Lack of association of rs3798220 with small apolipoprotein(a) isoforms and high lipoprotein(a) levels in East and Southeast Asians

Mahmoud Khalifa[a,b,1]*; Asma Noureen[a,b]*; Kathrin Ertelthalner[a]; Ahmad Reza Bandegi[b,2]; Rhena Delport[b,c]; Wance J.J Firdaus[b,3]; Finney S. Geethanjali[b,d]; Kalpana Luthra[b,e]; Orawan Makemaharn[b,f]; Richard WC Pang[b,g]; Abdel-Halim Salem[b,4]; Jun Sasaki[h]; Wulf Schiefenhoevel[i]; Arno Lingenhel[a,b]; Florian Kronenberg[a]; Gerd Utermann[b]; Konrad Schmidt[a,b,j,k †]

* contributed equally

[a] Division of Genetic Epidemiology, Department of Medical Genetics, Molecular and Clinical Pharmacology, Medical University of Innsbruck, Innsbruck, Austria.

[b] Division of Human Genetics, Department of Medical Genetics, Molecular and Clinical Pharmacology, Medical University of Innsbruck, Innsbruck, Austria.

[c] Department of Chemical Pathology, Faculty of Health Sciences, University of Pretoria, Pretoria, Republic of South Africa.

[d] Department of Clinical Biochemistry, Christian Medical College Vellore, Tamil Nadu, India.

[e] Department of Biochemistry, All India Institute of Medical Sciences, New Delhi, India.

[f] Department of Obstetrics and Gynaecology, Faculty of Medicine, Siriraj Hospital, Mahidol University, Bangkok, Thailand.

[g] Division of Clinical Biochemistry, Queen Mary Hospital, Hong Kong SAR, China.

[h] Graduate School of Pharmaceutical Medicine, International University of Health and Welfare, Fukuoka, Japan.

[i] Max-Planck-Institute for Ornithology, Andechs, Germany.

[j] Centre de Recherches Médicales de Lambaréné, Albert Schweitzer Hospital, Lambaréné, Gabon.

[k] Department for Tropical Medicine, Eberhard-Karls-University Tuebingen, Tuebingen, Germany.

[1] Molecular Biology Lab, Zoology Department, Faculty of Science, Al-Azhar University, Cairo, Egypt.

[2] Biochemistry Department, Faculty of Medicine, Semnan University of Medical Sciences, Semnan, Iran.

[3] Department Obstetrics/Gynaecology and Neurology, School of Medicine, John Hopkins University, Baltimore, Maryland, USA.

[4] Departments of Anatomy, College of Medicine and Medical Sciences, Arabian Gulf University, Manama, Kingdom of Bahrain.

[†] Address for correspondence:

Dr. Konrad Schmidt, M.D.
Medical University of Innsbruck
Division of Genetic Epidemiology
Department of Medical Genetics, Molecular and Clinical Pharmacology
Schöpfstraße 41, A-6020 Innsbruck, Austria
E-mail: konrad.schmidt@afrigene.com; konrad.schmidt@i-med.ac.at

This manuscript encloses 3 figures and 2 tables

Abbreviations

apo(a)	apolipoprotein(a)
CAD	coronary artery disease
CNV	copy number variation
GWAS	genome-wide association study
KIV	Kringle IV
KIV-2	Kringle IV type 2
LD	linkage disequilibrium
LPA	the apoliprotein(a) gene (MIM +152200)
Lp(a)	lipoprotein(a)
MAF	minor allele frequency
PFGE	pulsed-field gel electrophoresis
PLG	plasminogen
QTL	quantitative trait locus
1000G	1000 Genomes Project

Abstract

Objective: The variant allele of rs3798220 in the apolipoprotein(a) gene (*LPA*) is used to assess the risk for coronary artery disease (CAD) in Europeans, where it is associated with short alleles of the Kringle IV-2 (KIV-2) copy number variation (CNV) and high lipoprotein(a) (Lp(a)) concentrations. No association of rs3798220 with CAD was detected in a GWAS of East Asians. Our study investigated the association of rs3798220 with Lp(a) concentrations and KIV-2 CNV size in non-European populations to explain the missing association of the variant with CAD in Asians.

Methods: We screened three populations from Africa and seven from Asia by TaqMan Assay for rs3798220 and determined KIV-2 CNV sizes of *LPA* alleles by pulsed-field gel electrophoresis (PFGE). Additionally, CAD cases from India were analysed. To investigate the phylogenetic origin of rs3798220, 40 *LPA* alleles from Chinese individuals were separated by PFGE and haplotyped for further SNPs.

Results: The variant was not found in Africans. Allele frequencies in East and Southeast Asians ranged from 2.9% to 11.6%, and were very low (0.15%) in CAD cases and controls from India. The variant was neither associated with short KIV-2 CNV alleles nor elevated Lp(a) concentrations in Asians.

Conclusion: Our study shows that rs3798220 is no marker for short KIV-2 CNV alleles and high Lp(a) in East and Southeast Asians, although the haplotype background is shared with Europeans. It appears unlikely that this SNP confers atherogenic potential on its own. Furthermore, this SNP does not explain Lp(a) attributed risk for CAD in Asian Indians.

Keywords: copy number variation, coronary artery disease, genetic variation, lipoprotein(a), population genetics

Introduction

The human lipoprotein(a) locus (*LPA*; MIM +152200) has been identified as one of the strongest genetic risk factors for coronary heart disease (for review see [1]).

LPA consists of 10 plasminogen Kringle IV like domains (named KIV-1 to KIV-10), a Kringle V domain (KV), and a protease-like domain [2] (Figure 1). Only the number of KIV-2 copies varies (range from 1 to >40) [3], resulting in size differences between individual alleles of 5.5kb per KIV-2 copy [4,5]. As a consequence of the transcribed and translated KIV-2 copy number variation (KIV-2 CNV), the apolipoprotein(a) protein (apo(a)) displays a size polymorphism [6]. A pentanucleotide repeat polymorphism (5'PNRP) 1372bp upstream of *LPA* is frequently analysed for more detailed discrimination between *LPA* alleles but has no effect on gene expression [7,8].

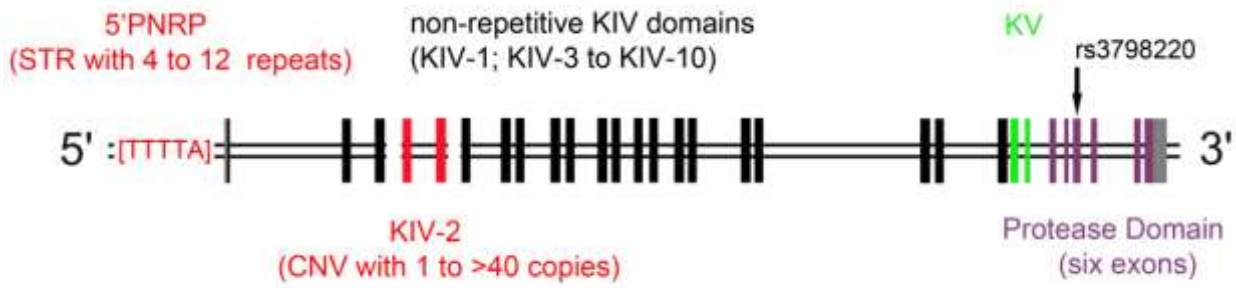


Figure 1: *LPA* gene structure

rs3798220 is located in the third exon of the protease-like domain, 72kb downstream of the KIV-2 CNV. 5'PNRP: Pentanucleotide Repeat Polymorphism, a short tandem repeat (STR) of "TTTTA" 1372bp upstream of the start codon; KIV-1, KIV-2 etc.: Kringle IV type 1, type 2 etc., all unique except KIV-2, which displays a copy number variation (CNV); KV: Kringle V.

Lp(a) consist of an LDL-like particle attached to apolipoprotein(a). *LPA* is the major quantitative trait locus controlling the lipoprotein(a) (Lp(a)) plasma concentrations [9], which are highly heritable in all populations studied so far [1], including Asians [10,11]. Individual

Lp(a) levels range from <0.1mg/dl to >200mg/dl. The KIV-2 CNV size is inversely correlated with Lp(a) concentrations in a causal manner, as longer apo(a) isoforms degrade at a higher rate in the hepatocytes prior to secretion [12,13]. However, even same-sized apo(a) isoforms vary in their associated Lp(a) levels between individuals [14], and, on average, between populations [15,16]. Several SNPs at the *LPA* locus have been identified which are associated with such individual variability of Lp(a) concentrations. For some, this has been related to a causal mechanism [17-19].

In individuals of European descent, high Lp(a) plasma concentrations and short apo(a) isoforms have been established as a risk factor for atherothrombotic disease [20]. In these populations the variant allele of rs3798220 is associated with high Lp(a) and short KIV-2 CNV alleles, and linked to an about 70% increased risk for coronary artery disease (CAD) [21-23]. An allele frequency of 13% was reported in German CAD patients undergoing lipid apheresis for very high Lp(a) [24]. However, minor allele frequencies (MAFs) of this SNP range from only approximately 1% to 2% in the general European population. Therefore, genotyping for the variant allele as a surrogate for the laborious laboratory assessment of apo(a) isoform size by pulsed-field gel [4,5] or SDS gel electrophoresis [25] captures only a small proportion of the individuals with small apo(a) isoforms and increased risk [26]. Nevertheless, rs3798220 has found commercial application in genetic risk evaluation for CAD [27].

Since rs3798220 induces an amino acid change from Isoleucine to Methionine (I4399M) in the protease-like domain of apo(a), it has been suggested that its pathophysiological significance extends beyond its role as a marker for short apo(a) isoforms but rather confers risk on its own [21,28].

The variant allele of rs3798220 is found at much higher frequencies in indigenous Asian populations, with MAFs of up to 10.5% in Southern Han Chinese. The highest frequencies are recorded for individuals in the Americas, with up to 35.3% in Peruvians. For individuals of African descent, the variant was only reported for African Americans (MAF 4.9%), while in indigenous Africans it was not found (all data from the 1000 Genomes Project (1000G)).

The marked differences in the allele frequencies between populations of Asian and European descent raise the question whether rs3798220 defines *LPA* risk alleles in Asian populations. The only genome-wide association study (GWAS) which included direct analysis of rs3798220 in East Asians did not find an association with CAD in Japanese [29]. As the strong linkage disequilibrium (LD) between rs3798220 and the KIV-2 CNV in Europeans could indicate that this haplotype is evolutionary rather young, the mutation possibly has occurred twice after the out of Africa-migration of modern human, once in Asian populations and once in Europeans. Consequently, the *LPA* haplotype of rs3798220 variant alleles might differ between populations, thus affecting this SNP's potential to tag alleles causally associated with Lp(a) and/or *LPA* conferred risk for CAD. Our study aimed at elucidating the association of rs3798220 with KIV-2 CNV allele size and Lp(a) concentrations in non-European populations and the possible origin of the mutation.

Additionally, we screened for any association of rs3798220 with coronary heart disease in two Asian Indian populations.

Material and Methods

Subjects

Samples came from seven Asian and three African populations, which were all sampled for previous *LPA* studies [16,30-32]. The African samples consisted of Khoi San of Angolan descent (n=263), sampled in Schmidtsdrift, Republic of South Africa, Gabonese Bantu from the district of Moyen-Ogooué (n=113), Gabon; and Egyptians from Ismailia (n=132), Egypt. Asian individuals were recruited in New Dehli (Asian Indians, North, n=122) and Tamil Nadu (Asian Indians, South, n=347), India; Bangkok (Thai, n=155), Thailand; Hong Kong (Hong Kong Chinese, n=198), China; Fukuoka (Japanese, n=200), Japan; Jakarta (Indonesians, n=188), Indonesia; and the Trobriand (Kiriwina) Islands (Trobriand Islanders, n=126). For additional comparison of SNP haplotypes between populations, samples from five Austrian individuals that had been previously identified as heterozygous for rs3798220 were also included in the study.

In addition to these individuals with no documented history of coronary artery disease, CAD cases from North India (n=122) and cases from South India with a history of CAD (n=97) were analysed. For these samples, the association of Lp(a) plasma concentration and the KIV-2 CNV size with CAD had been investigated previously, with the other Indians included in our study serving as controls [31].

Ethics statement

All samples were completely anonymized prior to inclusion in the analyses. Informed consent was obtained from all participants during the sampling process, in writing where possible. In

populations with a high percentage of illiterate participants, the study objectives and procedures were explained in their respective languages in understandable terms and the participants then gave verbal consent. Consent was documented by the investigator. The consent procedures and the study design were approved by the ethics committee of Medical University of Innsbruck, Austria (reference UN4485, 305/4.1). The analyses were performed in compliance with the Austrian Gene Technology Act, the federal law regulating genetic research in Austria.

LPA KIV-2 copy number genotyping

DNA was extracted from whole blood and stored in agarose plugs [4]. Data on KIV-2 CNV size had been previously assessed by pulsed-field gel electrophoresis (PFGE) after digestion of genomic DNA with the endonuclease KpnI, followed by Southern blotting using a KIV-2 specific probe [4]. We ran comparative PFGE gels covering samples from all populations over their complete size range of KIV-2 CNV alleles and ensured correct and consistent size evaluation by comparison with a molecular size ladder (λ ladder PFG marker; NEB) and standards for which the KIV-2 CNV size had been assessed previously by fibre-FISH [33]. The KIV copy numbers that are given in all tables and figures correspond to the total KIV number (hence the KIV CNV size) of alleles in order to be consistent with the nomenclature of previous publications. The number of KIV-2 repeats of any *LPA* allele can be calculated by subtracting nine repeats from the total KIV repeat number because the numbers of KIV-1 and KIV-3 to KIV-10 are invariable [3].

In the Trobriands Islanders, only a randomly chosen subset consisting of 28 individuals, and for the Indonesians, only a subset of samples including all variant carriers of rs3798220 had

been analysed for the KIV CNV size. For the individuals from Austria, DNA was not available in sufficient amount and quality to allow genotyping for the CNV size by PFGE.

Genotyping for rs3798220

Samples were typed for rs3798220 with the TaqMan® SNP genotyping assay C_25930271_10 from Applied Biosystems on a 7900HT Fast Real Time PCR System (Applied Biosystems) according to the manufacturer's protocol. Samples identified as homozygous C/C or T/T respectively heterozygous C/T in resequencing of that region (see below) served as standards on all plates. Two samples from Hong Kong could not be genotyped and were excluded from further analysis. The five individuals from Austria had been previously typed as heterozygous for rs3798220, which was confirmed by our PCR (see Supplemental Table S1, Supplemental Table S2).

Direct SNP haplotyping of *LPA* alleles

We have used a new PFGE protocol to experimentally haplotype complete *LPA* alleles profiting from the size difference conferred by the variable number of 5.5kb long KIV-2 repeats in individuals heterozygous for the KIV-2 CNV. For this procedure, endonuclease Kpn2I digested fragments containing the complete *LPA* alleles were cut from the gels and specific regions in *LPA* were amplified by PCR and sequenced (for detailed protocol, see Supplemental protocol PFGEs). As a proof of principle, this technique was compared with haplotyping by segregation analysis (Supplemental Figure S1). The fragments obtained by this approach cover the genomic area from Chr6:160,943,053 to Chr6:161,412,333 (positions according to **GRCh37:CM000668.1**), thus enclosing the complete *LPA* and *PLG* genes (Supplemental Figure S2).

Using this technique, 40 alleles from Hong Kong Chinese heterozygous for the KIV CNV by at least three KIV-2 copies difference between their two alleles were subjected to direct haplotyping of non-repetitive exons plus flanking intron sequences in *LPA* by resequencing, and the 5'PNRP 1372bp upstream of the ATG start codon by polyacrylamide gel electrophoresis [8] (Supplemental Table S1, Supplemental Table S2). These samples had been chosen randomly from the total population sample, from which they neither deviated significantly in KIV CNV size nor Lp(a) concentrations (Kolmogrov-Smirnov test).

Lp(a) concentrations

Lp(a) plasma concentrations were measured by ELISA with a specific monoclonal antibody to apo(a) as described previously [34]. This assay is highly accurate and has been used in several previous population studies [9,16,30-32]. For the Trobriand Islanders, only DNA samples were available.

KIV CNV adjusted Lp(a) concentrations

For the Indonesians, DNA and plasma samples had been made anonymous independently at the time of sampling, hence KIV CNV and Lp(a) data could not be analysed for their correlation in this sample.

To correct for the influence of the KIV CNV size, in all populations for which Lp(a) measurements were available, expected Lp(a) levels were calculated for LN transformed Lp(a) concentrations, as these did not deviate significantly from a normal distribution in the individual populations (Kolmogorov-Smirnov test). However, for the populations combined (Thai, Chinese, Japanese), the distribution still deviated significantly from a normal

distribution after LN transformation (p=0.037). The linear regression of the LN transformed Lp(a) concentrations on the sum of both KIV CNV alleles in each sample was highly significant ($p<0.001$; $R^2=0.589$ in Thai, 0.474 in Chinese, 0.339 in Japanese, and 0.443 for the combined population). The ratio of measured to expected Lp(a) levels was then used to assess the impact of rs3798220 on Lp(a) concentrations.

Statistical analyses

Standard descriptive and analytical statistical analyses were performed with IBM SPSS statistics 20.0. Pair wise linkage between the multiallelic KIV CNV and the biallelic rs3798220 was estimated with the programme MIDAS [35], and LD between SNPs as well as haplotype structures from 1000G panels with Haploview [36]. Since neither the frequency distribution of the KIV CNV nor that of the Lp(a) plasma concentrations were normally distributed, either nonparametric test were conducted or, as in case of the KIV CNV adjusted Lp(a) concentrations, data were LN-transformed prior to calculations.

Results

rs3798220 allele frequencies

The screening for rs3798220 revealed distinct differences in the presence of the variant C allele (4399M) between populations. The variant was absent in all three African populations (for the power to detect at least one variant allele see Table 1 panel a). In Southeast and East Asians, the variant allele was detected at frequencies ranging from 2.9% in the Trobriand islanders to 11.9% in the Hong Kong Chinese. In the two South Asian populations (i.e. the control groups of the Indian CAD study), only one sample from the North Indian population

Table 1: Genotype count and allele frequencies for rs3798220 in different populations

a) General population samples

		Population									
		Khoi San	**Gabonese Bantu**	**Egyptians**	**Asian Indians**	**Asian Indians**	**Thai**	**Chinese from Hong Kong**	**Japanese**	**Trobriand Islanders**	**Indonesians**
Genotype count	TT	263	113	132	121	347	133	152	166	116	177
	CT	0	0	0	1	0	22	43	32	10	11
	CC	0	0	0	0	0	0	1	2	0	0
MAF [a]	(rs3798220 C)	0	0	0	0.004	0	0.071	0.116	0.09	0.04	0.029
HWE	*p* value				0.964		0.342	0.266	0.743	0.643	0.679
Power	for 1% MAF	>0.99	0.9	0.9		>0.999					
	for 1% MAF	All Africans combined: >0.9999									
	for 0.5% MAF	All Africans combined: >0.99									

[a]: minor allele frequency for variant allele (rs3798220 C); HWE: probability of Hardy-Weinberg Equilibrium; Power gives probability to detect at least one variant allele for assumed true MAFs

b) CAD cases from India

		Population	
		Asian Indians North	**Asian Indians South**
Genotype count	TT	122	96
	CT	0	1
	CC	0	0
MAF[a]	(rs3798220 C)	0	0.005

[a]: minor allele frequency for variant allele (rs3798220 C)

carried the variant. For all populations, the observed genotype frequencies met the expectations under Hardy–Weinberg equilibrium.

rs3798220 genotype and KIV CNV Size

The frequency distribution of the KIV CNV allele size was not significantly different between Thai, Chinese, Japanese, and Trobriand Islanders; hence we also analysed their combined population sample for the association of rs3798220 with KIV CNV alleles. Overall, the mean and median KIV CNV allele size was slightly but significantly higher in individuals with the rs3798220 C/T genotype (Table 2 panel a; Figure 2). This was also seen in most of the individual populations, where these differences were significant in the Thai and the Japanese (Table 2, panel a, Supplemental Figures S3 and S4). CNV alleles with more than 33 KIV repeats were more frequent in variant carriers than in non-carriers (Supplemental Figure S3), and five of the six alleles carried by rs3798220 C/C homozygotes were larger than 32 KIV repeats, too. Samples with short KIV CNV alleles (i.e., by convention, alleles with up to 22 KIV repeats) were infrequent in all these populations. In Thai and Chinese, the frequency of short CNV alleles was higher in individuals with the rs3798220 T/T genotype (10.5% versus 4.5% and 6.9% versus 5.7%, respectively), while in the Japanese, short alleles were more frequent in the variant carriers (5.9% versus 2.1%) (Supplemental Figure S4). In the Trobriands and the Indonesians, all variant carriers analysed for CNV size harboured only alleles with far more than 22 KIV repeats. Linkage analysis confirmed the overrepresentation of rs3798220 C on long KIV CNV alleles in Asians. However, linkage disequilibrium of rs3798220 C with short alleles (19 to 21 KIV repeats) was also significant in the Thai and Japanese, but only before Yates correction for the low haplotype frequencies (Supplemental Table S3).

Table 2:

a) KIV CNV allele frequencies by rs3798220 genotypes in different populations

Population	Number of Individuals			KIV CNV size								
				mean			median			mode		
	rs3798220 genotype			rs3798220 genotype			rs3798220 genotype			rs3798220 genotype		
	TT	CT	CC	TT	CT	CC	TT	CT	CC	TT	CT	CC
Combined population[a]	476	100	3	30.7	32.2	32.7	32[1]	33[1]	33[1]	32	35	33
Thai	133	22	0	30.0	33		31[2]	34[2]		32	35	
Chinese	152	43	1	30.9	31.6	31	32	32	31	33	31	25; 37
Japanese	166	32	2	31.0	32.5	33.5	32[3]	34[3]	33[3]	32	35	33
Trobriand Islanders	25	3	0	30.8	31.7		32.5	30.5		33	28	
Indonesians	177	11	0	n.d.	33.4		n.d.	33.5		n.d	32;34	

[a] Thai, Chinese, Japanese, Trobriand Islanders combined; [1] $p < 0.0001$; [2] $p < 0.001$; [3] $p = 0.003$ (Mann-Whitney U test); n.d. : not determined

Sizes of the Kringle IV copy number variation (KIV CNV) as determined by pulsed-field gel electrophoresis (PFGE) are given for the different rs3798220 genotypes. For the Indonesians, only samples found heterozygous for rs3798220 were evaluated for KIV CNV size.

b) Lp(a) plasma concentrations by rs3798220 genotypes in different populations

Population	Number of Individuals			Lp(a) concentration [mg/dl]						KIV CNV-adjusted Lp(a) [ratio]					
				mean			median			mean			median		
	rs3798220 genotype			rs3798220 genotype			rs3798220 genotype			rs3798220 genotype			rs3798220 genotype		
	TT	CT	CC	TT	CT	CC	TT	CT	CC	TT	CT	CC	TT	CT	CC
Thai	133	22	0	21.3	13.2		12.3[1]	5.2[1]		1.03	0.84		1.03	1.03	
Chinese	152	43	1	28.3	22.0	6.3	17.4	13.3	6.3	1.00	0.95	0.65	1.01	1.00	0.65
Japanese	166	32	2	20.7	18.7	9.1	14.5	10.7	9.1	1.00	1.02	1.03	1.08	1.11	1.03
Combined population[a]	451	97	3	23.4	18.9	8.2	15.5	10.4	6.3	1.00	0.96	0.90	1.02	1.07	0.71

[a] Thai, Chinese, Japanese combined; [1] $p < 0.009$ (Mann-Whitney U test); $p >> 0.05$ all other comparisons

Measured Lp(a) concentrations are compared between the different rs3798220 genotypes in each population. To adjust for the causal impact of the KIV CNV on Lp(a) concentrations, the ratio of measured to expected LN-transformed Lp(a) concentrations was calculated. Expected LN-transformed concentrations for each KIV CNV genotype were calculated from linear regression analysis.

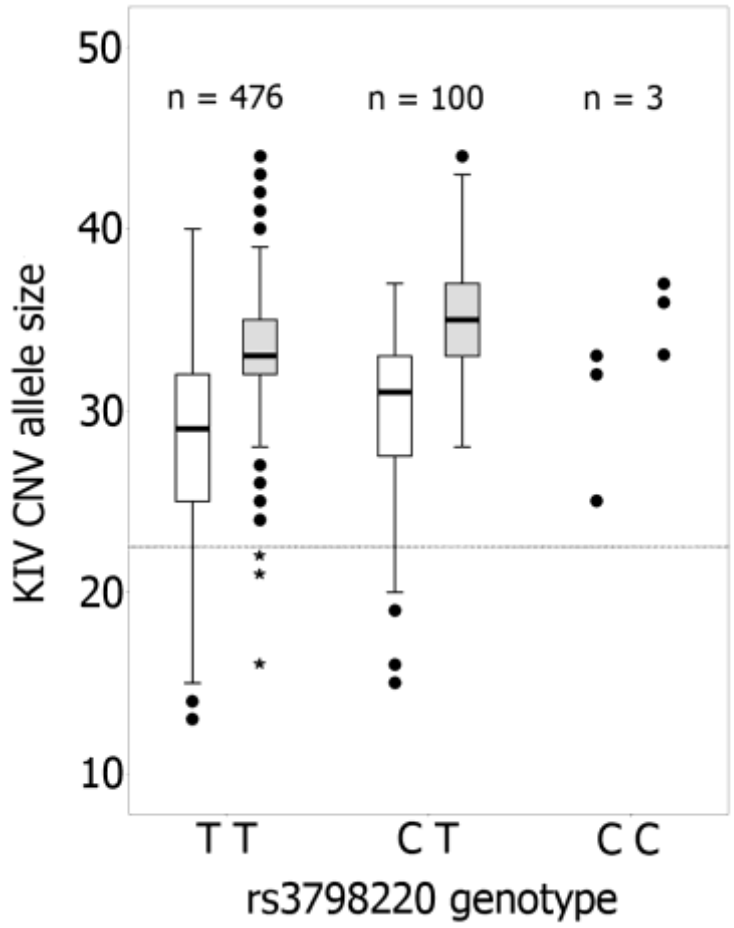


Figure 2: KIV CNV sizes in Asians depending on the rs3798220 genotype

For the combined Asian populations from Thailand, Hong Kong, Japan, and the Trobriand Islands, the size of the Kringle IV copy number variation (KIV CNV) for the shorter (white bars) and longer (grey bars) allele in each individual is shown depending on the rs3798220 genotype. The reference line gives the cut-off point for KIV CNV alleles conventionally classified as short (22 KIV repeats). Note that it is not depicted whether rs3798220 C is on the shorter or the longer CNV allele of individuals heterozygous for this SNP. Supplemental Figure S4 shows these distributions for all populations individually.

KIV CNV - *LPA* SNP haplotypes

Nine of the 40 *LPA* alleles from Hong Kong Chinese which we separated by PFGE harboured the rs3798220 C allele. Two of these variant alleles carried a KIV CNV of comparatively short size (24 respectively 25 KIV repeats, hence not short alleles by common definition), while the other seven came with KIV CNV allele sizes ranging from 29 to 41 repeats

(Supplemental Figure S5). Depending on their KIV CNV size, i.e. shorter (≤25 KIV) versus longer (≥29) alleles, the rs3798220 C carrying haplotypes also diverged in the additional sequence screened at the *LPA* locus. Two different "background haplotypes" were identified, composed of sequence variations upstream (5'PNRP, and SNPs rs1800769, rs2315129 in the promoter region of *LPA*) and downstream (indel rs3474744) of the KIV-2 CNV. For the rest of the resequenced regions, both haplotypes were indistinguishable (Supplemental Figure S5). LD between the SNPs captured in our sample further illustrates the breakup of these haplotypes (Supplemental Figure S6). The background haplotype found with the longer rs3798220 variant carrying alleles was frequent (42.5% of alleles), with two further alleles both deviating by only one additional rare SNP (rs150325648 for allele Hk17 II, and 6:g.161085594C>A (ENSG00000198670) for Hk1 I, respectively) from this background haplotype (Supplemental Figure S5). The LD pattern of our 40 haplotyped alleles matched well with that from the Asian populations of the 1000G panel (Supplemental Figure S7).

Genotyping for SNPs in five European individuals from Austria which were heterozygous for rs3798220 showed that two were homozygous for the variants of rs1800769, rs2315129, and rs3447443, while the other three were heterozygous at these positions. At the 5'PNRP, only one sample harboured an allele with 10 repeats, all others carrying 8 repeats. Consequently, their haplotype can match that of the rs3798220 C carrying alleles from the Chinese which are found with the larger KIV CNV sizes, but not that of the rarer shorter ones (Supplemental Figure S5).

Shorter core SNP haplotypes based upon calculation from 1000G data also indicate that the rs3798220 variant allele is frequently carried on the same background haplotype in all

populations. This method was limited to only 25kb of sequence downstream and 13.5kb upstream of rs3798220 (Supplemental Figure S8). Population frequencies of this core background haplotype vary considerably between continental groups, from only 7.1% in Europeans (GBR) and 40% to 50% in Asians. However, in Europeans, a haplotype which carries one additional SNP (rs6919346) in close vicinity to rs3798220 but is otherwise undistinguishable from the mentioned background haplotype constitutes another 16.5% of alleles.

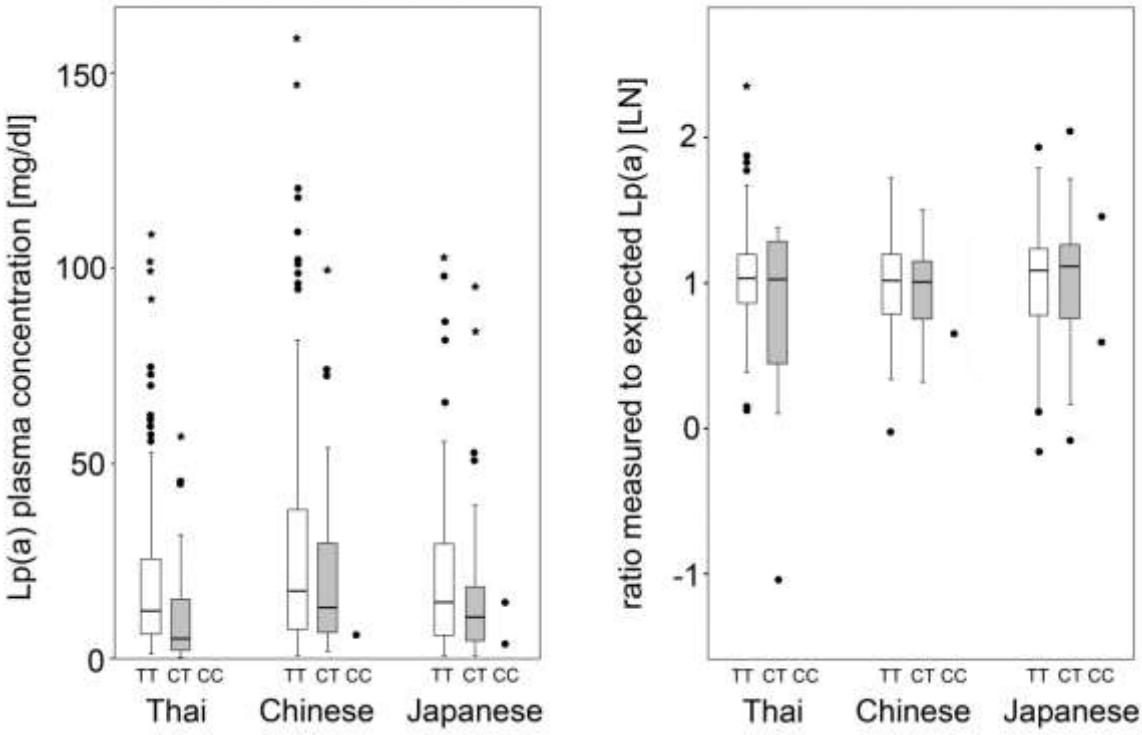


Figure 3: rs3798220 genotypes and Lp(a) concentrations

Lp(a) plasma concentrations were lower in individuals carrying the variant allele C of rs3798220 (left panel), but this difference disappeared upon correction for the effect of the size of the Kringle IV copy number variation (KIV CNV) (right panel). Lp(a) levels corrected for KIV CNV size were analysed as the ratio of measured to expected LN-transformed Lp(a) concentrations.

Association of rs3798220 with Lp(a) plasma concentrations

We investigated for the association of the rs3798220 genotype with Lp(a) plasma concentrations in our Chinese, Japanese, and Thai samples, where for all individuals screened for both the KIV CNV and rs3798220 full information on Lp(a) concentrations was available. In all three populations, Lp(a) concentrations were lower in individuals carrying the variant allele of rs3798220, but this difference was only significant in the Thai (Table 2, panel b; Figure 3). To correct for the effect of KIV CNV size on Lp(a) concentrations, we compared Lp(a) concentrations after adjusting for KIV CNV size. With this correction, no significant difference in Lp(a) concentrations was observed between rs3798220 genotypes (Figure 3).

rs3798220 and coronary artery disease in Asian Indians

In the CAD cases from North and South India, only one individual from the South Indian case group was heterozygous for C/T; all other cases carried the wild type T/T genotype (Table 1, panel b). As described above, in the respective control groups to these cases, the variant allele was also very rare (Table 1, panel a).

Discussion

Our data on the association of rs3798220 with KIV CNV sizes and Lp(a) plasma concentrations in East and Southeast Asian populations clearly show that the variant allele of this SNP is not associated with short KIV CNV alleles in these populations. Likewise, Lp(a) concentrations are not higher in carriers of the variant allele. This is in contrast to individuals of European descent, where the association of the C allele with small KIV CNV sizes and high Lp(a) concentrations was consistently demonstrated with similar numbers of variant

alleles studied [21,23,28]. In East and Southeast Asians, rs3798220 C rather seems to be associated with KIV CNV sizes larger than their population mean. This can explain the recently reported missing association of rs3798220 C with higher Lp(a) in a small sample of Americans of East Asian and Hispanic descent [37] and within Chinese CAD patients [38]. In the latter, Lp(a) concentrations were considerably lower among variant carriers. KIV CNV sizes were not measured in those studies comprising non-Europeans.

When we corrected for KIV CNV size, the association of the variant with lower Lp(a), which we could observe in Asians and which was significant in Thai, disappeared. Consequently, this association of the rs3798220 variant allele with Lp(a) can be explained by its linkage with KIV CNV size [12,13]. This constitutes a strong argument against the proposition that this variant itself has a causal effect on Lp(a) concentrations. We can, however, not rigorously rule out that such an effect was compensated for by other SNPs which rs3798220 might be in linkage disequilibrium with in Asians but not in Europeans. By the same argument, yet unidentified sequence variation in LD with both rs3798220 C and short KIV CNV alleles could be causal for the higher Lp(a) observed with this variant in Europeans. It is noteworthy that the KIV-2 region itself is still poorly analysed for sequence variations within the repetitive KIV-2 copies. Some variants within KIV-2 units appear to be in LD with KIV-2 CNV size [39].

The missing association of rs3798220 with CAD in Japanese which was reported by a GWAS [29] can be explained by this lack of association of variant carrying alleles with higher Lp(a) in East Asians. Case-control studies conducted in Chinese on the Lp(a)/apo(a) trait in cardiovascular disease consistently reported higher Lp(a) levels [40-42] and an

overrepresentation of short KIV CNV alleles [40,41] in cases compared to controls. Apparently, such high risk alleles are not tagged by rs3798220 in East Asians. The much higher population frequencies of the variant and its missing association with short KIV CNV sizes and high Lp(a) concentrations in populations of (South-) East Asian descent could explain why the predictive value of rs3798220 for CAD is considerably lower even when comparatively few non-Europeans are included in multi-ethnic studies [43]. Hence, while rs3798220 can serve as a marker for high risk *LPA* alleles in individuals of European descent [21-23, 28,44], where short KIV CNV alleles with high Lp(a) concentrations are associated with an increased risk for CAD (reviewed in [1]), testing for the presence of this variant irrespective of the ethnic background of an individual can lead to erroneous risk assessments.

The rs3798220 variant allele C is also unlikely to play a major role in the *LPA* attributed risk for CAD in Asian Indians. Its population frequency is by nearly an order of magnitude lower in these South Asians than in Europeans, and the frequencies were not different between CAD cases and controls of our study. In a previous study based on the same samples, it was shown for both North and South Indians that over the whole range of KIV CNV sizes Lp(a) concentrations were higher in CAD cases than in the respective controls [31]. Other studies also identified high Lp(a) to be associated with CAD risk in Asian Indians [42,45-47]. Given the sample size and the low minor allele frequency of rs3798220, our results do not exclude that rs3798220 variant allele frequencies are elevated in CAD cases from India. Consequently, rs3798220 C could indeed be associated with CAD risk in Asian Indians. However, with its very low allele frequency in the Asian Indians, rs3798220 C clearly does not contribute to the higher Lp(a) levels observed in the CAD cases of this study [31]. In two GWASs on CAD conducted in South Asian populations, rs3798220 was not included in the

statistical analyses due to its low population frequencies (<1% for Pakistani and for Indians alike) [48].

It has been speculated that the atherogenic potential of Lp(a) is also conferred by high oxidized phospholipids (OxPL) levels on apolipoprotein B-100 in lipoprotein(a)-particles, which were higher in carriers of small apo(a) isoforms i.e. shorter KIV CNV alleles [49] and also correlated with rs3798220 C [28]. Here the association of this variant with larger apo(a) isoforms in (South-) East Asians than in Europeans might prove important. Likewise, it has to be investigated whether the reported beneficial effect of aspirin on major cardiovascular events in variant carriers [44] might be applicable to Asian populations.

The high population frequency of the rs3798220 variant allele in the different East and Southeast Asian populations raises the question about the age and origin of the mutation. Its absence in all autochthonous African populations, including the Egyptians, is in accordance with the mutation having occurred after the initial out-of-Africa migration of modern humans. Interestingly, the geographically intermediate Asian Indians show by far lower MAFs than either Europeans or other Asians [48]. Under neutral expectations, the rather narrow range of KIV CNV sizes associated with rs3798220 C in Europeans [21,22] together with the variant's lower population frequency could be interpreted as a younger age of the mutation in Europe than in Asia, where it is far more frequent and associated with a wider CNV size range. Our haplotype analyses comprising other SNPs in *LPA* indicated that the more common background haplotype of rs3798220 variant alleles in Asians is comparatively old, as it was frequent (42.5%) and has acquired further mutations restricted to this haplotype. From our own experiments and the 1000G data of variant carriers, it appears that the rs3798220 variant

allele is mostly carried on the same background haplotype in Europeans and in Asians, despite their apparent differences in KIV-2 CNV sizes. Whether this is the case because rs3798220 is predating the split of non-African continental groups or because it has occurred as a recurrent mutation on the same old background haplotype cannot be answered with certainty. Considering the rather low frequency of this background haplotype in the European 1000G samples, the latter case appears less likely. This conclusion might be questioned as excluding rs6919346 from the haplotype definition would increase the respective background haplotype frequency in Europeans to 23.6%. However, the rs6919346 carrying haplotype seems to have diverged from the rs3798220 background haplotype earlier, as rs6919346 is also observed in autochthonous Africans and South Asians from 1000G, even on the same core haplotype across populations (Supplemental Figure S8).

In general, it is remarkable that the median KIV CNV size is far larger in East and Southeast Asians than in Europeans [15,16]. Significant bottlenecks during the peopling of East and Southeast Asia [50] constitute a likely scenario for the accumulation of the rs3798220 variant allele and longer KIV CNV alleles in the respective populations. Interestingly, the highest MAFs for rs3798220 are reported for admixed populations in the Americas. Possibly, bottleneck and drift effects during the migration out of Asia into the Americas enriched the variant further. On the other hand, since rs3798220 C is a non-synonymous SNP, introducing a substitution of isoleucine by methionine (I4399M) in the protease domain of apo(a), a functional effect of the variant resulting in positive selection might also have to be considered as a cause for such high population frequencies. So far, data on the KIV CNV size in rs3798220 variant alleles are missing for admixed Americans.

Conclusions

Our results show that in East and South East Asians, the rs3798220 variant allele is neither associated with short KIV CNV alleles nor with elevated Lp(a). This likely explains the lack of association of this variant with CAD in East Asians. Assessing the CAD risk profile by simple genotyping for rs3798220 is likely to return a high rate of false-positives in populations with a significant frequency of *LPA* alleles of Asian descent. More precise haplotype definition is needed to differentiate between rs3798220 variants carried on short KIV CNV alleles with high Lp(a) from those on long alleles without elevated Lp(a). In general, for genome-wide association studies the vastly different association of SNP haplotypes in *LPA* with KIV-2 CNV sizes found in unrelated populations has to be considered, as CNV size is causally related with Lp(a) concentrations. So far, proper CAD risk assessment of *LPA* should continue to include measurements of Lp(a) concentrations and apo(a) isoform sizes.

Acknowledgments

We are very grateful to all anonymous donors of the DNA samples used in this study

Funding sources

The study was funded by Research Project Fund P5030-011-011 "Clinical Genetics" of Medical University of Innsbruck to the Division of Human Genetics, Department of Medical Genetics, Molecular and Clinical Pharmacology.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Disclosures

The authors declare no conflicts of interest.

Reference List

[1] Kronenberg F, Utermann G. Lipoprotein(a): resurrected by genetics. *J Intern Med* 2013 January;273(1):6-30.

[2] McLean JW, Tomlinson JE, Kuang WJ, Eaton DL, Chen EY, Fless GM, Scanu AM, Lawn RM. cDNA sequence of human apolipoprotein(a) is homologous to plasminogen. *Nature* 1987 November 12;330(6144):132-7.

[3] Haibach C, Kraft HG, Kochl S, Abe A, Utermann G. The number of kringle IV repeats 3-10 is invariable in the human apo(a) gene. *Gene* 1998 February 27;208(2):253-8.

[4] Kraft HG, Kochl S, Menzel HJ, Sandholzer C, Utermann G. The apolipoprotein (a) gene: a transcribed hypervariable locus controlling plasma lipoprotein (a) concentration. *Hum Genet* 1992 November;90(3):220-30.

[5] Lackner C, Boerwinkle E, Leffert CC, Rahmig T, Hobbs HH. Molecular basis of apolipoprotein (a) isoform size heterogeneity as revealed by pulsed-field gel electrophoresis. *J Clin Invest* 1991 June;87(6):2153-61.

[6] Utermann G, Menzel HJ, Kraft HG, Duba HC, Kemmler HG, Seitz C. Lp(a) glycoprotein phenotypes. Inheritance and relation to Lp(a)-lipoprotein concentrations in plasma. *J Clin Invest* 1987 August;80(2):458-65.

[7] Mooser V, Mancini FP, Bopp S, Petho-Schramm A, Guerra R, Boerwinkle E, Muller HJ, Hobbs HH. Sequence polymorphisms in the apo(a) gene associated with specific levels of Lp(a) in plasma. Hum Mol Genet 1995 February;4(2):173-81.

[8] Trommsdorff M, Kochl S, Lingenhel A, Kronenberg F, Delport R, Vermaak H, Lemming L, Klausen IC, Faergeman O, Utermann G. A pentanucleotide repeat polymorphism in the 5' control region of the apolipoprotein(a) gene is associated with lipoprotein(a) plasma concentrations in Caucasians. *J Clin Invest* 1995 July;96(1):150-7.

[9] Boerwinkle E, Leffert CC, Lin J, Lackner C, Chiesa G, Hobbs HH. Apolipoprotein(a) gene accounts for greater than 90% of the variation in plasma lipoprotein(a) concentrations. *J Clin Invest* 1992 July;90(1):52-60.

[10] Chen TJ, Ji CY, Hu YH. Genetic and environmental influences on serum lipids and the effects of puberty: a Chinese twin study. *Acta Paediatr* 2009 June;98(6):1029-36.

[11] Rao VS, Khadrinarasimhaih NB, Kanjilal S, Mukerjee M, Kakkar VV. Genetic contribution to variation in atherothrombotic phenotypes in the Asian Indian population. The Indian Atherosclerosis Research Study. *Thromb Haemost* 2009 August;102(2):379-88.

[12] White AL, Hixson JE, Rainwater DL, Lanford RE. Molecular basis for "null" lipoprotein(a) phenotypes and the influence of apolipoprotein(a) size on plasma lipoprotein(a) level in the baboon. *J Biol Chem* 1994 March 25;269(12):9060-6.

[13] Brunner C, Lobentanz EM, Petho-Schramm A, Ernst A, Kang C, Dieplinger H, Muller HJ, Utermann G. The number of identical kringle IV repeats in apolipoprotein(a) affects its processing and secretion by HepG2 cells. *J Biol Chem* 1996 December 13;271(50):32403-10.

[14] Perombelon YF, Soutar AK, Knight BL. Variation in lipoprotein(a) concentration associated with different apolipoprotein(a) alleles. *J Clin Invest* 1994 April;93(4):1481-92.

[15] Gaw A, Boerwinkle E, Cohen JC, Hobbs HH. 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* 1994 June;93(6):2526-34.

[16] Kraft HG, Lingenhel A, Pang RW, Delport R, Trommsdorff M, Vermaak H, Janus ED, Utermann G. Frequency distributions of apolipoprotein(a) kringle IV repeat alleles and their effects on lipoprotein(a) levels in Caucasian, Asian, and African populations: the distribution of null alleles is non-random. *Eur J Hum Genet* 1996;4(2):74-87.

[17] Zysow BR, Lindahl GE, Wade DP, Knight BL, Lawn RM. C/T polymorphism in the 5' untranslated region of the apolipoprotein(a) gene introduces an upstream ATG and reduces in vitro translation. *Arterioscler Thromb Vasc Biol* 1995 January;15(1):58-64.

[18] Ogorelkova M, Gruber A, Utermann G. Molecular basis of congenital lp(a) deficiency: a frequent apo(a) 'null' mutation in caucasians. *Hum Mol Genet* 1999 October;8(11):2087-96.

[19] Parson W, Kraft HG, Niederstatter H, Lingenhel AW, Kochl S, Fresser F, Utermann G. A common nonsense mutation in the repetitive Kringle IV-2 domain of human apolipoprotein(a) results in a truncated protein and low plasma Lp(a). *Hum Mutat* 2004 December;24(6):474-80.

[20] Erqou S, Thompson A, Di Angelantonio E, Saleheen D, Kaptoge S, Marcovina S, Danesh J. Apolipoprotein(a) isoforms and the risk of vascular disease: systematic review of 40 studies involving 58,000 participants. *J Am Coll Cardiol* 2010 May 11;55(19):2160-7.

[21] Luke MM, Kane JP, Liu DM, Rowland CM, Shiffman D, Cassano J, Catanese JJ, Pullinger CR, Leong DU, Arellano AR, Tong CH, Movsesyan I, Naya-Vigne J, Noordhof C, Feric NT, Malloy MJ, Topol EJ, Koschinsky ML, Devlin JJ, Ellis SG. A Polymorphism in the Protease-Like Domain of Apolipoprotein(a) Is Associated With Severe Coronary Artery Disease. *Arterioscler Thromb Vasc Biol* 2007 June 14.

[22] Shiffman D, O'Meara ES, Bare LA, Rowland CM, Louie JZ, Arellano AR, Lumley T, Rice K, Iakoubova O, Luke MM, Young BA, Malloy MJ, Kane JP, Ellis SG, Tracy RP, Devlin JJ, Psaty BM. Association of gene variants with incident myocardial infarction in the Cardiovascular Health Study. *Arterioscler Thromb Vasc Biol* 2008 January;28(1):173-9.

[23] Clarke R, Peden JF, Hopewell JC, Kyriakou T, Goel A, Heath SC, Parish S, Barlera S, Franzosi MG, Rust S, Bennett D, Silveira A, Malarstig A, Green FR, Lathrop M, Gigante B, Leander K, de Faire U, Seedorf U, Hamsten A, Collins R, Watkins H, Farrall M. Genetic variants associated with Lp(a) lipoprotein level and coronary disease. *N Engl J Med* 2009 December 24;361(26):2518-28.

[24] Leebmann J, Roeseler E, Julius U, Heigl F, Spitthoever R, Heutling D, Breitenberger P, Maerz W, Lehmacher W, Heibges A, Klingel R. Lipoprotein Apheresis in Patients With Maximally Tolerated Lipid-Lowering Therapy, Lipoprotein(a)-Hyperlipoproteinemia, and Progressive Cardiovascular Disease: Prospective Observational Multicenter Study. *Circulation* 2013 December 17;128(24):2567-76.

[25] Kraft HG, Lingenhel A, Bader G, Kostner GM, Utermann G. The relative electrophoretic mobility of apo(a) isoforms depends on the gel system: proposal of a nomenclature for apo(a) phenotypes. *Atherosclerosis* 1996 August 23;125(1):53-61.

[26] Kronenberg F. Genetic determination of lipoprotein(a) and its association with cardiovascular disease: convenient does not always mean better. *J Intern Med* 2014 September;276(3):243-7.

[27] Shiffman D, Slawsky K, Fusfeld L, Devlin JJ, Goss TF. Cost-effectiveness model of use of genetic testing as an aid in assessing the likely benefit of aspirin therapy for primary prevention of cardiovascular disease. *Clin Ther* 2012 June;34(6):1387-94.

[28] Arai K, Luke MM, Koschinsky ML, Miller ER, Pullinger CR, Witztum JL, Kane JP, Tsimikas S. The I4399M variant of apolipoprotein(a) is associated with increased oxidized phospholipids on apolipoprotein B-100 particles. *Atherosclerosis* 2010 April;209(2):498-503.

[29] Takeuchi F, Yokota M, Yamamoto K, Nakashima E, Katsuya T, Asano H, Isono M, Nabika T, Sugiyama T, Fujioka A, Awata N, Ohnaka K, Nakatochi M, Kitajima H, Rakugi H, Nakamura J, Ohkubo T, Imai Y, Shimamoto K, Yamori Y, Yamaguchi S, Kobayashi S, Takayanagi R, Ogihara T, Kato N. Genome-wide association study of coronary artery disease in the Japanese. *Eur J Hum Genet* 2012 March;20(3):333-40.

[30] Scholz M, Kraft HG, Lingenhel A, Delport R, Vorster EH, Bickeboller H, Utermann G. Genetic control of lipoprotein(a) concentrations is different in Africans and Caucasians. *Eur J Hum Genet* 1999 February;7(2):169-78.

[31] Geethanjali FS, Luthra K, Lingenhel A, Kanagasaba-Pathy AS, Jacob J, Srivastava LM, Vasisht S, Kraft HG, Utermann G. Analysis of the apo(a) size polymorphism in Asian Indian populations: association with Lp(a) concentration and coronary heart disease. *Atherosclerosis* 2003 July;169(1):121-30.

[32] Schmidt K, Kraft HG, Parson W, Utermann G. Genetics of the Lp(a)/apo(a) system in an autochthonous Black African population from the Gabon. *Eur J Hum Genet* 2006 February;14(2):190-201.

[33] Erdel M, Hubalek M, Lingenhel A, Kofler K, Duba HC, Utermann G. Counting the repetitive kringle-IV repeats in the gene encoding human apolipoprotein(a) by fibre-FISH. *Nat Genet* 1999 April;21(4):357-8.

[34] Kronenberg F, Lobentanz EM, Konig P, Utermann G, Dieplinger H. Effect of sample storage on the measurement of lipoprotein[a], apolipoproteins B and A-IV, total and high density lipoprotein cholesterol and triglycerides. *J Lipid Res* 1994 July;35(7):1318-28.

[35] Gaunt TR, Rodriguez S, Zapata C, Day IN. MIDAS: software for analysis and visualisation of interallelic disequilibrium between multiallelic markers. *BMC Bioinformatics* 2006 April 27;7:227.:227.

[36] Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005 January 15;21(2):263-5.

[37] Rowland CM, Pullinger CR, Luke MM, Shiffman D, Green L, Movsesyan I, Devlin JJ, Malloy MJ, Kane JP, Undas A. Lipoprotein (a), LPA Ile4399Met, and fibrin clot properties. *Thromb Res* 2014 May;133(5):863-7.

[38] Li ZG, Li G, Zhou YL, Chen ZJ, Yang JQ, Zhang Y, Sun S, Zhong SL. Lack of association between lipoprotein(a) genetic variants and subsequent cardiovascular events in Chinese Han patients with coronary artery disease after percutaneous coronary intervention. *Lipids Health Dis* 2013 August 27;12:127. doi: 10.1186/1476-511X-12-127.:127-12.

[39] Noureen A, Fresser F, Utermann G, Schmidt K. Sequence Variation within the KIV-2 Copy Number Polymorphism of the Human LPA Gene in African, Asian, and European Populations. PLoS One 2015;10(3):e0121582.

[40] Sandholzer C, Boerwinkle E, Saha N, Tong MC, Utermann G. Apolipoprotein(a) phenotypes, Lp(a) concentration and plasma lipid levels in relation to coronary heart disease in a Chinese population: evidence for the role of the apo(a) gene in coronary heart disease. *J Clin Invest* 1992 March;89(3):1040-6.

[41] Qin S, Wang S, Li C. [Apolipoprotein (a) polymorphism in relation to coronary heart disease in Chinese Han nationality]. *Zhonghua Yi Xue Za Zhi* 1995 October;75(10):588.

[42] Banerjee D, Wong EC, Shin J, Fortmann SP, Palaniappan L. Racial and Ethnic Variation in Lipoprotein (a) Levels among Asian Indian and Chinese Patients. *J Lipids* 2011;2011:291954. doi: 10.1155/2011/291954. Epub 2011 May 23.:291954.

[43] Anderson JL, Knight S, May HT, Horne BD, Bair TL, Huntinghouse JA, Rollo JS, Muhlestein JB, Carlquist JF. Validation and quantification of genetic determinants of lipoprotein-a levels and predictive value for angiographic coronary artery disease. *Am J Cardiol* 2013 September 15;112(6):799-804.

[44] Chasman DI, Shiffman D, Zee RY, Louie JZ, Luke MM, Rowland CM, Catanese JJ, Buring JE, Devlin JJ, Ridker PM. Polymorphism in the apolipoprotein(a) gene, plasma lipoprotein(a), cardiovascular disease, and low-dose aspirin therapy. *Atherosclerosis* 2009 April;203(2):371-6.

[45] Sandholzer C, Saha N, Kark JD, Rees A, Jaross W, Dieplinger H, Hoppichler F, Boerwinkle E, Utermann G. Apo(a) isoforms predict risk for coronary heart disease. A study in six populations. *Arterioscler Thromb* 1992 October;12(10):1214-26.

[46] Hoogeveen RC, Gambhir JK, Gambhir DS, Kimball KT, Ghazzaly K, Gaubatz JW, Vaduganathan M, Rao RS, Koschinsky M, Morrisett JD. Evaluation of Lp[a] and other independent risk factors for CHD in Asian Indians and their USA counterparts. *J Lipid Res* 2001 April;42(4):631-8.

[47] Gambhir JK, Kaur H, Prabhu KM, Morrisett JD, Gambhir DS. Association between lipoprotein(a) levels, apo(a) isoforms and family history of premature CAD in young Asian Indians. *Clin Biochem* 2008 February 5.

[48] The Coronary Artery Disease (C4D) Genetics Consortium. A genome-wide association study in Europeans and South Asians identifies five new loci for coronary artery disease. *Nat Genet* 2011 March 6;43(4):339-44.

[49] Tsimikas S, Clopton P, Brilakis ES, Marcovina SM, Khera A, Miller ER, de Lemos JA, Witztum JL. Relationship of oxidized phospholipids on apolipoprotein B-100 particles to race/ethnicity, apolipoprotein(a) isoform size, and cardiovascular risk factors: results from the Dallas Heart Study. *Circulation* 2009 April 7;119(13):1711-9.

[50] Keinan A, Mullikin JC, Patterson N, Reich D. Measurement of the human allele frequency spectrum demonstrates greater genetic drift in East Asians than in Europeans. Nat Genet 2007 October;39(10):1251-5.

Supplemental protocol PFGEs:

Detailed PFGE protocols

Pulsed-field gel electrophoresis (PFGE) was used to determine the size of the KIV-2 CNV, and to separate complete *LPA* alleles according to the size of their KIV-2 CNV for direct SNP haplotyping. Samples were run on PFGE after digesting the genomic DNA with KpnI for KIV-2 CNV size determination, and after digestion with Kpn2I for allele separation and SNP haplotyping.

DNA was extracted in Innsbruck from whole blood and stored in agarose plugs [1]. Data on KIV size had been previously assessed by pulsed-field gel-electrophoresis (PFGE) after digestion of genomic DNA with the endonuclease KpnI, followed by Southern blotting using a KIV-2 specific probe [1]. For this study, we ran comparative PFGE gels covering samples from all populations over their complete size range of KIV CNV alleles and ensured correct and consistent KIV size evaluation by comparison with a molecular size ladder (λ ladder PFG marker; NEB) and standards for which the CNV size had been assessed previously by fibre-FISH [2].

KIV allele size determination after KpnI digestion

For all samples, half an agarose plug, containing approximately 25µg of DNA within 100µl volume, was washed twice for 30min in 1xTE buffer. Afterwards the plugs were transferred into 1,5ml Eppendorff tubes and incubated for 2 hours at 37°C with a mix of 5µl KpnI (MBI Fermentas, 10U/µl), 20µl KpnI Buffer with BSA (MBI Fermentas) and 125µl Aqua

bidestillata. Incubation was continued for another 2 hours after addition of a further 5µl KpnI. Tubes were then shortly cooled on ice, and the plugs were then transferred on a 20cm x 14cm TAE gel (agarose concentration dependent on the PFGE programme, see below). PFGE was conducted on a Bio-Rad Chef Mapper System in 0.5xTAE with the following programmes:

For alleles size determination in samples harboring at least one allele shorter than (by the previous size determination, see Material and Methods) 22 KIV repeats, the gel concentration was 1% (2g SeaKem LE agarose (Lonza) in 200ml 0.5xTAE Buffer) an the following programme was used: run duration: 18hrs 23min; temperature: 14°C; Angle: 120°; Initial switch time: 1.73s; Final switch time: 14.92s; Gradient: 6.00; Ramping constant: 0.

Size determination for all other samples was conducted with a gel concentration of 1.25% (2.5g SeaKem LE agarose (Lonza) in 200ml 0.5x TAE Buffer) by the following programme: run duration: 25hrs; temperature: 14°C; Angle: 120°; Initial switch time: 7.86s; Final switch time: 28.46s Gradient: 6.00; Ramping constant: -1.360. 2.5g. Samples heterozygous for one allele ≤22 KIV repeats and one allele ≥35 KIV were run for size determination with both programmes.

Allele separation after Kpn2I digestion

We have used a new PFGE protocol to experimentally haplotype complete *LPA* alleles profiting from the size difference conferred by the variable number of 5.5kb long KIV-2 repeats in individuals heterozygous for the KIV-2 CNV. The following protocol was applied

for this approach: Two half plugs, each containing approximately 25µg of DNA within 100µl volume, were washed 2x for 30min in 1x TE buffer. Afterwards the plugs were transferred into 1,5ml Eppendorff tubes and incubated for 2 hours at 55°C with a mix of 5µl Kpn2I (MBI Fermentas, 10U/µl), 20µl Buffer Tango (MBI Fermentas) and 125µl Aqua bidestillata. Incubation was continued for another 2 hours after addition of a further 5µl Kpn2I. Tubes were then shortly cooled on ice, and the plugs were then transferred on a 20 cm x 14 cm 1.8% TBE gel (3.6 g SeaKem LE agarose (Lonza) in 200 ml 0.5x TBE Buffer). PFGE was conducted on a Bio-Rad Chef Mapper System in 0.5x TBE with the following programme:

Run duration: 80hrs 36min; Temperature: 14°C; Angle: 120°; Initial switch time: 1min 1.53s; Final switch time: 1min 18.41s; Gradient: 6.00; Ramping constant: -1.357

Yeast chromosome (New England Biolabs) was used as a size marker in addition to a size standard with 19 and 26 KIV previously typed by Fiber-FISH [2].

For cutting the alleles, two half plugs were digested according to the protocol given above. Each sample was applied twice on the same gel which was divided into two halves, one for cutting the alleles and one for blotting. On the part of gel for cutting the alleles, the samples were loaded in alternate wells so as to minimize the chances of cross contamination between samples while later cutting the alleles. After PFGE, the gel was cut into two; one part was used for detection of the alleles by Southern blotting with a KIV-2 intron specific probe, while the other part was stored in a sealed plastic bag at 4°C. After identifying the position of alleles on the blotting part, the corresponding positions of the alleles on the other part of the

gel were excised with a razor blade (for the same method applied to KpnI digested plugs, see [3]). Additional slices were cut above and below each main slice and used in case that the main slice did not give positive PCR results, which would have indicated that a slight shift in the position of the bands had occurred during PFGE across the gel resulting in the allele having been missed in the main slice. The haploid DNA of the separated KIV-2 CNV alleles was extracted from the gel slices by centrifugation in Ultrafree-DA Centrifugal Unit tubes (Merck Millipore) after washing them twice in 1x TE buffer for 1 hour. This DNA was stored for downstream applications at -20°C until further usage.

Supplemental Table S1: Primers for PCR and sequencing

	Primer	Sequence 5'-3'	Chromosomal positions
PCR primers			
	LPA-320kb U	GGTTGTCTCCAATTCTTGTG	161,407,340 - 161,407,359
	LPA-320kb L	GCCCTATATCACCTATCTGC	161,407,878 - 161,407,897
	PNRP_U	GAATTCATTTGCGGAAAGATTG	161,086,673 - 161,086,616
	PNRP_L	CTTCAACCGGGGTGAGAGTCTC	161,086,595 - 161,086,694
	PNP 670nt U	GGCTTGAGTTCCAGATCAGC	161,087,049 - 161,087,068
	PNP 670nt L	GCCAAGACAGGTGGATCACA	161,086,399 - 161,086,418
	P1 U	CTTATGTTGTCATTTCGGCACAG	161,085,842 - 161,085,864
	P1 L	TACATTCACAATCACACACGGTG	161,084,537 - 161,084,559
	411 U	CACGTCTGTCTGCCTGCT	161,071,668 - 161,071,651
	411 L	GTTAGCTTGACGCACACC	161,071,293 - 161,071,310
	431 U	GAGCGCACTTTGCAGTGAGAAG	161,033,009 - 161,033,030
	431 L	AAACTCCAATCCCTCTCCTCTGC	161,031,735 - 161,031,757
	432 U	CTTTTAAGTATGAGAAATGTGCAGAC	161,027,859 - 161,027,884
	432 L	CAGCATCGAAACGTGTAGGTT	161,027,468 - 161,027,488
	441 U	GTGGCTTCACATGTTTGTCTCT	161,026,280 - 161,026,301
	441 L	CTGACCCTTCCTTCACTTATGG	161,026,038 - 161,026,059
	442 U	CAAAATATACAGTGTTCCTGCATATGTA	161,022,169 - 161,022,196
	442 L	CCCACATGGCAGACCCAG	161,021,821 - 161,021,838
	451 U	ATGTCATAAATGGCTTCCTGTATT	161,020,741 - 161,020,764
	451 L	ACTTCCATTGGCCCTTCC	161,020,485 - 161,020,502
	452 U	AGTTGAGGACCATTTGTGCATA	161,016,697 - 161,016,718
	452 L	TGTGTGAGCATGGAAGGCTT	161,016,291 - 161,016,310
	461 U	TCTCAGAGACAATGATCCTGTT	161,015,224 - 161,015,245
	461 L	ATGTAAGGGGGCTGCTGT	161,014,943 - 161,014,960
	462 U	GGATGTCTGGTGCCCTTG	161,012,180 - 161,012,197
	462 L	CAGATTCCCGTGCAGCAT	161,011,874 - 161,011,891
	471 U	GCATGTCCATCTGCCTACTAATG	161,010,864 - 161,010,886
	471 L	CAGACCCTGTGCCCAAAGC	161,010,429 - 161,010,447
	472 U	CATAGAGGGTGCCAGGTGGT	161,007,713 - 161,007,732
	472 L	GCTAGGACACTAATATGATTCGTTTG	161,007,203 - 161,007,228
	481 U	TCAGATTATTCCACTTTGTGTGTC	161,006,302 - 161,006,325
	481 L	GCTTGTAGCATGGGCCAG	161,006,004 - 161,006,021
	482 U	CCTTGAATATTCTCCCATC	160,999,769 - 160,999,787
	482 L	CCAGTATATAGATGTCTAACC	160,999,521 - 160,999,541
	491 U	ATCTACTGGGTCTTTTCACATCC	160,998,458 - 160,998,480
	491 L	CTTAAAGCATGGGTCTTCTAACC	160,998,095 - 160,998,117
	492 U	GAGGACCGTTTTTGCTTTTC	160,978,740 - 160,978,759
	492 L	TCCATCTCTTTTCATCCCAAT	160,978,364 - 160,978,384

	4101 U	TGCTTCACATGTTTCTGCATACTAGTAG	160,977,246 - 160,977,273
	4101 L	GAGCCTTGTAGAATGGCACTCTAAGTA	160,976,760 - 160,976,786
	4102 U	TGAAAGGCATGCACAAATGG	160,969,853 - 160,969,872
	4102 L	AGAGAGGTATGTGCTAGACAAGGTAAG	160,969,468 - 160,969,494
	51 U	TGCCTGGCTGTCTGTAATCA	160,968,998 - 160,969,017
	51 L	TAACAGGCCACAGATAATTACC	160,968,503 - 160,968,524
	52 U	GCCTCATAACTCCTTTTGACTATG	160,966,745 - 160,966,768
	52 L	AGGGCAGGTTCAAGGTGG	160,966,304 - 160,966,321
	PD1 U	CTTGGACTCAAATGCGACTT	160,963,863 - 160,963,882
	PD1 L	TTGTGGGGCTTACATGGA	160,963,667 - 160,963,684
	PD2 U	TGTGCCTATTCTTGCCTCAG	160,962,429 - 160,962,448
	PD2 L	CTGGAAATTCTGCCTCAGG	160,961,961 - 160,961,979
	PD3 U	GAGGGACAGCGCCAAGTT	160,961,282 - 160,961,299
	PD3 L	ATGTGGCAGGGCTGGGGT	160,960,907 - 160,960,924
	PD4 U	GCCTGGTTGTAGGTACTGCCA	160,959,040 - 160,959,060
	PD4 L	GACTCAAGGAAATGGCACCG	160,958,682 - 160,958,701
	PD5 U	GAACAAGCTTTGGCCTCAG	160,953,712 - 160,953,730
	PD5 L	GGAGAGCACAAGACTTTGATCT	160,953,525 - 160,953,546
	PD6 U	TGGATAAGAGAAGGATGGCA	160,952,983 - 160,953,002
	PD6 L	TGGGATGAAAAGAGCATGAA	160,952,467 - 160,952,486
Sequencing primers			
	P1seq 311 F	CTAATCAGGAAAGATGAAGGTC	161,085,533 - 161,085,554
	P1seq 780 R	CATATACAAGATTTTGAACTGG	161,085,063 - 161,085,084
	P1seq 1625 U	GGTGATCCATCTGCCGTGG	161,086,006 - 161,086,024
	P1seq 1625 L	GATTTGGGAATTTGGTCACGAG	161,084,400 - 161,084,421
	Seq431 F1	CAGTTCTAGATTGGTAGGAATC	161,032,927 - 161,032,948
	Seq432to441 U337	ATTACCCCGATTCCAAGC	161,027,531 - 161,027,548
	451 L100	CATAATGGTAGTAGCAGTCCTG	161,020,644 - 161,020,665
	451 L 208	TAGTTTTCTGGGGTCCG	161,020,541 - 161,020,557
	452 Seq F 300	CCAGATCCAAGCACAGAG	161,016,402 - 161,016,419
	452 Seq R 187	ACCAAGGGCGAATCTCAG	161,016,515 - 161,016,532
	472 Seq F145	TCAGATGGGAGTACTGCAAC	161,007,569 - 161,007,588
	472 Seq R 431	ACTCAACATCCCAGCTCTG	161,007,302 - 161,007,320
	Seq472to481 U421	GGATGTTGAGTGCCTTGTG	161,007,294 - 161,007,312
	Seq51 Seq 303L	ATTACGCTTAGCCTCCTG	160,968,715 - 160,968,732
	Seq52 L1	AAGCACTCAGCTCAAAAGCC	160,966,372 - 160,966,391
	Seq52 L2	TTTTTCCAAATCCATATTAAGG	160,966,415 - 160,966,436

Chromosomal positions are according to **GRCh37:CM000668.1**

Supplemental Table S2: PCRs used for resequencing of exons 1, 2, and 17 to 39 of *LPA* and size determination of the 5'PNRP

	PCR					
	LPA-320kb[a]	**PNP670**	**P1**	**432**	**441**	**451**
Product length	558bp	645 to 685bp	1328bp	417bp	264bp	280bp
Total volume	25µl	25µl	25µl	25µl	25µl	25µl
Upper Primer	LPA-320kb U	PNP 670nt U	P1U	432U	441U	451U
Lower Primer	LPA-320kb L	PNP 670nt L	P1L	432L	441L	451L
Primer (100µM)	0.25µl	0.15µl	0.25µl	0.15µl	0.15µl	0.15µl
Polymerase	Firepol (5U/µl)	Qiagen Multiplex Master Mix 2x	Qiagen HotStar (5U/µl)	DyNAzyme II (2U/µl)	DyNAzyme II (2U/µl)	DyNAzyme II (2U/µl)
Taq amount	0.15µl		0.125µl	0.375µl	0.375µl	0.375µl
Buffer type	10X Firepol		10X Qiagen with MgCl2	10X DyNAzyme with $MgCl_2$	10X DyNAzyme with $MgCl_2$	10X DyNAzyme with $MgCl_2$
Buffer amount	2.5µl		2.5µl	2.5µl	2.5µl	2.5µl
Additional $MgCl_2$	2.5µl		-	-	-	-
dNTP Mix	(1,25 mM each dNTP)		(1,25 mM each dNTP)	(1,25 mM each dNTP)	(1,25 mM each dNTP)	(1,25 mM each dNTP)
dNTP mix amount	4µl		4µl	4µl	4µl	4µl
Template amount	2µl	2µl	2µl	2µl	2µl	2µl
PCR programme						
Initial temperature	3 min at 95°C	15min at 95°C	15 min at 95°C	3 min at 95°C	3 min at 95°C	3 min at 95°C
Number of cycles	40	40	40	40	40	40
Denaturation	40 sec. at 95°C	40 sec. at 95°C	1 min at 95°C	40 sec. at 95°C	30 sec. at 95°C	30 sec. at 95°C
Annealing	40 sec. at 60°C	40 sec. at 67°C	1 min at 68.5°C	40 sec. at 56°C	30 sec. at 60°C	30 sec. at 60°C
Extension	1 min at 72°C	1 min at 72°C	1.5 min at 72°C	1 min at 72°C	40 sec at 72°C	40 sec at 72°C
Final extension	10 min at 72°C	10 min at 72°C	10 min at 72°C	10 min at 72°C	10 min at 72°C	10 min at 72°C

[a] only used to ascertain fragment length of separated alleles after Kpn2I digestion, see Supplemental Figure S2.

	PCR					
	452	**461**	**471**	**472**	**482**	**491**
Product length	428bp	303bp	458bp	530bp	267bp	386bp
Total volume	25µl	25µl	25µl	25µl	25µl	25µl
Upper Primer	452U	461U	471U	472U	482U	491U
Lower Primer	452L	461L	471L	472L	482L	491L
Primer (100µM)	0.15µl	0.15µl	0.15µl	0.25µl	0.25µl	0.08µl
Polymerase	DyNAzyme II (2U/µl)	DyNAzyme II (2U/µl)	Hotfire (5U/ul)	Biotherm (5U/µl)	DyNAzyme II (2U/µl)	Biotherm (5U/µl)
Taq amount	0.375µl	0.375µl	0.15µl	0.15µl	0.375µl	0.15µl
Buffer type	10X DyNAzyme with $MgCl_2$	10X DyNAzyme with $MgCl_2$	10X Firepol	10X Biotherm with $MgCl_2$	10X DyNAzyme with $MgCl_2$	10X Biotherm with $MgCl_2$
Buffer amount	2.5µl	2.5µl	2.5µl	2.5µl	2.5µl	2.5µl
Additional $MgCl_2$	-	-	2.5µl	-	-	-
dNTP Mix	(1,25 mM each dNTP)	(1,25 mM each dNTP)	(1,25 mM each dNTP)	(1,25 mM each dNTP)	(1,25 mM each dNTP)	(1,25 mM each dNTP)
dNTP mix amount	4µl	4µl	4µl	4µl	4µl	4µl
Template amount	2µl	2µl	2µl	1µl	2µl	2µl
PCR programme						
Initial temperature	3 min at 95°C	3 min at 95°C	3 min at 95°C	3 min at 95°C	3 min at 95°C	3 min at 95°C
Number of cycles	40	40	40	40	40	40
Denaturation	30 sec. at 95°C	40 sec. at 95°C	40 sec. at 95°C	40 sec. at 95°C	40 sec. at 95°C	40 sec. at 95°C
Annealing	30 sec. at 60°C	40 sec. at 64°C	40 sec. at 67°C	40 sec. at 65.2°C	40 sec. at 57°C	40 sec. at 60°C
Extension	40 sec at 72°C	1 min at 72°C	1 min at 72°C	1 min at 72°C	1 min at 72°C	1 min at 72°C
Final extension	10 min at 72°C	10 min at 72°C	10 min at 72°C	10 min at 72°C	10 min at 72°C	10 min at 72°C

	PCR					
	492	**4101**	**51**	**52**	**PD1**	**PD2**
Product length	396bp	514bp	515bp	465 bp	216bp	488bp
Total volume	25µl	25µl	25µl	25µl	25µl	25µl
Upper Primer	492U	4101U	51U	52U	PD1U	PD2U
Lower Primer	492L	4101L	51L	52L	PD1L	PD2L
Primer (100µM)	0.15µl	0.15µl	0.25µl	0.15µl	0.15µl	0.125µl
Polymerase	Biotherm (5U/µl)	Hotfire (5U/ul)	DyNAzyme II (2U/µl)	Biotherm (5U/µl)	Biotherm (5U/µl)	Biotherm (5U/µl)
Taq amount	0.15µl	0.15µl	0.375µl	0.15µl	0.15µl	0.15µl
Buffer type	10X Biotherm with $MgCl_2$	10X Firepol	10X DyNAzyme with $MgCl_2$	10X Biotherm with $MgCl_2$	10X Biotherm with $MgCl_2$	10X Biotherm with $MgCl_2$
Buffer amount	2.5µl	2.5µl	2.5µl	2.5µl	2.5µl	2.5µl
Additional $MgCl_2$	-	2.5ul	-	-	-	-
dNTP Mix	(1,25 mM each dNTP)	(1,25 mM each dNTP)	(1,25 mM each dNTP)	(1,25 mM each dNTP)	(1,25 mM each dNTP)	(1,25 mM each dNTP)
dNTP mix amount	4µl	4µl	4µl	4µl	4µl	4µl
Template amount	2µl	2µl	2µl	2µl	2µl	1µl
PCR programme						
Initial temperature	3 min at 95°C	3 min at 95°C	3 min at 95°C	3 min at 95°C	3 min at 95°C	3 min at 95°C
Number of cycles	40	40	40	40	40	40
Denaturation	40 sec. at 95°C	40 sec. at 95°C	40 sec. at 95°C	40 sec. at 95°C	30 sec. at 95°C	40 sec. at 95°C
Annealing	40 sec. at 55°C	40 sec. at 57°C	40 sec. at 57°C	40 sec. at 62°C	30 sec. at 60°C	40 sec. at 64°C
Extension	1 min at 72°C	1 min at 72°C	40 sec at 72°C	1 min at 72°C	40 sec. at 72°C	1 min at 72°C
Final extension	10 min at 72°C	10 min at 72°C	10 min at 72°C	10 min at 72°C	10 min at 72°C	10 min at 72°C

	PCR			
	PD4	**PD5**	**PD6**	**431**[b]
Product length	379bp	206bp	536bp	1296bp
Total volume	25µl	25µl	25µl	25µl
Upper Primer	PD4U	PD5U	PD6U	431U
Lower Primer	PD4L	PD5L	PD6L	431L
Primer (100µM)	0.15µl	0.25µl	0.25µl	0.15µl
Polymerase	DyNAzyme II (2U/µl)	DyNAzyme II (2U/µl)	DyNAzyme II (2U/µl)	Qiagen HotStar (5U/µl)
Taq amount	0.375µl	0.375µl	0.375µl	0.25µl
Buffer type	10X DyNAzyme with $MgCl_2$	10X DyNAzyme with $MgCl_2$	10X DyNAzyme with $MgCl_2$	10X Qiagen with MgCl2
Buffer amount	2.5µl	2.5µl	2.5µl	2.5µl
Additional $MgCl_2$	-	-	-	-
dNTP Mix	(1,25 mM each dNTP)	(1,25 mM each dNTP)	(1,25 mM each dNTP)	(1,25 mM each dNTP)
dNTP mix amount	4µl	4µl	4µl	4µl
Template amount	2µl	1µl	2µl	2µl
PCR programme				
Initial temperature	3 min at 95°C	3 min at 95°C	3 min at 95°C	15 min at 95°C
Number of cycles	40	40	40	40
Denaturation	40 sec. at 95°C	40 sec. at 95°C	40 sec. at 95°C	40 sec. at 95°C
Annealing	40 sec. at 62°C	40 sec. at 62°C	40 sec. at 60°C	40 sec. at 64°C
Extension	1 min at 72°C	1 min at 72°C	1 min at 72°C	1.5 min at 72°C
Final extension	10 min at 72°C	10 min at 72°C	10 min at 72°C	10 min at 72°C

[b] Only used to ascertain validity of haplotyping by separated alleles after Kpn2I digestion in comparison to segregation analysis, see Supplemental Figure S1.

	PCR (Touchdown)			
	5‘PNRP	**411**	**442**	**462**
Product length	75 to 115bp	376bp	376bp	324bp
Total volume	25µl	25µl	25µl	25µl
Upper Primer	PNRP_U	411U	442U	462U
Lower Primer	PNRP_L	411L	442L	462L
Primer (10 µM)	0.25µl	0.25µl	0.08µl	0.15µl
Polymerase	Qiagen HotStar (5U/ul)	Biotherm (5U/µl)	Biotherm (5U/µl)	DyNAzyme II (2U/µl)
Taq amount	0.125µl	0.15µl	0.15µl	0.375µl
Buffer type	10X Qiagen with $MgCl_2$	10X Biotherm with $MgCl_2$	10X Biotherm with $MgCl_2$	10X DyNAzyme with $MgCl_2$
Buffer amount	2.5µl + 5ul Q-Solution	2.5µl	2.5µl	2.5µl
Additional $MgCl_2$	6ul	-	-	-
dNTP Mix	(1,25 mM each dNTP)	(1,25 mM each dNTP)	(1,25 mM each dNTP)	(1,25 mM each dNTP)
dNTP mix amount	8µl	4µl	4µl	4µl
Template amount	2µl	2µl	2µl	2µl
PCR programme				
Initial temperature	5 min at 95°C	3 min at 95°C	3 min at 95°C	3 min at 95°C
Number of cycles	10	9	9	12
Denaturation	30sec. at 95°C	40sec. at 95°C	40sec. at 95°C	40 sec. at 95°C
Annealing	40sec. at 63.8°C (-0.5°C/cycle)	40sec. at 65°C (-1°C/cycle)	40sec. at 68°C (-1°C/cycle)	40 sec. at 62°C (-1°C/cycle)
Extension	60sec. at 72°C	40sec. at 72°C	40sec. at 72°C	1 min at 72°C
Number of cycles	30	35	30	30
Denaturation	30sec. at 95°C	40sec. at 95°C	40sec. at 95°C	40 sec. at 95°C
Annealing	40sec. at 58.8°C	40sec. at 57°C	40sec. at 60°C	40 sec. at 51°C
Extension	60sec. at 72°C	40sec. at 72°C	40sec. at 72°C	1 min at 72°C
Final extension	10 min at 72°C	10 min at 72°C	10 min at 72°C	10 min at 72°C

	PCR (Touchdown)		
	PD3	**481**	**4102**
Product length	393bp	322bp	405bp
Total volume	25µl	25µl	25µl
Upper Primer	PD3U	481U	4102U
Lower Primer	PD3L	481L	4102L
Primer (100µM)	0.25µl	0.15µl	0.15µl
Polymerase	DyNAzyme II (2U/µl)	Biotherm (5U/µl)	DyNAzyme II (2U/µl)
Taq amount	0.375µl	0.15µl	0.375µl
Buffer type	10X DyNAzyme with $MgCl_2$	10X Biotherm with $MgCl_2$	10X DyNAzyme with $MgCl_2$
Buffer amount	2.5µl	2.5µl	2.5µl
Additional $MgCl_2$	-	-	-
dNTP Mix	(1,25 mM each dNTP)	(1,25 mM each dNTP)	(1,25 mM each dNTP)
dNTP mix amount	4µl	4µl	4µl
Template amount	1µl	2µl	2µl
PCR programme			
Initial temperature	3 min at 95°C	3 min at 95°C	3 min at 95°C
Number of cycles	8	9	16
Denaturation	40 sec. at 95°C	40 sec. at 95°C	40 sec. at 95°C
Annealing (touchdown)	40 sec. at 69.1°C (-0.3°C/cycle)	40 sec. at 69°C (-1°C/cycle)	40 sec. at 68°C (-0.5°C/cycle)
Extension	1 min at 72°C	40 sec. at 72°C	40 sec. at 72°C
Number of cycles	35	31	30
Denaturation	40 sec. at 95°C	40 sec. at 95°C	40 sec. at 95°C
Annealing	40 sec. at 67°C	40 sec. at 60°C	40 sec. at 60°C
Extension	1 min at 72°C	40 sec. at 72°C	40 sec at 72°C
Final extension	10 min at 72°C	10 min at 72°C	10 min at 72°C

For chromosomal positions of primer annealing sites, see Supplemental Table S1. Reaction volume of 25µl was completed by adding aqua bidestillata.

Supplemental Table S3: Linkage Disequilibrium between KIV CNV and rs3798220

Population	KIV CNV - rs3798220 haplotype	Haplotype frequency [%]	LD		p	p corrected (Yates)
			D'	r^2		
Thai (n=155)						
	20 - C	0.6	0.42	0.03	<0.05	n.s.
	35 - C	1.9	0.20	0.04	<0.05	n.s.
	37 - C	1.2	0.31	0.04	<0.05	n.s.
	38 - C	0.6	0.61	0.05	<0.01	n.s.
	43 - C	0.3	1	0.04	<0.05	n.s.
Chinese (n=196)						
	38 - C	0.7	0.48	0.02	<0.05	n.s.
	39 - C	0.9	0.36	0.02	<0.05	n.s.
Japanese (n=200)						
	19 – C	0.3	1	0.03	<0.05	n.s.
	21 – C	0.3	1	0.02	<0.05	n.s.
	35 - C	2.0	0.18	0.03	<0.05	n.s.
	37 - C	1.3	0.31	0.04	<0.001	p<0.05
	44 - C	0.3	1	0.03	<0.05	n.s.
Trobriand Islanders (n=28)						
	28 - C	5.0	1	0.58	<0.0001	p<0.05
	39 - C	1.8	1	0.32	<0.01	n.s.
Combined Asians (n=579)[a]						
	32 -C	0.3	-	0.01	<0.05	n.s.
	35 - C	1.5	0.11	0.01	<0.05	<0.05
	37 - C	0.7	0.16	0.01	<0.05	n.s.
	38 - C	0.5	0.35	0.02	<0.01	<0.05
	39 - C	0.4	0.25	0.01	<0.05	n.s.

[a] Thai, Chinese, Japanese, Trobriand Islanders combined

Pair wise linkage between the multiallelic KIV CNV and the biallelic rs3798220 was estimated with the programme MIDAS [4]. p values after Yates correction consider haplotype frequencies.

Legends to Supplemental Figures:

Supplemental Figure S1: Haplotyping in *LPA* by allele separation and segregation analysis

The accuracy of SNP haplotyping within *LPA* was compared between two methods: Allele separation of complete *LPA* alleles by pulsed-field gel electrophoresis (PFGE) using our new protocol (see Supplemental protocol PFGEs) for Kpn2I digested genomic DNA (left panel) and segregation analysis in a family from South Africa (not part of this study) where segregation of the *LPA* alleles was determined by PFGE after KpnI digestion. The SNP to be phased was rs7754522, a tri-allelic SNP located in the KIV-3 flanking intron region of *LPA*. Sample S13 was heterozygous A/T at this location (note: the bases are given as on the reverse = coding strand of *LPA*) and carried alleles with a KIV CNV size of 20 and 27 KIV repeats on the short and long alleles, respectively. This individual is the father of individuals 31c, 31e, and 31g, whose mother is individual 31a with two KIV CNV alleles of size 28 and the reference allele C of rs7754522. Both methods unambiguously allocated rs7754522 variant T on the short (S13 I) and variant A on the long (S13 II) allele of S13. The same size standard with a CNV size of 19 and 26 KIV as previously determined by fibre-FISH [2] was run on the gels for allele separation (Kpn2I digestion) and KIV CNV size assessment (KpnI digestion), as shown on the Southern Blots after detection with a KIV-2 specific probe [1].

Supplemental Figure S2: Fragment size recovered from allele separation by PFGE after Kpn2I digestion

The fragments separated by PFGE after Kpn2I digestion according to the size of the KIV-2 CNV in *LPA* (in the reference allele (ENSG00000198670) depicted here with six KIV-2 copies, 5.5kb long each) span from Chr6:160,943,053 to Chr6:161,412,333, thus enclosing

the complete *LPA* and *PLG* genes. The fragments are not cut at the Kpn2I recognition site TCCGGA at Chr6:161,139,041 within *PLG*, indicating methylation at this site. Fragment lengths vary according to the size of the KIV-2 CNV from approximately 469 kb to 680 kb, depending on the number of KIV-2 copies (1 to >40). Inactivation of the recognition site within *PLG* can be derived from the length of the fragments in comparison with a molecular size ladder (see Supplemental protocol PFGEs) and was demonstrated by the possibility to phase SNPs both at positions Chr6:161,407,819 (rs673468) and Chr6:160,952,838 (rs3124784) after specific PCRs (Supplemental Tables S1 and S2) for those regions on separated alleles. After identification of a heterozygous SNP, phasing it on the short allele is sufficient, as shown for rs3124784 in sample H10.

Supplemental Figure S3: KIV CNV allele frequencies depending on the rs3798220 genotype

The frequency distribution of KIV CNV alleles (by number of KIV repeats) is depicted for individuals depending on their rs3798220 genotype and by population. In all three Asian populations, alleles with more than 33 KIV repeats are more frequent in individuals carrying at least one LPA allele with rs3798220 C. Which of the two alleles in individuals heterozygous for rs3798220 is harbouring the variant allele cannot be derived from these distributions. The frequency distributions are significantly different between the rs3798220 genotypes in Thai and Japanese, but not in Chinese ($p<0.001$ and $p=0.301$, respectively, Kolomogorov-Smirnov test).

Supplemental Figure S4: KIV CNV genotype by rs3798220 genotype in individual Asian populations

The KIV CNV size for the shorter (left bar) and longer (right bar) allele in each individual is shown depending on his or her rs3798220 genotype. For the Indonesians, only individuals carrying rs3798220 C had been typed for the CNV size. The reference line gives the cutoff point for KIV CNV alleles conventionally classified as being small/short (22 KIV or 13 KIV-2 repeats). In Thai and Japanese, short KIV CNV alleles are more frequent in individuals that are homozygous for the reference allele of rs3798220.

Note that it is not depicted whether rs3798220 C is on the short or the long CNV allele of heterozygous individuals for this SNP.

Supplemental Figure S5: SNP haplotypes from resequencing of *LPA* domains

Haplotypes for 40 alleles from Chinese were experimentally assessed after allele separation by PFGE (upper panel). *LPA* domains resequenced comprised exons 1, 2, and 17 to 39 and flanking sequences (see Figure 1 and materials and methods). Haplotyping included the 5'PNRP and the KIV CNV. The length of the black bars correlates to the number of TTTTA repeats and KIV-2 copies, respectively. Blue squares indicate the reference allele at the respective position, yellow squares the variant allele. Roman numbers refer to the short and long allele of sample IDs.

5 Austrian were resequenced for some of these regions with the same PCRs at the genotype level (lower panel). Uncovered regions are marked in white. Red squares indicate that the sample was heterozygous at the respective position, yellow squares indicate homozygosity. The samples are not phased for KIV CNV and 5'PNRP sizes, and their KIV CNV size is based upon phenotyping by Western Blot, thus samples with only one detectable isoforms can carry a non-expressed allele or be homozygous for the CNV.

Chinese alleles carrying the rs3798220 variant allele (at the lower end of the panel) group into two haplotypes, which are distinguished by SNPs rs1800769, rs2315129, rs3474744 as well as by the 5'PNRP and the range of the KIV CNV size. In the region resequenced, the Austrian samples are either homozygous or heterozygous for the variants defining the larger group of Chinese rs3798220 variant alleles.

Supplemental Figure S6: LD between rs3798220 and other SNPs within *LPA* in phased alleles from Hong Kong Chinese

For all SNPs covered in our sample of 40 Chinese alleles haplotyped by PFGE LD was calculated by Haploview [5]. Colouring according to the "alternate D'/LOD" scheme (low D' white; high D' & high LOD: black; high D' & low LOD: shades of pink), numbers within squares give r^2.

Supplemental Figure S7: LD between rs3798220 and other SNPs within *LPA*

For all SNPs covered both in our sample of 40 Chinese alleles haplotyped by PFGE (panel A) and in individual genotypes from the 1000G project in Southern Han Chinese (CHS) (panel B), Chinese from Bejing, China (CHB) (panel C), Chinese Dai from Xishuangbanna, China

(CDX) (panel D), Japanese from Tokyo, Japan (panel E), and Kinh from Ho Chi Minh City, Vietnam (KHV) (panel F), LD was calculated by Haploview [5]. Colouring is according to the "alternate D'/LOD" scheme (low D' white; high D' & high LOD: black; high D' & low LOD: shades of pink), numbers within squares give r^2. Haplotype blocks are here defined by solid spine of LD.

Supplemental Figure S8: Haplotypes calculated from 1000G data

From the individual genotypes provided by the 1000 Genomes Project (release 16), core haplotypes around rs3798220 were calculated by Haploview [5] according to the programme's "confidence interval" method [6] for all the Asian (Chinese Dai in Xishuangbanna, China (CDX); Han Chinese in Bejing, China (CHB); Southern Han Chinese (CHS), Kinh in Ho Chi Minh City, Vietnam (KHV); Japanese in Tokyo, Japan (JPT)) populations and the only European population (British in England and Scotland (GBR)) which harboured more than one rs3798220 variant allele. While otherwise markers below a MAF of 5% were excluded for haplotype definition, rs3798220 was included though being of lower frequency in some populations.

For easier graphical interpretation, colours depicting wild type (blue) and variant bases (red) were adjusted in all populations to the definition of wild type and variant in the GBR panel. For this comparison, the core region was restricted to the region assigned to one haplotype block in the GBR panel by Haploview. Where within this region haplotype blocks were shorter in other populations (CHS, JPT), only the haplotype for the block including rs3798220 is depicted.

Blanks indicate SNPs which were not included in the haplotype output from Haploview, though these SNPs are biallelic in the respective populations (with low minor allele

frequencies). Positions at which populations are homogenous in the 1000G data, while others show variation are shown in the appropriate colours in all populations.

Panel A:

Haplotypes carrying the rs3798220 variant allele (C) are the same in all populations for the region captured. Frequencies for these haplotypes are given for each population. Below these rs3798220 C haplotypes, the rs3798220 T haplotypes carrying the same or very closely related background haplotypes are also depicted by population. While the latter are frequent in the Asian populations, in the GBR panel the exactly same background haplotype to the rs3798220 variant carrying haplotype is only seen in 4.4% of samples. However, a closely related haplotype with rs6919346 as an additional SNP reaches a frequency of 16.5% in GBR. Since this SNP is also seen in autochthonous African populations (Gambian in Western Divisions in the Gambia (GWD); Mende in Sierra Leone (MSL); Luhya in Webuye, Kenya (LWK)) and South Asian (Punjabi from Lahore, Pakistan (PJL)), and haplotype calculation by Haploview puts this SNP on the same background haplotype in GBR, GWD, and PJL, the respective haplotype can be assumed to be old and possibly predate rs3798220. The assignment of very different haplotypes to this SNP in MSL and LWK can be explained by the very low frequency of rs6919346 in these populations (one allele only), rendering calculation of haplotypes covering this rare variant notoriously untrustworthy.

Panel B:

For the Asian populations and the British from the same 1000G data, all other calculated haplotypes which are neither carrying the rs3798220 variant allele nor showing a corresponding background haplotype are depicted together with their respective population frequencies.

Reference List

[1] Kraft HG, Kochl S, Menzel HJ, Sandholzer C, Utermann G. The apolipoprotein (a) gene: a transcribed hypervariable locus controlling plasma lipoprotein (a) concentration. *Hum Genet* 1992 November;90(3):220-30.

[2] Erdel M, Hubalek M, Lingenhel A, Kofler K, Duba HC, Utermann G. Counting the repetitive kringle-IV repeats in the gene encoding human apolipoprotein(a) by fibre-FISH. *Nat Genet* 1999 April;21(4):357-8.

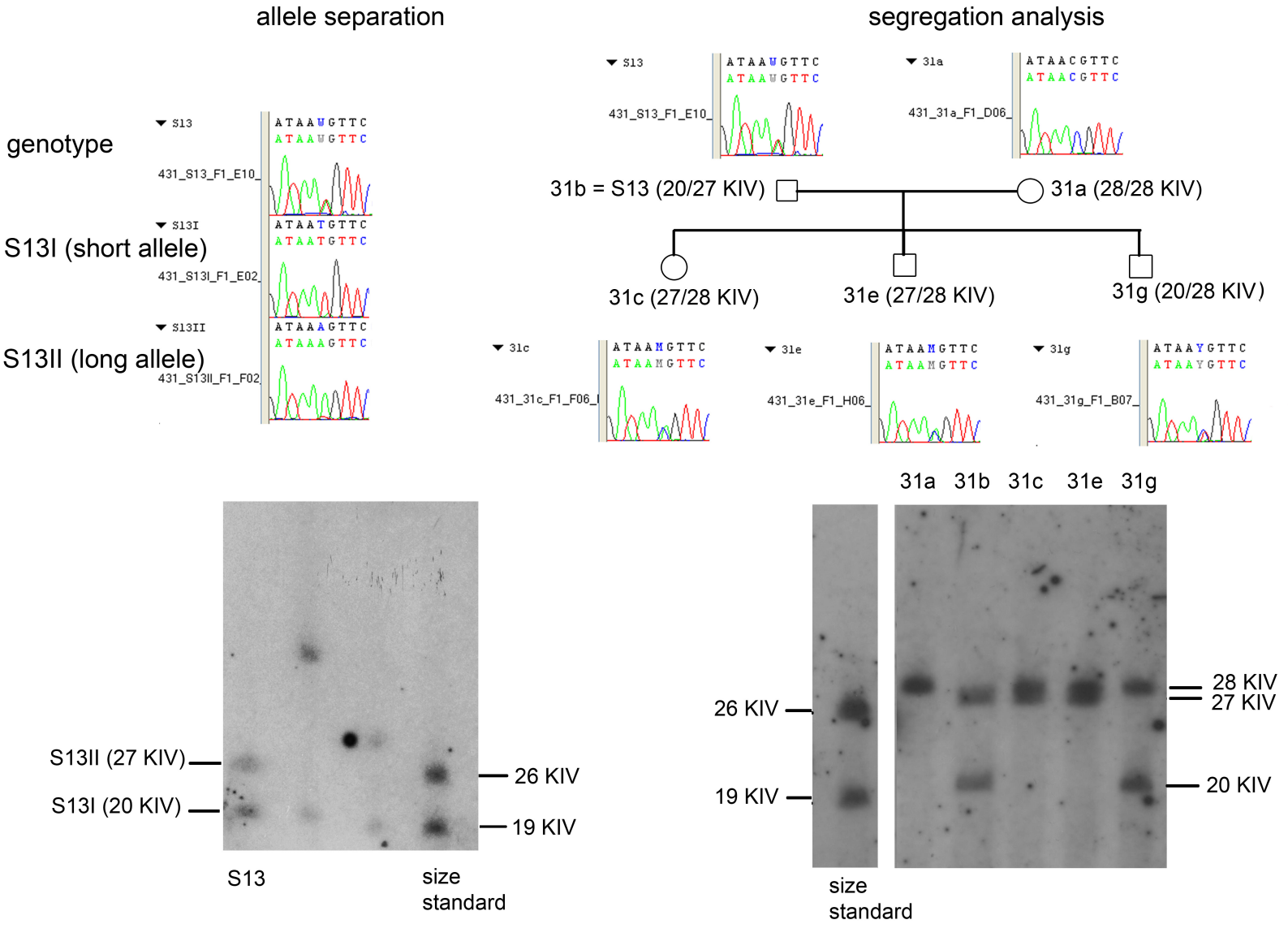
[3] Noureen A, Fresser F, Utermann G, Schmidt K. Sequence Variation within the KIV-2 Copy Number Polymorphism of the Human LPA Gene in African, Asian, and European Populations. *PLoS One* 2015;10(3):e0121582.

[4] Gaunt TR, Rodriguez S, Zapata C, Day IN. MIDAS: software for analysis and visualisation of interallelic disequilibrium between multiallelic markers. *BMC Bioinformatics* 2006 April 27;7:227.:227.

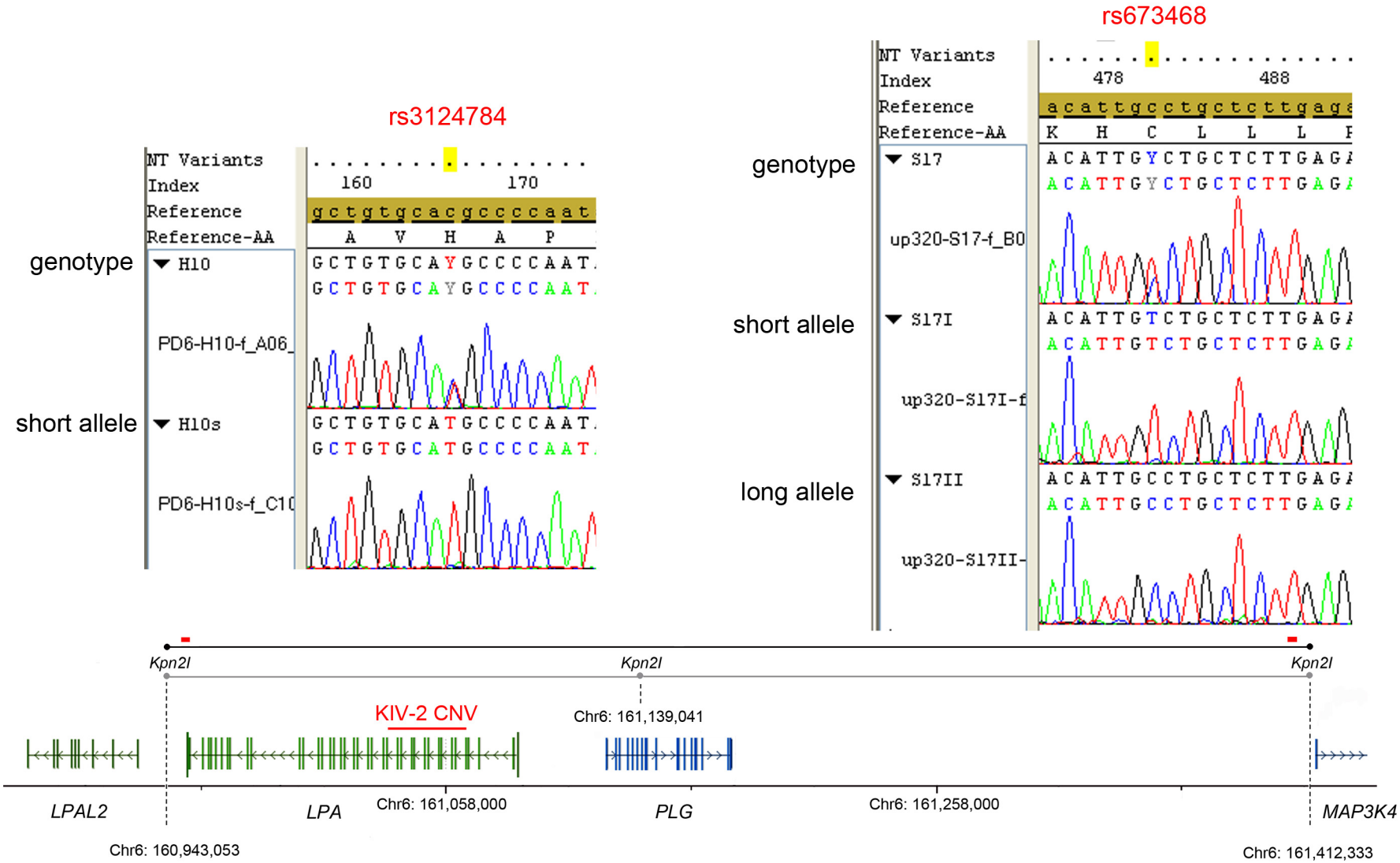
[5] Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005 January 15;21(2):263-5.

[6] Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D. The structure of haplotype blocks in the human genome. *Science* 2002 June 21;296(5576):2225-9.

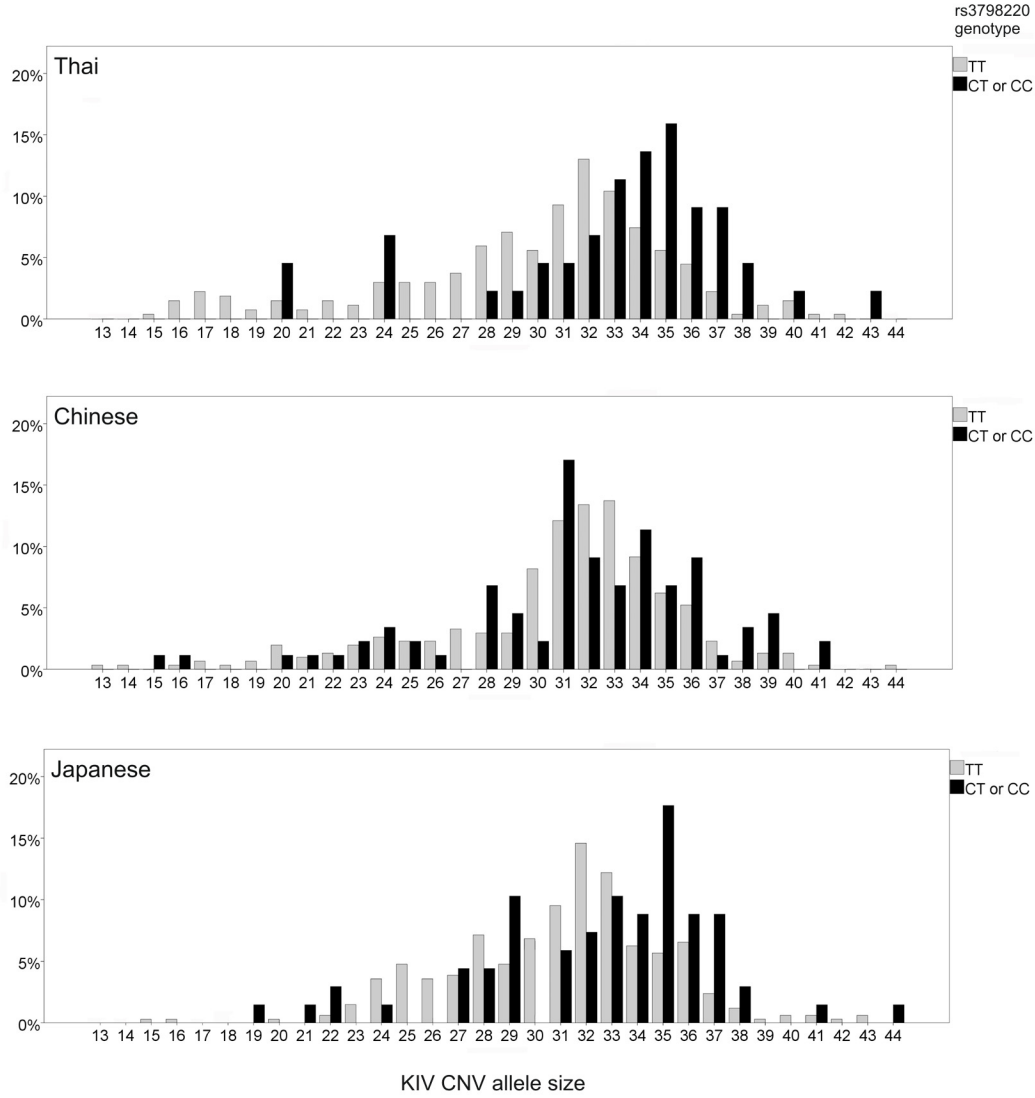
Supplemental Figure S1: Haplotyping in LPA by allele separation and segregation analysis



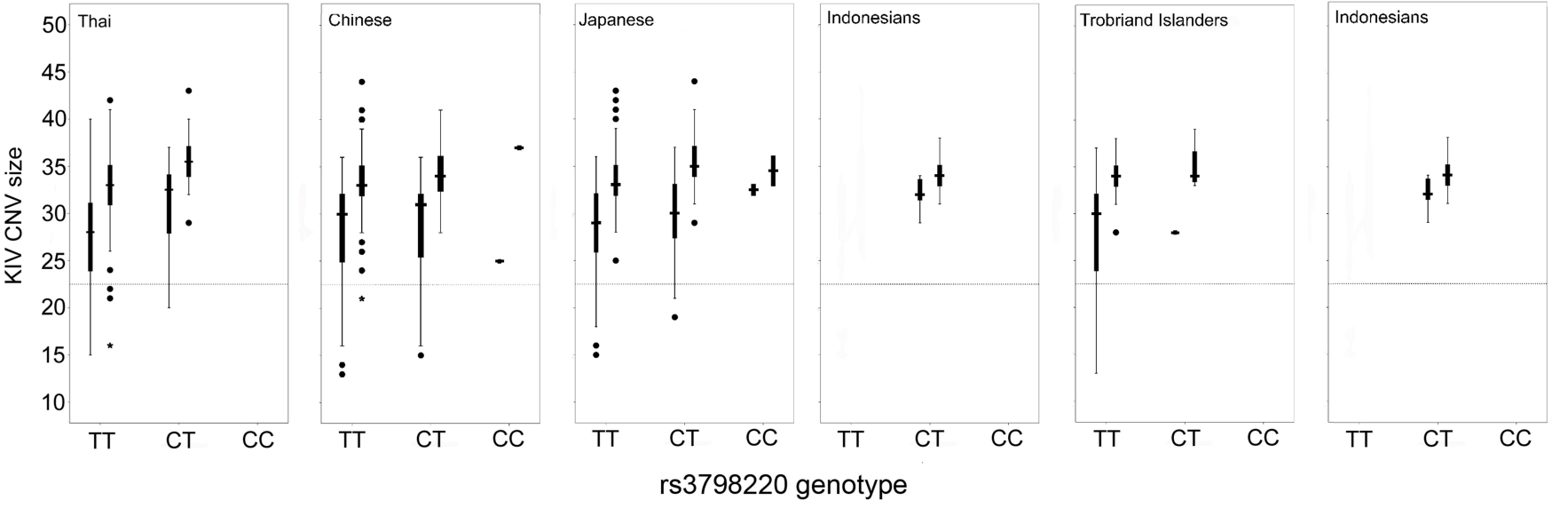
Supplemental Figure S2: Fragment size recovered from allele separation by PFGE after Kpn2I digestion



Supplemental Figure S3: KIV CNV allele frequencies depending on the rs3798220 genotype



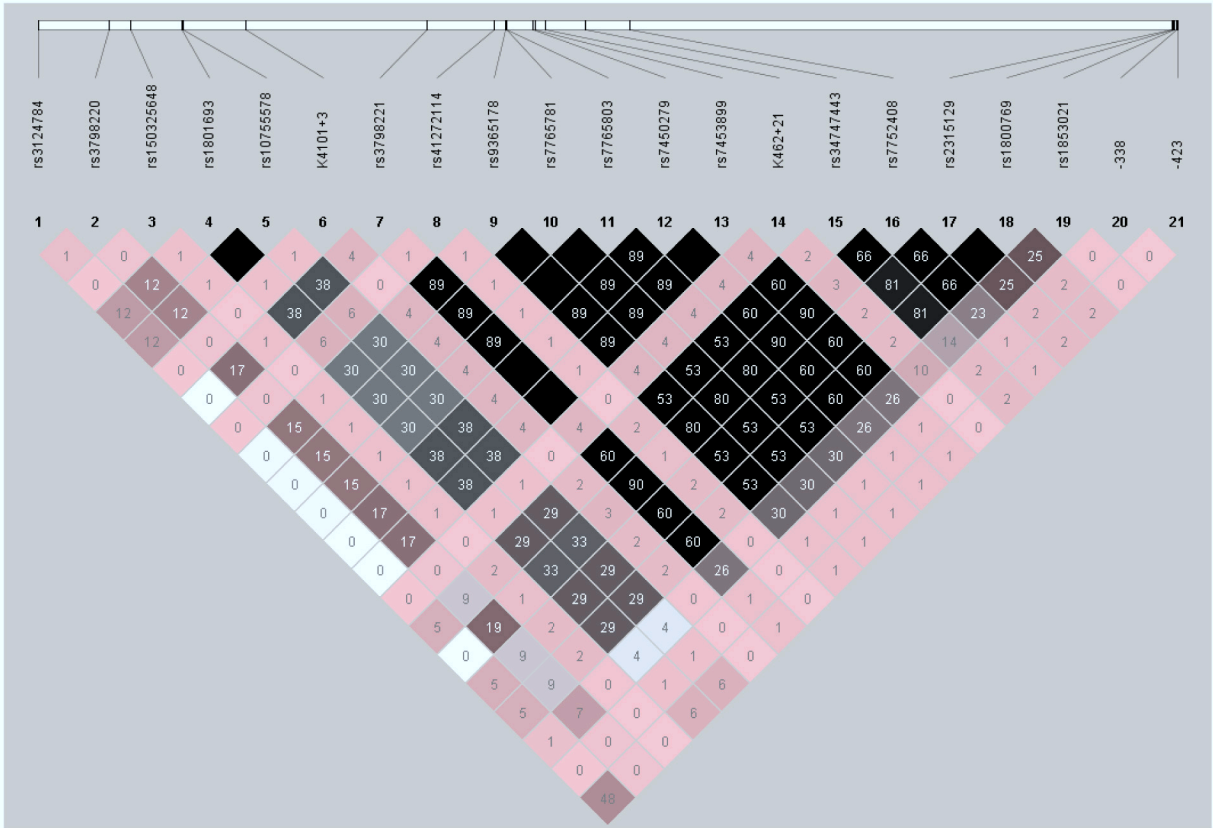
Supplemental Figure S4: KIV CNV genotype by rs3798220 genotype in individual Asian populations



Supplemental Figure S5:

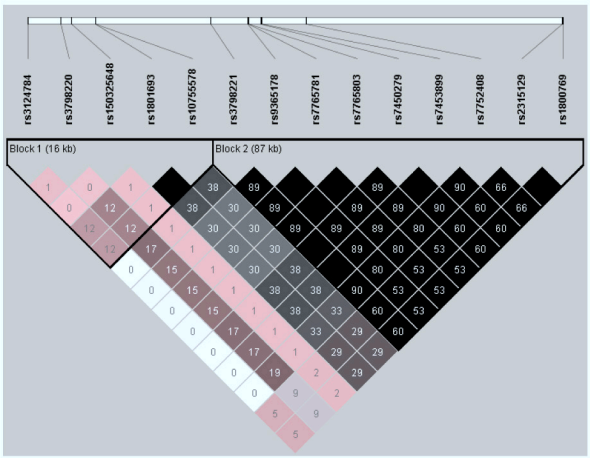
Supplemental Figure S6: LD between rs3798220 and other SNPs within *LPA* in phased alleles from Hong Kong Chinese

40 haplotyped alleles from Hong Kong Chinese (own data, all SNPs):

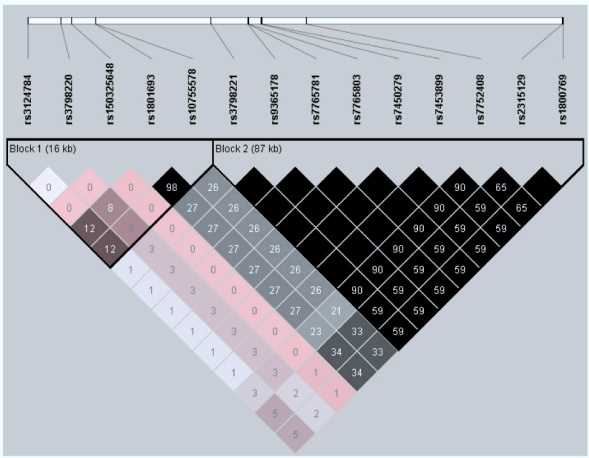


Supplemental Figure S7: LD between rs3798220 and other SNPs within *LPA*

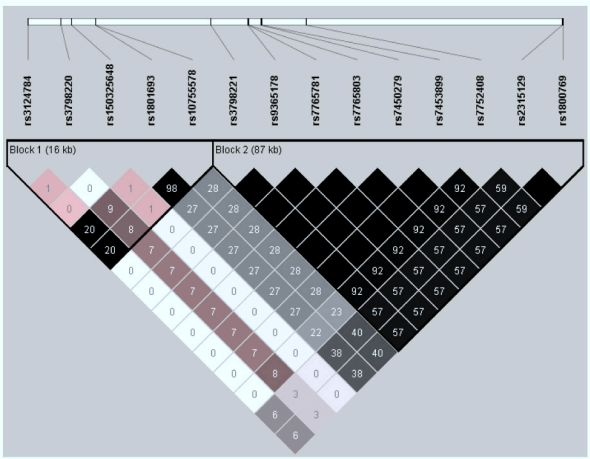
A: own Hong Kong Chinese



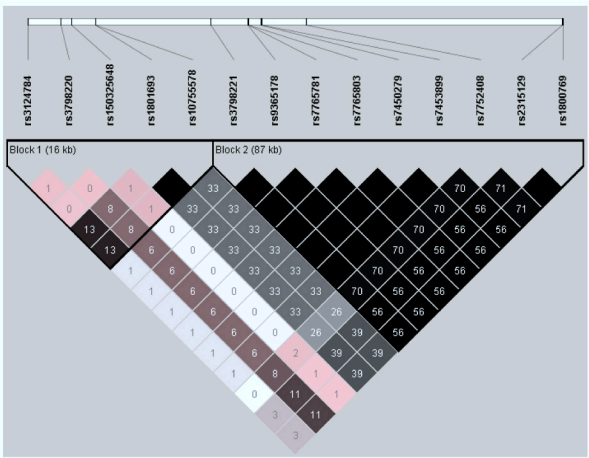
B: 1000G CHS panel



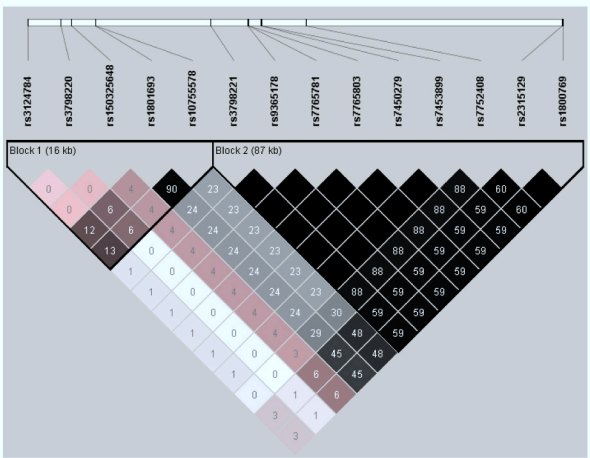
C: 1000G CHB panel



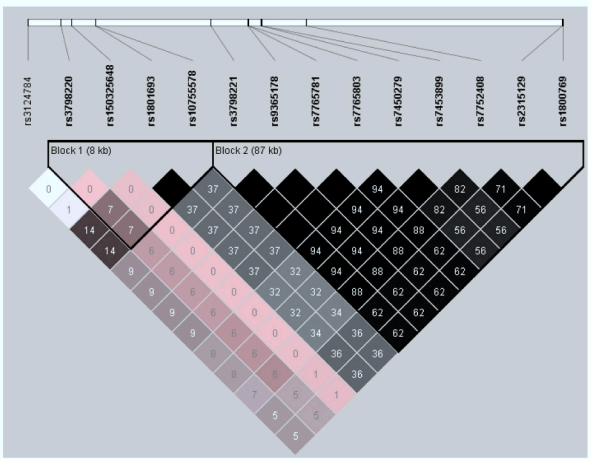
D: 1000G CDX panel



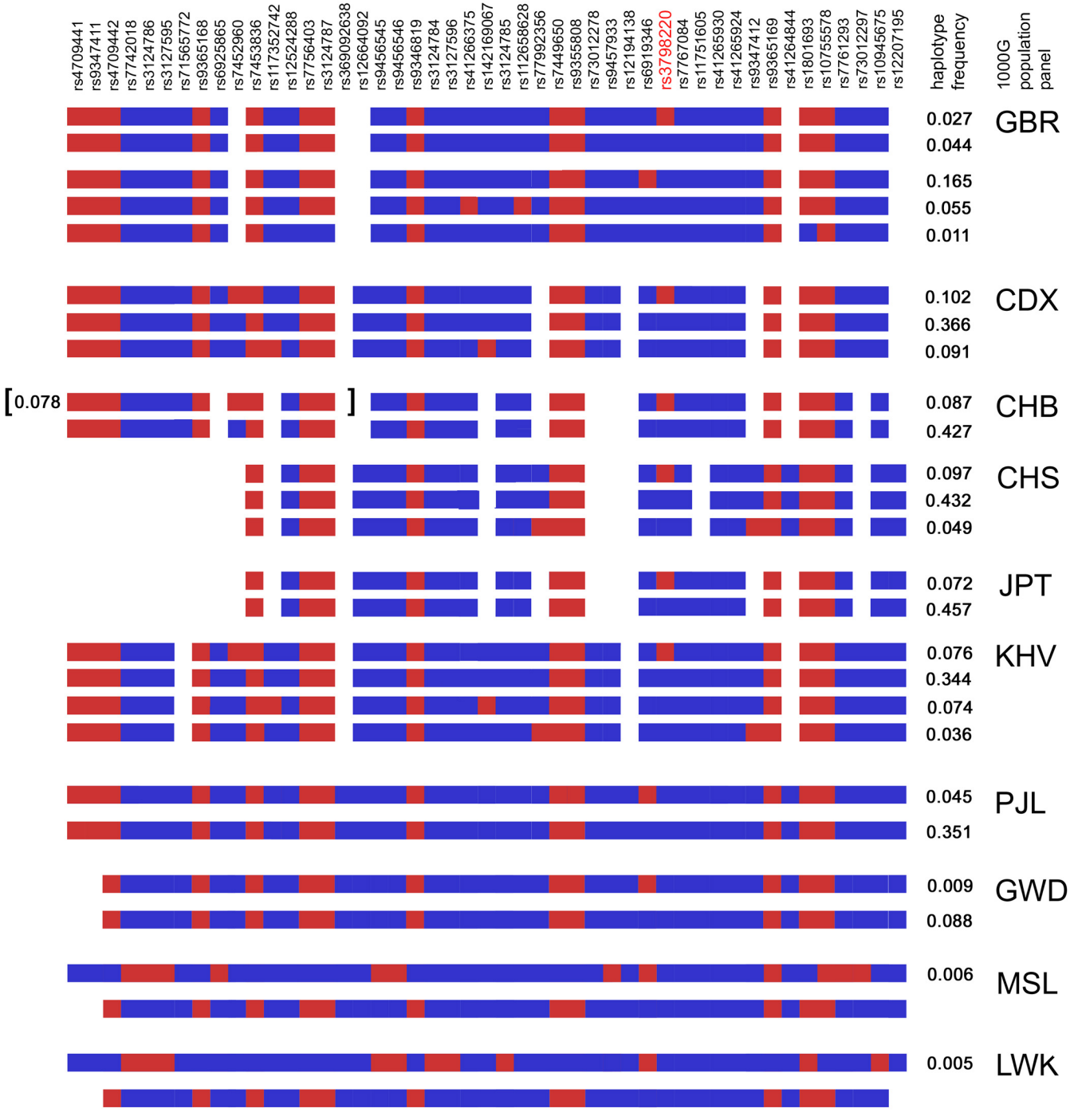
E: 1000G JPT panel



F: 1000G KHV panel



Supplemental Figure S8 Panel A:



Supplemental Figure S8 Panel B: