a constant region of so called unique kringles, one kringle 5, and the protease domain (9). It is difficult to envisage that strong allelic associations exist over the entire apo(a) gene, e.g., 5' and 3' of the K4 repeat. The association of Lp(a) levels with sequence variation in the 5' region of the gene may therefore suggest that the causal sequence variation is also located here. This narrows the region to be searched for such sequence variation. After submission of this manuscript Zysow et al. (40) reported that a C/T transition in the 5' untranslated region of the apolipoprotein(a) gene at position +93 from the transcription start affects apo(a) promoter activity in a reporter gene assay. It is presently unknown whether this sequence variation is in allelic association with PNR alleles. It is tempting to speculate that the C/T polymorphism is the causal sequence variation responsible for the effect observed here. Our data also indicate that the observed effect may be caused to a large degree by apo(a) alleles with higher PNR numbers (9–11), which are less frequent in the populations. We have observed that PNRs with high repeat numbers (9–11) are in allelic association with K4 repeats with low repeat numbers (Trommsdorff, M., H. G. Kraft, and G. Utermann, unpublished observations). This indicates that a few alleles with high pentanucleotide and low K4 repeat numbers may be associated with low Lp(a) levels. However, there is no evidence that a single PNR type or a single apo(a) allele has a very strong impact and explains the entire impact of the PNRP.

The surprisingly clear inverse relation between plasma Lp(a) concentrations and the added number (from both alleles) of PNRs in the 5' apo(a) region is striking and might contradict this notion. It might together suggest a direct and additive effect of PNR alleles on Lp(a) concentrations. As outlined, such a direct effect is not compatible with the complete absence of an effect in Blacks (Figs. 5 and 6). How then can this almost linear relation in Caucasians be explained? One possibility is that a mutation affecting Lp(a) concentrations occurred on a "Caucasian" allele with an extreme PNR number. Changes in repeat length during evolution then resulted in the spread of this mutation and in a frequency gradient along PNR alleles. Most likely this mutation occurred on an allele with a high PNR number (e.g., 10 or 11 repeats) and had a negative effect on Lp(a) concentrations (see Fig. 5).

The relation of apo(a) gene variation to the concentrations of its products apo(a) and Lp(a) is presently unparalleled: an extreme size polymorphism in the coding region of the gene, which results from a variable number of large transcribed repeats together with a repeat polymorphism in the 5' untranslated promoter region, explains a large fraction of the variation in plasma Lp(a) concentrations in Caucasians.

The relationship between the K4 length variation and plasma Lp(a) concentrations seems to be causal. White et al. (41, 42) have carefully studied the intracellular processing, secretion, and assembly of apo(a)/Lp(a) in baboons. In primary cultures of baboon hepatocytes they found that the transition times of apo(a) through the endoplasmic reticulum and the secretion

rates from the cells are a function of apo(a) size (41). Transfection experiments with recombinant human apo(a), which differed only in K4 repeat number, have also provided direct evidence that differences in K4 repeat number are sufficient to cause differences in secretion rates of recombinant apo(a) from HepG2 cells (Brunner, Ch., H. J. Müller, and G. Utermann, manuscript in preparation). No such evidence exists for the PNRP. Apo(a) promoter fragments which differ in PNR number exhibit comparable activities in HepG2 cells (Bopp, S., S. Köchl, F. Acquati, P. Magnaghi, A. Pethö-Schramm, H. G. Kraft, G. Utermann, H. J. Müller, and R. Taramelli, manuscript submitted for publication). The sequences responsible for the PNR-associated effect still need to be identified.

Acknowledgments

We thank C. Haibach, L. Fineder, and A. Freudenstein for excellent technical assistance.

This work was supported by the Austrian Fonds zur Förderung der Wissenschaftlichen Forschung (FWF) grant S7109 to G. Utermann, and grant P9355 to H. G. Kraft.

References

1. Utermann, G. 1989. The mysteries of lipoprotein(a). *Science (Wash. DC)*. 246:904-910.
2. Sandholzer, C., D. M. Hallman, N. Saha, G. Sigurdsson, C. Lackner, A. Császár, E. Boerwinkle, and G. Utermann. 1991. Effects of the apolipoprotein(a) size polymorphism on the lipoprotein(a) concentration in 7 ethnic groups. *Hum. Genet.* 86:607-614.
3. Gaw, A., E. Boerwinkle, J. C. Cohen, and H. H. Hobbs. 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.
4. Utermann, G., H. J. Menzel, H. G. Kraft, C. H. Duba, H. G. Kemmler, and C. Seitz. 1987. Lp(a) glycoprotein phenotype. Inheritance and relation to Lp(a)-lipoprotein concentrations in plasma. *J. Clin. Invest.* 80:458-465.
5. Kamboh, M. I., R. E. Ferrell, and B. A. Kottke. 1991. Expressed hypervariable polymorphism of apolipoprotein (a). *Am. J. Hum. Genet.* 49:1063-1074.
6. Lackner, C., J. C. Cohen, and H. H. Hobbs. 1993. Molecular definition of the extreme size polymorphism in apolipoprotein(a). *Hum. Mol. Genet.* 2:933-940.
7. Kraft, H. G., S. Köchl, H. J. Menzel, C. Sandholzer, and G. Utermann. 1992. The apolipoprotein(a) gene. A transcribed hypervariable locus controlling plasma lipoprotein(a) concentration. *Hum. Genet.* 90:220-230.
8. Koschinsky, M. L., U. Beisiegel, D. Hennebruns, D. L. Eaton, and R. M. Lawn. 1990. Apolipoprotein(a) size heterogeneity is related to variable number of repeat sequences in its messenger RNA. *Biochemistry.* 29:640-644.
9. McLean, J. W., J. E. Tomlinson, W. Kuang, D. L. Eaton, E. Y. Chen, G. M. Fless, A. M. Scanu, and R. M. Lawn. 1987. cDNA sequence of human apolipoprotein(a) is homologous to plasminogen. *Nature (Lond.)*. 300:132-137.
10. Scanu, A. M., and G. M. Fless. 1990. Lipoprotein(a). Heterogeneity and biological relevance. *J. Clin. Invest.* 85:1709-1715.
11. Sandholzer, C., N. Saha, J. D. Kark, A. Rees, W. Jaross, H. Dieplinger, F. Hoppichler, E. Boerwinkle, and G. Utermann. 1992. Apo(a) isoforms predict risk for coronary heart disease. A study in 6 populations. *Arterioscler. Thromb.* 12:1214-1226.
12. Loscalzo, J. 1990. Lipoprotein(a). A unique risk factor for atherothrombotic disease. *Arteriosclerosis.* 10:672-679.
13. Harpel, P., B. Gordon, and T. Parker. 1989. Plasmin catalyzes binding of lipoprotein(a) to immobilized fibrinogen and fibrin. *Proc. Natl. Acad. Sci. USA* 86:3847-3851.
14. Grainger, D. J., P. R. Kemp, A. C. Liu, R. M. Lawn, and J. C. Metcalfe. 1994. Activation of transforming growth factor- β is inhibited in transgenic apolipoprotein(a) mice. *Nature (Lond.)*. 370:460-462.
15. Grainger, D. J., H. L. Kirschenlohr, J. C. Metcalfe, P. L. Weissberg, D. P. Wade, and R. M. Lawn. 1993. Proliferation of human smooth muscle cells promoted by lipoprotein(a). *Science (Wash. DC)*. 260:1655-1658.
16. Ross, R. 1993. The pathogenesis of atherosclerosis. A perspective for the 1990s. *Nature (Lond.)*. 362:801-809.
17. Boerwinkle, E., C. C. Leffert, J. Lin, C. Lackner, G. Chiesa, and H. H. Hobbs. 1992. Apolipoprotein(a) gene accounts for greater than 90% of the variation in plasma lipoprotein(a) concentrations. *J. Clin. Invest.* 90:52-60.
18. Boerwinkle, E., H. J. Menzel, H. G. Kraft, and G. Utermann. 1989. Genetics of the quantitative Lp(a) lipoprotein trait. III. Contribution of Lp(a) glycoprotein phenotypes to normal lipid values. *Hum. Genet.* 82:73-78.
19. Parra, H. J., I. Luyéyé, C. Bouramoué, C. Demarquilly, and J. C. Fruchart. 1987. Black-White differences in serum Lp(a) lipoprotein levels. *Clin. Chim. Acta.* 167:27-31.
20. Margaretti, N., F. Acquati, P. Magnaghi, L. Bruno, M. Pontoglio, M. Rocchi, S. Saccone, G. Dellavalle, M. Durso, D. LePaslier, S. Ottolenghi, and R. Taramelli. 1992. Characterization by yeast artificial chromosome cloning of the linked apolipoprotein(a) and plasminogen genes and identification of the apolipoprotein(a) 5' flanking region. *Proc. Natl. Acad. Sci. USA*. 89:11584-11588.
21. Wade, D. P., J. G. Clarke, G. E. Lindahl, A. C. Liu, B. R. Zysow, K. Meer, K. Schwartz, and R. M. Lawn. 1993. Genetic influences on lipoprotein(a) concentration. *Biochem. Soc. Trans.* 21:499-502.
22. Kraft, H. G., N. Margaretti, S. Köchl, F. Acquati, G. Utermann, and R. Taramelli. 1993. Demonstration of physical linkage between the promoter region and the polymorphic kringle-IV domain in the apo(a) gene by pulsed-field gel electrophoresis. *Genomics.* 17:260-262.
23. Sanger, F., S. Nicklen, and A. R. Coulson. 1995. DNA sequencing with chain-terminating inhibitors. *Proc Natl. Acad. Sci. USA*. 74:5463-5467.
24. Menzel, H. J., H. Dieplinger, K. Lackner, F. Hoppichler, J. K. Lloyd, D. P. R. Muller, C. Labeur, P. J. Talmud, and G. Utermann. 1990. Abetalipoproteinemia with an apoB-100-lipoprotein(a) glycoprotein complex in plasma: indication for an assembly defect. *J. Biol. Chem.* 265:981-986.
25. Guo, W. S. and E. A. Thompson. 1989. Monte-Carlo estimation of p values for sparse contingency tables. Technical report#187. Department of Statistics, University of Washington, Seattle, WA.
26. Budowle, B., A. M. Giusti, J. S. Wayne, F. S. Baechtel, R. M. Fourney, D. E. Adams, L. A. Presley, H. A. Deadman, and K. L. Monson. 1991. Fixed-bin analysis for statistical evaluation of continuous distributions of allelic data from VNTR loci, for use in forensic comparisons. *Am. J. Hum. Genet.* 48:841-855.
27. Neter, J., W. Wassermann, and M. H. Kutner. 1990. Applied Linear Statistical Models. Irwin Inc., Homewood, IL.
28. Parra, H. J., D. Arveiler, A. E. Evans, J. P. Cambou, P. Amouyel, A. Bingham, D. Mcmaster, P. Schaffer, P. Dousteblazy, G. Luc, et al. 1992. A case control study of lipoprotein particles in 2 populations at contrasting risk for coronary heart disease. The ECTIM study. *Arterioscler. Thromb.* 12:701-707.
29. Conover, W. J. 1980. Practical Nonparametric Statistics. John Wiley & Sons, Inc., New York.
30. Chen, H., M. Kalaitzidakis, A. C. Warren, D. Avramopoulos, and S. Antonarakis. 1993. A novel zinc finger cDNA with a polymorphic pentanucleotide repeat (ATTTT)_n maps on human chromosome 19p. *Genomics.* 15:621-625.
31. Spritz, A. R. 1981. Duplication/deletion polymorphism 5' to the human β globin gene. *Nucleic Acids Res.* 5037-5047.
32. Schlötterer, C., and D. Tautz. 1992. Slipping synthesis of simple sequence DNA. *Nucleic Acids Res.* 20:211-215.
33. Vigilant, L., M. Stoneking, H. Harpending, K. Hawkes, and A. C. Wilson. 1991. African populations and the evolution of human mitochondrial DNA. *Science (Wash. DC)*. 253:1503-1507.
34. Takahashi, N., K. Hiyama, M. Kodaira, and C. Satoh. 1991. The length polymorphism in the 5' flanking region of the human globin gene with denaturing gradient gel electrophoresis in a Japanese population. *Hum. Genet.* 87:219-220.
35. Kraft, H. G., C. Haibach, A. Lingenhel, C. Brunner, M. Trommsdorff, F. Kronenberg, H. J. Müller, and G. Utermann. 1995. Sequence polymorphism in kringleIV-37 in linkage disequilibrium with the apolipoprotein(a) size polymorphism. *Hum. Genet.* 95:275-282.
36. Kidd, K. K. 1993. Associations of disease with genetic markers: déjà vu all over again. *Am. J. Med. Genet.* 48:71-73.
37. Rosenberg, L., and K. K. Kidd. 1977. HLA and disease susceptibility. A primer. *N. Engl. J. Med.* 297:1060-1062.
38. Pieretti, M., F. Zhang, Y. H. Fu, S. T. Warren, B. A. Oostra, C. T. Caskey, and D. L. Nelson. 1991. Absence of expression of the FMR-1 gene in fragile X syndrome. *Cell* 66:817-822.
39. Verheij, C., C. E. Bakker, E. de Graff, J. Keulemans, R. Willemsen, A. J. M. H. Verkerk, H. Galjaard, A. J. J. Reuser, A. T. Hoogeveen, and B. A. Oostra. 1993. Characterization and localization of the FMR1 gene product. *Nature (Lond.)*. 363:722-724.
40. Zysow, B. R., G. E. Lindahl, D. P. Wade, B. L. Knight, and R. M. Lawn. 1995. C/T polymorphism in the 5' untranslated region of the apolipoprotein(a) gene introduces an upstream ATG and reduces in vitro translation. *Arteriosclerosis, Thrombosis, and Vascular Biology.* 15:58-64.
41. White, A. L., J. E. Hixson, D. L. Rainwater, and R. E. Lanford. 1995. Molecular basis for "null" lipoprotein(a) phenotypes and the influence of apolipoprotein(a) size on plasma lipoprotein(a) level in baboon. *J. Biol. Chem.* 269:9060-9066.
42. White, A. L., and R. E. Lanford. 1994. Cell surface assembly of lipoprotein(a) in primary cultures of baboon hepatocytes. *J. Biol. Chem.* 269:28716-28723.