

RESEARCH ARTICLE

A genome-wide analysis of DNA methylation identifies a novel association signal for Lp(a) concentrations in the *LPA* promoter

Stefan Coassin¹, Natascha Hermann-Kleiter², Margot Haun¹, Simone Wahl^{3,4}, Rory Wilson^{3,4}, Bernhard Paulweber⁵, Sonja Kunze^{3,4}, Thomas Meitinger^{6,7,8}, Konstantin Strauch^{9,10,11}, Annette Peters^{4,6}, Melanie Waldenberger^{3,4,6}, Florian Kronenberg¹, Claudia Lamina^{1*}

1 Department of Genetics and Pharmacology, Institute of Genetic Epidemiology, Medical University of Innsbruck, Innsbruck, Austria, **2** Department of Genetics and Pharmacology, Institute of Cell Genetics, Medical University of Innsbruck, Innsbruck, Austria, **3** Research Unit of Molecular Epidemiology, Helmholtz Zentrum München—German Research Center for Environmental Health, Neuherberg, Germany, **4** Institute of Epidemiology II, Helmholtz Zentrum München—German Research Center for Environmental Health, Neuherberg, Germany, **5** First Department of Internal Medicine, Paracelsus Private Medical University, Salzburg, Austria, **6** German Center for Cardiovascular Research (DZHK), Partner Site Munich Heart Alliance, Munich, Germany, **7** Institute of Human Genetics, Technische Universität München, Munich, Germany, **8** German Research Center for Environmental Health, Institute of Human Genetics, Helmholtz Zentrum München, Neuherberg, Germany, **9** German Research Center for Environmental Health, Institute of Genetic Epidemiology, Helmholtz Zentrum München, Neuherberg, Germany, **10** Institute of Medical Informatics, Biometry, and Epidemiology, Ludwig-Maximilians-Universität, Munich, Germany **11** Institute of Medical Biostatistics, Epidemiology and Informatics (IMBEI), University Medical Center, Johannes Gutenberg University, Mainz, Germany

* claudia.lamina@i-med.ac.at



OPEN ACCESS

Citation: Coassin S, Hermann-Kleiter N, Haun M, Wahl S, Wilson R, Paulweber B, et al. (2020) A genome-wide analysis of DNA methylation identifies a novel association signal for Lp(a) concentrations in the *LPA* promoter. PLoS ONE 15(4): e0232073. <https://doi.org/10.1371/journal.pone.0232073>

Editor: Dana C. Crawford, Case Western Reserve University, UNITED STATES

Received: February 12, 2020

Accepted: March 19, 2020

Published: April 28, 2020

Copyright: © 2020 Coassin et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant summarized data are within the manuscript and its Supporting Information files. Individual level data of the KORA studies are subject to national data protection laws and restrictions were imposed by the ethics committee of the Bavarian Medical Association (Bayerische Landesärztekammer) to ensure data privacy of the study participants. Access to individual participant data in the KORA studies can be applied for at: <https://epi.helmholtz-muenchen.de/>

Abstract

Lipoprotein(a) [Lp(a)] is a major cardiovascular risk factor, which is largely genetically determined by one major gene locus, the *LPA* gene. Many aspects of the transcriptional regulation of *LPA* are poorly understood and the role of epigenetics has not been addressed yet. Therefore, we conducted an epigenome-wide analysis of DNA methylation on Lp(a) levels in two population-based studies (total $n = 2208$). We identified a CpG site in the *LPA* promoter which was significantly associated with Lp(a) concentrations. Surprisingly, the identified CpG site was found to overlap the SNP rs76735376. We genotyped this SNP de-novo in three studies (total $n = 7512$). The minor allele of rs76735376 (1.1% minor allele frequency) was associated with increased Lp(a) values ($p = 1.01e-59$) and explained 3.5% of the variation of Lp(a). Statistical mediation analysis showed that the effect on Lp(a) is rather originating from the base change itself and is not mediated by DNA methylation levels. This finding is supported by eQTL data from 208 liver tissue samples from the GTEx project, which shows a significant association of the rs76735376 minor allele with increased *LPA* expression. To evaluate, whether the association signal at rs76735376 may actually be derived from a stronger eQTL signal in LD with this SNP, eQTL association results of all correlated SNPs ($r^2 \geq 0.1$) were integrated with genetic association results. This analysis pinpointed to rs10455872 as the potential trigger of the effect of rs76735376. Furthermore, both SNPs coincide with short apo(a) isoforms. Adjusting for both, rs10455872 and the apo

Funding: The study was supported by the Austrian Science Fund (FWF) Project Number P266600-B13 to CL. The KORA study was initiated and financed by the Helmholtz Zentrum München – German Research Center for Environmental Health, which is funded by the German Federal Ministry of Education and Research (BMBF) and by the State of Bavaria. This work was supported by a grant (WA 4081/1-1) from the German Research Foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

(a) isoforms diminished the effect size of rs76735376 to 5.38 mg/dL ($p = 0.0463$). This indicates that the effect of rs76735376 can be explained by both an independent effect of the SNP and a strong correlation with rs10455872 and apo(a) isoforms.

Introduction

High Lipoprotein(a) [Lp(a)] concentrations represent a major risk factor for several cardiovascular diseases that is widely present in the population [1–3]. Depending on the cut-off used to define the normal range (30 mg/dL (reviewed in [4]) or 50 mg/dL[5]), 15% to 25% percent of the Caucasian population present Lp(a) concentrations that put them at increased cardiovascular risk[1–3]. Increased Lp(a) concentrations are consistently associated with cardiovascular diseases such as coronary artery disease[6,7] and myocardial infarction[8], stroke[9], peripheral artery disease[7,10], calcified aortic valve stenosis [7,11,12] and heart failure[13] and even cardiovascular as well as total mortality[7,14]. Conversely, very low Lp(a) concentrations have been shown to be associated with type 2 diabetes[7,15]. The associations with cardiovascular phenotypes as well as the unexpected association with diabetes[15] have been supported by several Mendelian randomization studies [6,12,13,16–19], underscoring an likely causal impact of Lp(a) on these phenotypes.

Lp(a) concentrations are mostly genetically determined by one major gene locus, the *LPA* gene[20], which presents a peculiar structure consisting of several so-called kringle domains (kringle 4 type one to type 10 and kringle 5; abbreviated KIV-1 to -10 and KV). The KIV-2 domain (consisting of two exons) is encoded by a hypervariable copy number variation, which can be present in 1 to >40 copies per allele[1] and thus generates >40 isoforms of the protein [apolipoprotein(a)]. In Caucasians, low molecular weight isoforms (11–22 KIV repeats) are associated with 4–5 times higher Lp(a) than high molecular isoforms with >22 repeats.

The apo(a) isoform determined by the KIV-2 repeat number is the strongest contributor to Lp(a) concentration variance and explains ≈ 30 –70% of the total trait variance, with the complete *LPA* locus explaining >90%[7,20,21]. Interestingly, Lp(a) levels of the same-sized isoforms vary by 200-fold in the general population[1] but <2.5-fold within families[22], suggesting the existence of genetic variation strongly modifying the impact of the isoform. Indeed, genome-wide SNP-bound heritability has been recently estimated to be $\approx 49\%$ in a German population[23].

Some high impact loss-of-function variants have been described[24–28] but regulatory variants have been elusive so far. The search for regulatory variants has so far focused on the promoter region[29–31] and a known enhancer region located 15 to 25 kb upstream of *LPA*[32], as well as on genome-wide association study efforts[23,33,34]. Three variants in the enhancer region have been found to regulate *LPA* expression in reporter assays[32]. Conversely, the promoter region of *LPA* is less well defined. Depending on whether a reference transcript with a leading non-coding first exon[29] (UCSC Genome browser hg38 annotation) or without it [30,31] (NCBI NM_005577.2) was used, two different genomic regions located ≈ 4 kb apart from each other have been investigated as putative promoters. The most extensively studied region extends about 1.4 kb upstream of the translation start. Some regulatory variants[35,36] and liver-specific regulatory elements[37] have been described in this region, as well as a pentanucleotide repeat (PNR) polymorphism with 5–12 repeats[30]. The latter is significantly associated with Lp(a) levels in Caucasians[38,39]. Early reports indicated a causal impact on promoter activity[30], which was not replicated in later studies[31]. Since the association of

the PNR with Lp(a) is, however, well replicated[38,39], it is likely that the PNR is simply in linkage disequilibrium with a regulatory variant[39].

Epigenetic regulation presents an additional layer of genetic regulation acting via chemical modifications of either histones[40] or the DNA sequence itself[41]. DNA methylation is a process by which methyl groups are added to the DNA molecule[41]. 5-Methylcytosine can be activating, when occurring in gene bodies, or repressing, when occurring in promoters[41] (albeit the repressing action of promoter methylation extends into the 5' region of the gene body[42]). DNA methylation may act by modifying the binding affinity of transcription factors[43] or recruiting silencing factors causing heterochromatinization[41]. Variation in the methylation level can thus affect both binding affinity of transcription factors and chromatin organization.

Allele-specific methylation (ASM) at SNPs represents a special case of gene regulation via methylation. Up to 50,000, mostly cis-acting, ASM-SNPs have been reported[42,44] so far. The large majority was located in CpG dinucleotides[45], where the SNP either creates or abolishes a CpG site (CpG-SNP). Such ASM-SNPs have been reported for e.g. inflammatory bowel disease[46], osteoarthritis susceptibility[47], atopic dermatitis[48], triglyceride levels[49] and diabetes[50]. Even methylation close to a SNP site may already modulate the allele specific binding of transcription factors to the SNP site itself[51].

The impact of methylation on regulation of lipid phenotypes has been studied to a lesser extent than germline SNPs on these phenotypes. The studies performed so far reported clear associations with lipid phenotypes such as VLDL-C, HDL-C, LDL-C, triglycerides and total cholesterol and even lipid-related diseases (reviewed in[52]), but none of these studies has addressed Lp(a). Therefore we report here the first epigenome-wide methylation association study on Lp(a) concentrations.

Results

Methylation analysis and methylation association study

Descriptive statistics of the participating studies are given in [S3 Table](#). One CpG site (cg17028067) located in the promoter region of the *LPA* gene was significantly associated with Lp(a) in the discovery sample KORA F4 ($p = 6.04e-11$, [Table 1](#)). This site was replicated in KORA F3 with direction-consistent effect size ($p = 0.0001$ in KORA F3, meta-analysis results: $p = 2.61e-14$; [Table 1](#)). This association result did not change after additionally adjusting for LDL-C in KORA F4 ($\beta(\text{se}) = -7.49(0.93)$, $p = 5.11e-11$).

Since low methylation beta-values of cg17028067 (mean \pm sd in KORA F4 = 0.811 ± 0.068 ; in KORA F3: 0.782 ± 0.061) were occurring predominantly in low molecular weight apo(a) isoforms with 18 to 20 KIV repeats ([S4 Fig](#)), the analysis was extended as follows ([Table 1](#)): first additionally adjusting for the lower of both apo(a) isoforms and secondly analyzing only the subgroup of participants with one isoform equal to 18–20 repeats (only in KORA F4 due to low sample size in KORA F3). Although the association was partially attenuated after adjustment for apo(a) isoforms ($p = 4.20e-7$, $n = 2187$) and in the subgroup of isoforms 18–20 ($p = 0.00265$, $n = 223$), it still remained highly significant, indicating that the association is not only a consequence of a correlation of the SNP with small apo(a) isoforms.

SNP genotyping

Surprisingly, the identified cg site was found to overlap a SNP (rs76735376). This C>T (reverse strand, i.e. transcribed strand of *LPA*) polymorphism is located exactly at the cytosine base of the CpG site cg17028067. The minor allele thus abolishes one potential methylation site by replacing the cytosine base of the CpG by a thymidine, thus effectively reducing the

Table 1. Results of the linear model of cg site cg17028067 on Lp(a).

	Study	n	Beta	Se	p-value
Model 1*	KORA F4	1724	-7.58	0.94	6.04e-11
	KORA F3	484	-11.08	1.91	0.0001
	Meta-analysis	2208	-8.26	0.84	2.61e-14
Model 2 [†] (adjusted for the smaller of both apo(a) isoforms)	KORA F4	1710	-4.66	0.82	7.24e-05
	KORA F3	477	-8.72	1.67	0.0007
	Meta-analysis	2187	-5.45	0.74	4.196e-07
Model 1* (in the subgroup of participants with one isoform equal to 18, 19 or 20 KIV repeats)	KORA F4	223	-3.17	1.72	0.00265

The p-value is derived from the log-transformed model, whereas beta estimate and standard error (se) are derived from the model on the original scale of Lp(a) to ease interpretability and refer to a change of 0.1 in methylation beta-value.

*Model 1: adjusted for age, sex, Houseman variables [87], 20 principal components of the control probes

[†]Model 2: as Model 1 + additionally adjusted for isoforms

<https://doi.org/10.1371/journal.pone.0232073.t001>

amount of available methylation target sites (Fig 1). This SNP had not been genotyped directly with the earlier SNP microarray but had been imputed with a low imputation quality (imputation quality info = 0.4) and a reported very low minor allele frequency of $\approx 0.1\%$. Therefore, this cg site had not been removed from the primary analysis dataset when SNP positions were masked. Even more surprising, subsequent de novo genotyping and confirmatory Sanger sequencing in KORA F3, KORA F4 and SAPHIR revealed that this SNP was considerably more frequent than reported in the imputed data with a true MAF of 1.1% and a carrier frequency of 2.2% (S4 Table).

Validation by genotyping and bisulfite sequencing

The methylation of the region was validated by bisulfite sequencing in 8 samples of the SAPHIR study (4 homozygotes for the major allele of rs76735376 and 4 heterozygotes, S5A Fig). Besides cg17028067 nine additional CpG sites nearby cg17028067 were all nearly 100% methylated (S2 Fig). None of these sites were represented on the HumanMethylation450

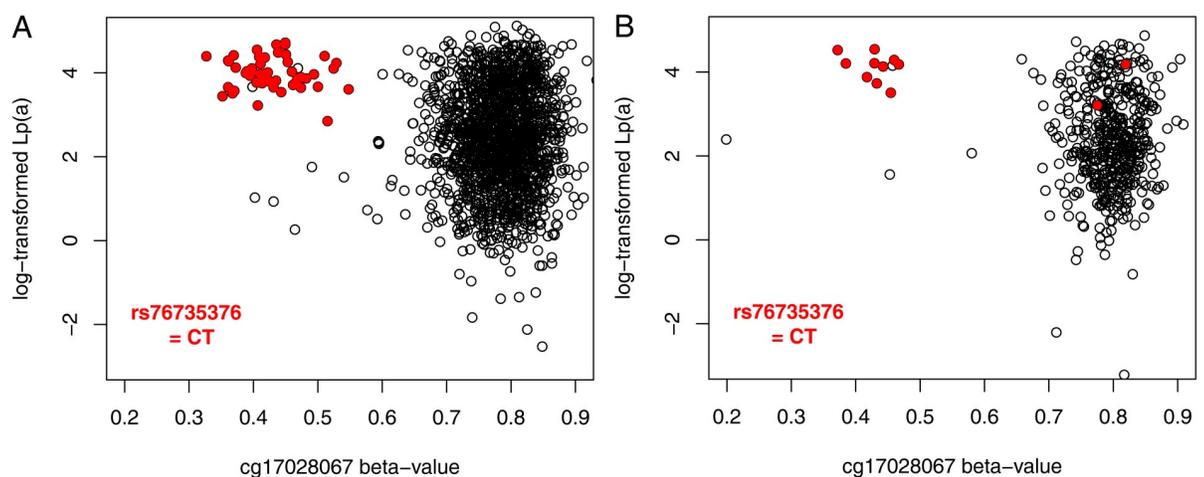


Fig 1. Association of methylation levels with Lp(a) values. Relationship of methylation level at cg17028067 (x axis), Lp(a) values (log-transformed (y axis) and rs76735376 genotype in the KORA F4 (panel A, n = 1724) and KORA F3 study (panel B, n = 484). Heterozygous carriers of rs76735376 (n = 49 in KORA F4, n = 12 in KORA F3) show roughly halved methylation levels and higher Lp(a) levels with less variance.

<https://doi.org/10.1371/journal.pone.0232073.g001>

BeadChip array. The methylation level of the cg sites nearby was independent from the status of cg17028067. This indicates that the whole region might contain CpG dinucleotides, which are target of DNA methylation and can thus potentially affect *LPA* expression. In heterozygous carriers of the SNP rs76735376 the methylation at cg17028067 was reduced by 50% as would be expected (S5B Fig).

SNP association analysis

The minor allele of rs76735376 was associated with increased Lp(a) values (Fig 2) with consistent effect sizes in all three cohorts ($p = 1.01e-59$, Table 2). Each copy of the minor allele was associated with an increase of Lp(a) concentrations by 37 mg/dL. The SNP explained 3.45% of the variance of Lp(a) concentrations. One part of the SNP effect seems to be due to correlation with apo(a) isoforms. As already observed for the methylation levels, isoforms 18–20 are over-represented in carriers of the minor allele (T) (S6A Fig). Additional adjusting for apo(a) isoforms leads to a reduction of the effect size ($\beta = 20.72$; $p = 1.72e-21$, Table 2).

Association with the pentanucleotide-repeat (PNR) polymorphism

Since rs76735376 is in close vicinity to the well-known *LPA* pentanucleotide-repeat (PNR) polymorphism [30,38,39,53], we evaluated in KORA F4 and SAPHIR, whether the association of the PNR with Lp(a) concentrations can be explained by rs76735376 or vice versa. In both studies, 89% of all participants present one allele with 8 repeats of this PNR, whereas all carriers of the minor allele of rs76735376 have this repeat number at least once (S5 Table, $p = 0.01$, Fisher's exact test in both studies combined). However, in a mixed effects model using data from both studies, the effect of the PNR on log-Lp(a) ($\beta = -0.15$, $se = 0.04$, $p = 0.0003$) is mediated only marginally by rs76735376 (results of PNR adjusted for rs76735376: $\beta = -0.14$, $se = 0.04$, $p = 0.0012$). The SNP-effect is not affected at all by adding the PNR to the model (effect of SNP only in both studies on log-Lp(a): $\beta = 1.55$, $se = 0.12$, $p = 3.82e-37$; Adjusted for PNR: $\beta = 1.53$, $se = 0.12$, $p = 1.62e-36$).

Mediation analysis

Since rs76735376 was identified actually by chance in a genome-wide methylation analysis, we investigated whether the effect of the SNP on Lp(a) is mediated through the methylation signal. A formal mediation analysis revealed no significant causal mediation effect of the methylation level of cg17028067 on log-Lp(a) ($p = 0.13$, S7 Fig). The total effect of the SNP on Lp(a) is only marginally altered by inclusion of the beta-value of cg17028067 in the model. In a regression model with both, rs76735376 and cg17028067, only the SNP remains significant. This might indicate that the true causal factor is the base change at rs76735376 rather than the methylation level at cg17028067.

Effect of rs76735376 on *LPA* transcription

LPA is expressed exclusively in liver tissue, which is hard to obtain for a large number of samples, especially if they need to be controlled for defined rare *LPA* genotypes. To assess whether rs76735376 indeed affects *LPA* expression in vivo, we performed a lookup in the data of the GTEx consortium V8 data [54]. GTEx is a multinational project providing both genome-wide SNP data and complete transcriptome data for >17,000 individuals and 54 different tissues, including eQTL data from 208 liver samples. The minor allele rs76735376 was significantly associated with increased *LPA* expression in liver tissue ($\beta(se) = +0.64 (0.18)$, $p = 0.00038$) despite only 15 heterozygous carriers in the data. Accordingly, bioinformatic transcription

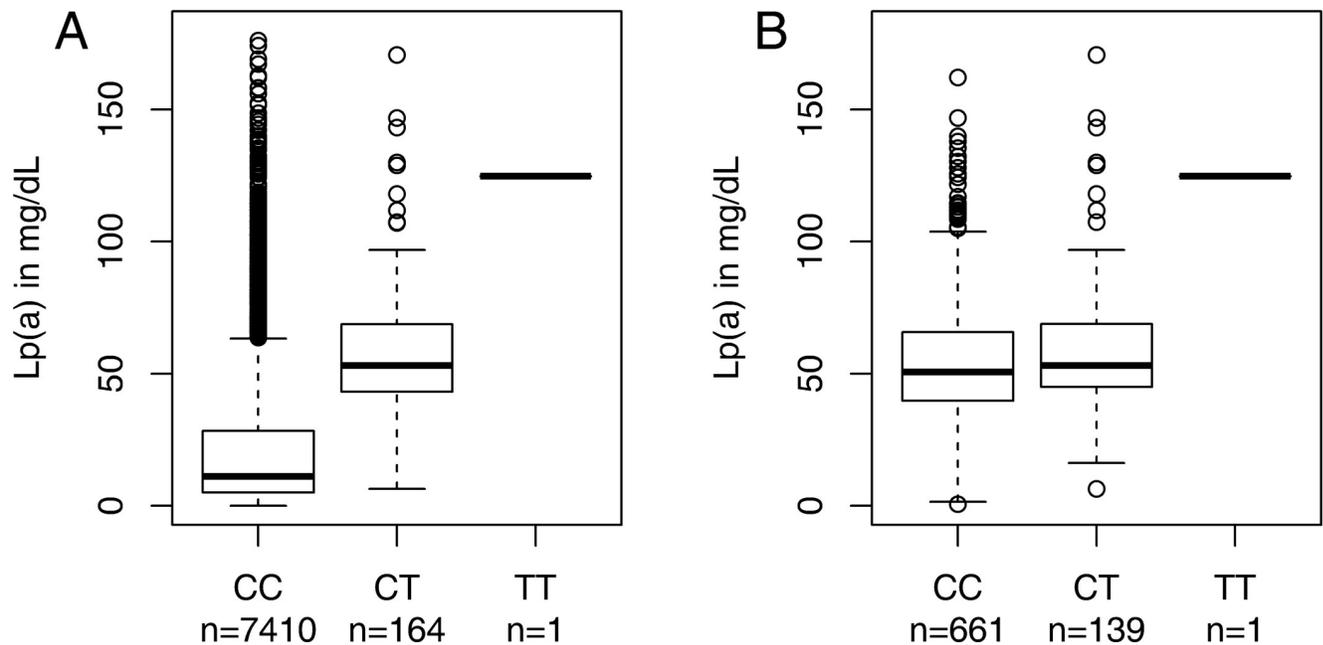


Fig 2. Distribution of Lp(a) stratified for genotypes of SNP rs76735376. In all participants of all three cohorts (panel A) and in subgroup of participants with at least one minor allele of rs10455872 (panel B).

<https://doi.org/10.1371/journal.pone.0232073.g002>

factor binding analysis using TRANSFAC[55] revealed a direct DNA transcription factor binding site for Oct factors *POU2F1* and *POU5F1* (also known as Oct1 and Oct3/4) within the T-allele of the rs76735376 SNP but not the C- or 5-mC allele (S3 Fig). Moreover, 5 bases besides the Oct binding site, a *CEBPB* was predicted (S3 Fig). Pilot experiments using electrophoretic mobility shift assays (EMSA) with probes encompassing the SNP site with C, 5-mC and T allele oligonucleotides suggests that *CEBPB* but not the Oct factors *POU2F1* and *POU5F1* indeed bind to both the C and 5-mC allele, but not to the T-allele (S8 Fig).

Integrating GWAS results with eQTL results and further association analyses

eQTL association results were further derived for all SNPs in LD ($r^2 \geq 0.1$) with rs76735376 (10 SNPs). S9 Fig shows a scatterplot of these SNPs between p-values derived from a genetic association analysis on Lp(a) versus eQTL p-values. The maximum ($-\log_{10}$)-transformed p-values for both analyses can be found the SNPs rs10455872[6] and rs118039278[23]. Since both are in perfect LD ($r^2 = 1$), we included only rs10455872 into the further analysis. While r^2 between rs76735376 and rs10455872 is only 0.1 and 0.17 in KORA F3 and KORA F4 due to different allele frequencies, almost all carriers of the minor allele of rs76735376 carry also at least one minor allele of rs10455872 (S6 Table), leading to a D' of 0.71 in KORA F3 and 0.95 in KORA F4. To evaluate, whether the association signal of rs76735376 can be explained by rs10455872, the linear regression model on Lp(a) was further adjusted for rs10455872, which reduced the effect size to $\approx +9$ mg/dL ($p = 4e-04$, Table 2).

Since rs76735376 coincides with isoforms 18 to 21 and since also rs10455872 was shown to be correlated with small isoforms (S6B Table), the analysis was further adjusted for both rs10455872 and apo(a) isoforms, leading to an increase of Lp(a) by 5.38 mg/dL ($p = 0.0463$) for each minor allele of rs76735376 (Table 2).

Table 2. Association analysis of rs76735376 on Lp(a) concentrations.

Study	Results of linear regression				Variance explained
	n	beta	se	p-value	
Adjusted for age and sex					
KORA F4	2986	36.42	2.81	3.23e-25	0.0384
KORA F3	3080	38.06	3.14	2.89e-27	0.0342
SAPHIR	1446	36.84	5.23	9.81e-11	0.0273
Combined*	7512	37.12	1.96	1.01e-59	0.0346[†]
Adjusted for age, sex and apo(a) isoforms:					
KORA F4	2986	19.65	2.49	1.37e-10	0.0062
KORA F3	3080	22.23	2.77	4.70e-09	0.0109
SAPHIR	1446	20.11	4.62	8.09e-05	0.0104
Combined*	7512	20.71	1.73	1.73e-21	0.0089[†]
Adjusted for age, sex and rs10455872:					
KORA F4	2857	5.94	2.82	0.1257	0.0026
KORA F3	2998	14.60	3.15	2.50e-04	0.0030
SAPHIR	1426	2.74	5.13	0.5740	0
Combined*	7281	8.95	1.96	4.00e-04	0.0016[†]
Adjusted for age, sex, rs10455872 and apo(a) isoforms:					
KORA F4	1831	2.94	2.59	0.2588	0
KORA F3	1976	9.66	2.83	0.0156	0.0017
SAPHIR	1413	1.12	4.71	0.8120	0
Combined*	7220	5.38	1.80	0.0463	0[†]

Models are adjusted for age and sex and additionally also for apo(a) isoforms and/or rs10455872. The p-value and variance explained are derived from the log-transformed model, whereas beta estimate and standard error (se) are derived from the model on the original scale of Lp(a) to ease interpretability and refer to the minor allele (T) of rs76735376.

*Linear mixed effects model with “study” as random effect;

[†]weighted average of study-specific variance explained

<https://doi.org/10.1371/journal.pone.0232073.t002>

Discussion

Lp(a) is a major genetically determined cardiovascular risk factor. Research on Lp(a) has experienced two revivals: the first by genetic studies providing strong support for a causal association with cardiovascular outcomes and the second, more recently, due to Lp(a)-lowering agents becoming available in the form of PCSK9 inhibitors and mRNA antisense therapy. Especially the advent of an antisense oligonucleotide therapy capable of reducing Lp(a) by >80%[56] now puts the validation of the “Lp(a) hypothesis” within reach[57]. Different studies reported that Lp(a) would have to be reduced by 50 mg/dL to even above 100 mg/dL [58–61] to produce the same risk reduction as LDL-C reduction by 1 mmol (38.67 mg/dL).

Despite the strong impact of Lp(a) on CVD risk, many controversies and uncertainties remain about its metabolism and regulation[62]. The association of Lp(a) concentrations with the apo(a) isoform size is well accepted, but far from being simple and linear[1,63]. Even within the same isoform group, levels may vary by 200-fold[22], suggesting the existence of modifier variants that act on top of the apo(a) isoforms. Some of these have been identified [24,25,28], but they all act via disturbing protein synthesis.

Regulatory variants and elements are less defined. Therefore, we performed the first genome-wide DNA methylation study for Lp(a). This identified the SNP rs76735376 located in a CpG site in the *LPA* promoter. Each minor allele abolishes a methylation site in a CpG

dinucleotide and is associated with an increase in Lp(a) of about 37 mg/dL and 9 mg/dL when restricted to individuals with 18–20 KIV repeats who are the main carriers of this methylation site. To dissect the effect of the base from the methylation site we performed a statistical mediation analysis which indicated that the effect observed in the association study is driven by the base change and not by the methylation status of the locus. This was supported by pilot EMSA experiments using the three possible allele states C, 5m-C and T), which showed no clearly defined difference between the C and the 5m-C allele, but a difference between the C and T alleles.

The location in the promoter suggests an effect on the transcriptional regulation of *LPA*. While Lp(a) is exclusively expressed in the liver and native tissue is hard to obtain in numbers sufficient to account for the low MAF of rs76735376. The recent large-scale multi-tissue and eQTL profiling initiative GTEx, however, provides a convenient way to still investigate the effects of rs76735376 on *LPA* expression. Indeed, a look-up for in the GTEx data revealed that the rs76735376 T-allele is associated with significantly increased *LPA* expression, which was in line with the observed increase of Lp(a) concentrations in these individuals.

Interestingly, rs76735376 was located close to the *LPA* pentanucleotide repeat (PNR). This led us to speculate that it might be the causal factor underlying the previously reported [30], but not clearly replicated [31] impact of the PNR on *LPA* expression. This was, however, not the case as the association signal at rs76735376 was only marginally affected by including the PNR in the model and vice versa. Of note, all carriers of the minor allele of rs76735376 carried a PNR allele with 8 copies and virtually all minor allele carriers presented an apo(a) allele with 19 to 20 KIV repeat (S6 Fig). Rs76735376 therefore adds to the other putative functional SNPs in *LPA*, which are confined to certain isoform ranges [28,36,64].

This is also true for the widely studied SNP rs10455872 [6], which coincides with rather small isoforms, too. Integrating eQTL association results with genetic association results, pinpointed to rs10455872 as the potential trigger of the effect of rs76735376. Indeed, adjusting for both, rs10455872 and the apo(a) isoforms diminished the effect size of rs76735376 from 37.1 mg/dL to 5.38 mg/dL. Although the p-value for this analysis is still <0.05 , it cannot be deemed significant in a strict sense. The methylation signal was identified via a genome-wide methylation analysis and the samples for the subsequent genotype association analysis were only partly independent and therefore, all results have to be accounted for this multiple testing situation.

Altogether, our results indicate that a major part of the effect of rs76735376 can be explained by correlation with rs10455872 and apo(a) isoforms. This underscores the complex genetic make-up of the Lp(a) trait, consisting of an intricate interplay between the KIV-2 repeat number, transcriptional regulation by enhancer and promoter regions, modifier SNPs, which may be in turn confined only to certain isoform sizes, and trans-regulating factors like apoE [23,65]. Given the large impact of the underlying apo(a) isoform on Lp(a) levels, a full assessment of variants influencing Lp(a) is likely to require complete knowledge about the isoforms of the investigated samples. Failure to account for the isoforms has been shown to even potentially mask true genetic effects [28,36].

In light of these observations are our findings still relevant and worth being discussed? Indeed, our study highlights several important messages for the community. First, effect size from the isoform- and rs10455872-adjusted analysis is still 5.69 mg/dL, which is about 50% of the median level of Lp(a) and therefore cannot be deemed to be irrelevant. The only person homozygous for the rare allele has an Lp(a) level of 124.75 mg/dL, which corresponds to the 99.4% percentile. Therefore it cannot be excluded that this SNP still might have some independent function or following a recessive mode of inheritance.

Secondly, rs10455872 explains a considerable part of the effect of rs76735376, despite presenting an r^2 of only 0.1 and 0.17. However, the high D' values of 0.71 and 0.95 are sufficient to

take up most of the signal. Indeed the D' values are often neglected when evaluating LD structures and it is often implicitly assumed that only a high r^2 can trigger an indirect signal. We exemplify here that this is not always the case. Moreover, it is also important to note how promising functional data may not necessarily translate to a true impact on the phenotype. Indeed mRNA levels do not necessarily always correlate strongly with the protein levels[66].

The third message is a rather technical one. Indeed, when we originally set out to investigate the impact of DNA methylation on Lp(a), the quality control procedure was supposed to purge any CpG sites affected by a SNP by using imputed 1000Genomes genotype data filtered at 5% MAF to focus only on differential methylation but not common ASM-SNPs. However, rs76735376 escaped filtering because it was badly imputed and thus incorrectly reported with a MAF of 0.1%. Since such a low MAF could not have triggered the association signal observed, we genotyped it de-novo, revealing a ten times higher true MAF. This MAF is now confirmed also by new reference datasets that became available only very recently (like TOPMed[67], <https://bravo.sph.umich.edu/freeze5/hg38/>). This exemplifies that, when drilling down from GWAS data to single SNPs, comparing the MAF from still widely used 1000 Genomes imputed data with that in the most recent resources like TOPMed can help to detect inconsistencies, which could otherwise bias the results if not handled properly.

Our study presents some limitations which need to be mentioned. Firstly, the methylation dataset was derived from peripheral blood, while *LPA* expression is limited to hepatic tissue. This is an unavoidable limitation of population-scale genome wide methylation analyses since hepatic tissue is not accessible at the sample numbers required for genome-wide studies. Indeed, even the GTEx initiative, which contains data from $\approx 17,000$ individuals and 54 different tissues (GTEx V8 as of December 2019), reports only 228 liver samples, thereof only 208 having also SNP genotype data available. While methylation is clearly a tissue-specific marker, studies also have shown that a consistent portion of the DNA methylation patterns is still conserved across tissues[68,69]. Therefore, while we might have missed some very tissue-specific regulators, we can be confident that a consistent portion of methylation-based regulation has been assayed. Importantly, the methylation status has been validated independently by Sanger sequencing. Unfortunately, due to limited amounts of biobank material available for bisulfite conversion, we acknowledge that we could perform this technical validation could only in a relatively low number of samples ($n = 8$).

Secondly, our dataset was of Caucasian origin. Given the known differences in allele frequencies and effects of *LPA* SNPs between ethnicities, investigations in other populations are required to establish the meaning of our findings in other ethnicities[70,71].

Thirdly, it needs to be mentioned that also GTEx presents limitations[72]. Most relevant for this work is that even eQTL signals still represent a form of association. Therefore, ultimately only functional studies like CRISPR editing can provide assess the true causality of SNPs on expression. They are, however, currently limited by the fact that *LPA* expression is blunted in cell culture for still unknown reasons and that the relationship of mRNA expression to secreted Lp(a) is not fully elucidated[73].

In conclusion, this study represents a first step assessing the epigenetic regulation of Lp(a). We showed that the analyzed region of the Lp(a) promoter is subjected to DNA methylation and that the identified CpG site cg17028067, respectively rs76735376, is associated with increased Lp(a) concentrations. This newly identified variant, however, is another example of the complex LD structures in the Lp(a) gene, since, despite clear eQTL signals, major parts of the effect could be finally explained by correlation with a complex interplay of apo(a) isoforms and rs10455872 genotype. This exemplifies the intricate genetic architecture of Lp(a).

Materials and methods

Study design and populations

The overall study design is outlined in [S1 Fig](#). Three population-based studies were available for this analysis: The KORA F3, KORA F4 and SAPHIR studies. The KORA study (Cooperative health research in the region of Augsburg) consists of independent, non-overlapping, population-based samples from the general population living in the region of Augsburg, Southern Germany, and was conducted in the years 2004/2005 (KORA F3) and 2006 to 2008 (KORA F4)[[74](#)]. A total of 3080 subjects with ages ranging from 32 to 81 years participated in the KORA F4 examination (stratified by sex-and age-groups). The genome-wide DNA methylation patterns were determined in a random subgroup of 1802 subjects (n = 1724 in analysis dataset with non-missing values in all relevant variables). In KORA F3 (n = 3184, age-range between 35 and 84 years), DNA methylation data were available from 484 subjects (comprised of 250 smokers and 250 non-smokers; 16 samples are missing due to data preprocessing or missing values in one of the relevant variables). For additional de novo genotyping, samples and data from 3080 KORA F3 and 2986 KORA F4 study participants were available, as well as samples from the SAPHIR study. For the SAPHIR study (Salzburg Atherosclerosis Prevention Program in subjects at High Individual Risk), participants were recruited by health screening programs in large companies in and around the city of Salzburg, Austria (n = 1770, age-range 39–67 years). Genotypes and relevant phenotype data were available for 1446 participants. All study participants provided a signed informed consent and the studies were approved by their respective local ethics committee (Bayerische Landesärztekammer for the KORA studies, Ethical Committee of Salzburg for SAPHIR). The PNR was measured in a subset of the KORA F4 (n = 2956) and SAPHIR (n = 1428) studies.

Genome-wide DNA methylation

Genome-wide DNA methylation patterns in the discovery and replication cohorts were analyzed using the Infinium HumanMethylation450 BeadChip Array (Illumina) and DNA derived from peripheral blood. KORA F3 and F4 samples were measured and processed separately [[75](#)]. KORA F3/F4 samples were processed on 20 respectively 7 96-well plates in 9 respectively 4 batches; plate and batch effects were investigated using principal component analysis and eigenR2 analysis[[76](#)]. DNA methylation data were preprocessed following the CPACOR pipeline[[77](#)]. Background correction was performed using the R package minfi, v1.6.0[[78](#)]. Signals with detection p-values (the probability of a signal being detected above the background) ≥ 0.01 were removed, as they indicate unreliable signals, as were signals summarized from fewer than three functional beads on the chip. Observations with less than 95% of CpG sites providing reliable signals were excluded.

To reduce the non-biological variability between observations, data were normalized using quantile normalization (QN) on the raw signal intensities: QN was performed on a stratification of the probe categories into 6 types, based on probe type and color channel, using the R package limma, v3.16.5[[79](#)].

The percentage of methylation at a given cytosine is reported as a beta-value, which is a continuous variable between 0 and 1 corresponding to the ratio of the methylated signal over the sum of the methylated and unmethylated signals of the particular cytosine site. Since it was the primary idea to detect the effect of differential methylation on Lp(a) and since probe binding might be affected by SNPs in the binding area, sites representing or being located in a 50 bp proximity to SNPs with a minor allele frequency (MAF) of at least 5% were excluded from

the data set[80]. Nine CpG sites in the *LPA* gene locus were present on the microarray (S1 Table).

SNP validation and *LPA* pentanucleotide repeat genotyping

KORA F4 samples were genotyped using the Affymetrix Axiom chip array. Genotypes were called with the Affymetrix software and were annotated to NCBI build 37. Imputation was performed with IMPUTE v2.3.0 using the 1000G phase1 (v3) reference panel[81]. Additional de novo genotyping for rs76735376 was performed in KORA F3, KORA F4 and the SAPHIR using a commercial Taqman assay (ThermoFisher Scientific).

Due to the low MAF, samples being heterozygous and homozygous for the rare allele were confirmed by Sanger sequencing using the primers 5' -TACAGGACAGAGACTAACT-3' and 5' -GCATAGTATCAATCTTTCCG-3'. The PCR conditions were: Qiagen Taq DNA Polymerase (#201207), 0.2 mM dNTP, 0.5 μ M primer, 1x Qiagen Q-Solution reagent, 40 ng input DNA amount. Cycling conditions were 94°C 3' initial denaturation, 40 cycles of 94°C 30 sec, 60°C 30 sec, 72°C 30 sec, 72°C 10 minutes final elongation. Sequencing was done using ABI BigDye 1.1 chemistry (Thermo Fisher Scientific Scientific, Waltham, MA, USA).

The *LPA* pentanucleotide repeat (PNR) is a microsatellite with 5-12 repeats of a TTTTA element located at hg19:chr6:161,086,617–161,086,663 (according to the RepeatMasker tool in the UCSC Genome Browser[82]) and thus only \approx 50 bp downstream of rs76735376. It has been genotyped by fragment analysis in SAPHIR and KORA F4. Details of the PNR genotyping via fragment analysis are given in reference[53]. In brief, a 181 bp PCR fragment encompassing the *LPA* pentanucleotide repeat was amplified in KORA F4 (data available in $n = 2956$) and SAPHIR (data available in $n = 1429$) in 384 well plates in a total volume of 5 μ l with 2.5 μ l Qiagen Multiplex PCR Plus Kit (Qiagen, Hilden, Germany) and 1 μ l Qiagen Q-solution using a Yakima Yellow-labelled primer. Fragment length was detected by fluorescent capillary fragment analysis on an ABI 3730s Genetic Analyzer and all data analyzed using the GeneMapper software (all Thermo Fisher Scientific). Data was quality controlled for null allele overrepresentation and deviation from Hardy-Weinberg Equilibrium using Micro-Checker [83] (<http://www.microchecker.hull.ac.uk/>) and ARLEQUIN (<http://cmpg.unibe.ch/software/arlequin3/>) [84].

Bisulfite sequencing

Bisulfite Sanger sequencing was used to validate the methylation signal at rs76735376 and map the methylation status of other CpG sites nearby in 8 selected samples from SAPHIR (4 homozygotes for the major allele and 4 heterozygotes). S2 Fig provides an overview of the region with location of the PNR, the primers, the amplicons and the CpG affected by rs76735376, as well as the neighboring CpG sites covered (respectively not covered) by our bisulfite sequencing.

500 ng sample DNA were bisulfite treated using the kit EZ DNA Methylation-Lightning Kit (Zymo Research, Irvine, CA, USA) and eluted in two times 10 μ l. 1 μ l eluted converted DNA was used as template for PCR. The region was amplified using the DNA Polymerase of the Agena PCR Accessory and Enzyme Set (Agena Bioscience, San Diego, CA, USA). PCR conditions were: 0.4 u Agena DNA Polymerase, 0.5 mM dNTP, 0.2 μ M primer, 2 mM MgCl₂, 1 μ l converted DNA. Cycling conditions were 94°C 4 minutes initial denaturation, 45 cycles of 94°C 20 sec, Annealing 30 sec, 72°C 60 sec, final elongation at 72°C 3 minutes. Primers for bisulfite-converted DNA were designed using the Agena EpiTyper Software Suite (see S2 Table for sequences and annealing temperature). All sequencing was done using ABI BigDye 1.1 chemistry.

Measurement of Lp(a) concentrations and apo(a) isoform size

Lp(a) concentration were determined in plasma using the same standardized sandwich ELISA assay in all samples. The assay has been described in detail previously[63,85]. In brief, ELISA plates were coated with an affinity-purified rabbit anti-human apo(a) antibody in a final concentration of 5 µg/ml in 1x PBS containing 1 mg/mL NaN₃. The plates were incubated with 100 µl antibody dilution (3 h, 37°C), washed three times (1x PBS + 0.05% v/v Tween-20) and blocked with 200 µl 0.1% w/v casein in 1x PBS pH 7.3 (30 min, 37°C). To ensure measuring each sample within the linear range of optical density, all samples were diluted into the ELISA plate twice (1:150 and 1:1,500 in Assay Buffer (Microcoat, Bernried, DE) out of 1:30 and 1:1000 predilutions in 1x PBS, pH 7.3). A 7-point standard curve ranging from 0.32 mg/dL to 5 µg/dL was created (with an additional blank representing the zero point). Duplicate determinations of four reference samples were used as longitudinal inter-assay controls. The coated plates were incubated with the analyte for 1 hour at 37°C. After several washing steps, the bound Lp(a) was detected using a horseradish-peroxidase-conjugated monoclonal antibody (1A2; in 0.1% w/v casein, 1x PBS, pH 7.3) directed against the KIV-2 domain and not cross-reacting with plasminogen[86] and the Blue Star TMB substrate (Adaltis, Guidonia Montecelio, IT) (30 minutes incubation at room temperature). Reaction was stopped by adding 50 µl 0.5 M sulfuric acid. Measurement of the absorption (dual wavelength, analyte: 450 nm, reference: 690 nm) was done using a Microplate Reader (BioRad Benchmark Plus, Bio-Rad Laboratories, Hercules, USA) and concentrations were calculated based on the standard curve (expressed as mg/dL). All dilution and pipetting steps were done using liquid handling robotics (TECAN, Männedorf, CH).

Lp(a) isoform was determined by Western blotting in all samples as described[63]. In brief, 150 ng Lp(a) as determined in the ELISA were loaded on a 1.46% agarose gel with 0.08% SDS and separated for 18 h at 0.04 A (constant current). A size standard consisting of a mixture of five plasma samples collected from individuals expressing each only one apo(a) isoform (13, 19, 23, 27, 35 KIV repeats determined by Pulsed Field Gel Electrophoresis) was applied in every seventh well of the gel. The gel was subsequently blotted to a PVDF membrane (Immobilion-P, Millipore, Darmstadt, DE) by semi-dry blotting (Perfect Blue Semi Dry Blotter, VWR, Vienna, A). After blocking (85 mM NaCl, 10 mM TRIS, 0.2% Triton X-100, 1% BSA; ≥30 min at 37°C), the membrane was incubated with horseradish-peroxidase-conjugated 1A2 antibody (1A2 concentration was 168 ng/ml in Buffer C) for 2 h at room temperature on a shaker. After washing with TTBS (3x 15 min), ECL substrate (WesternBright Chemilumineszenz Spray, Biozym, Hessisch Oldendorf, DE) was added and signals recorded on an Amersham Hyperfilm™ ECL™ (GE Healthcare, Vienna, A).

In heterozygous individuals, the smaller (and thus the commonly Lp(a) concentration-determining) of the two alleles was used for isoform-adjustment of statistical models.

Statistical methods

Genome wide methylation analysis. Linear regression models were applied to evaluate the association between DNA methylation beta values and Lp(a). Lp(a) levels were log-transformed to account for the skewed distribution. Age, sex and the first 20 principal components of the Illumina control probes (to adjust for technical confounding) were included as potential confounders. In addition, since whole blood DNA samples were used, cell heterogeneity had to be considered as a confounder. As no measured cell count information was available, sample-specific estimates of the proportion of the major white blood cell types were obtained using a statistical method described by Houseman et al[87]. To correct for multiple comparisons, a Bonferroni corrected significance level of 1.03e-07 (= 0.05/485512) was used. Genome-

wide significant findings were replicated in KORA F3 using the same adjustment model as in the discovery stage. Fixed-effects inverse-variance meta-analysis was applied to summarize the findings in KORA F3 and F4. The analyses were repeated additionally adjusting for apo(a) isoforms or in subgroups of apo(a) isoforms. Since it was shown previously that differential methylation can be a consequence of variation of lipid levels [88], the analyses were also adjusted for LDL cholesterol (LDL-C).

SNP association and mediation analysis. Association of the identified SNP with log-transformed Lp(a) was evaluated in all three studies individually (KORA F3, KORA F4, SAPHIR) using a linear regression model and in all three studies combined using a linear mixed effects model with “study” as a random effect and adjusting for age and sex. The genotypes of the SNP were additively coded, that is, effect sizes refer to one copy of the minor allele. This analysis was repeated additionally adjusting for apo(a) isoforms or in subgroups of apo(a) isoforms. Furthermore, as a consequence of the results of the analysis described in the next paragraph (“Integrating GWAS results with eQTL results”), the association analysis was further adjusted for rs10455872 and in subgroups of individuals with at least one minor allele of rs10455872.

In addition to the models with log-transformed Lp(a) values as the dependent variable, which were used to ensure the normal-distribution assumption to derive the test-statistic and p-values, Lp(a) on its original scale was used. From these models, beta estimates and standard errors (se) were derived to ease interpretability of the effect sizes.

Due to the close vicinity of the cg-site/SNP with the well-known PNR polymorphism, the association of the PNR with Lp(a) concentrations was evaluated and whether this effect can be explained by the SNP or vice versa.

Since both, the identified SNP and the methylation site are significantly associated with Lp(a) and since the methylation level is a consequence of the allelic variation of the SNP, it might be conceivable that there is no direct effect of the SNP on Lp(a), but that this effect is mediated by the methylation levels. To assess whether the effect of rs76735376 on Lp(a) is mediated through methylation, a formal mediation analysis was applied using the R library “mediation”.

Integrating GWAS results with eQTL results. To evaluate, whether the association signal at rs76735376 may actually be derived from a stronger eQTL signal in LD with this SNP, a list of SNPs with $r^2 \geq 0.1$ with rs76735376 was derived. This SNP-list was integrated with all SNPs, for which eQTL data from GTEx (see next paragraph) are available and with results derived from a genome-wide association study (GWAS) on Lp(a). The GWAS results are based on the KORA F4 study ($n = 2926$) and were part of a published meta-analysis[23]. Since rs76735376 was not well imputed in this dataset, results for rs76735376 were replaced with the results from the de novo genotyped SNP as described before in “SNP validation and LPA pentanucleotide repeat genotyping”. All other results are based on 1000G imputed SNPs as described in [23].

Follow-up in GTEx and transcription factor binding site prediction

The effect of rs76735376 on *LPA* gene expression in native liver tissue was assessed using the GTEx[89] data (www.gtexportal.com), which contains expression quantitative trait locus (eQTL) data with whole genome SNP data and whole transcriptome RNA-Seq data for 208 liver samples. The correlation of rs76735376 genotype with *LPA* expression in liver was extracted from the GTEx V8 dataset by downloading all pairwise SNP-expression correlations in liver tissue and using the SNP-ID “chr6_160665684” (chromosome and position of rs76735376 in hg38) and the Ensembl gene identifier for *LPA* “ENSG00000198670” to extract the data for rs76735376 and *LPA*. The dataset is available at https://console.cloud.google.com/storage/browser/_details/gtex-resources/GTEX_Analysis_v8_QTLs/GTEX_Analysis_v8_

eQTL_all_associations/Liver.allpairs.txt.gz, accessible via the link “eQTL Tissue-Specific All SNP Gene Associations” at <https://gtexportal.org/home/datasets>, respectively at <https://console.cloud.google.com/storage/browser/gtex-resources> (accessed on November, 26th 2019)

TRANSFAC (<http://genexplain.com/transfac/>) [55], a database predicting the genomic binding sites and DNA-binding profiles of eukaryotic transcription factors, was used in order to determine transcription factor binding sites at the genomic region of rs76735376.

Gel mobility shift assay

To perform a pilot validation of the mediation analysis results, single-stranded oligonucleotides spanning the SNP site were synthesized by Eurofins MWG Operon and then annealed to generate double-stranded probes. The single-stranded oligonucleotide sequences for the C, T and 5-mC allele of rs76735376 respectively are shown in [S3 Fig](#). 5-mC bases were introduced directly by the oligonucleotide provider during oligonucleotide synthesis. Annealing was performed by heating a mix of both oligos (2.5 μ g in 100 μ l with 1X annealing buffer: 20 mM TRIS pH 8.0, 1 mM EDTA, 50 mM NaCl) to 65° C for 5 minutes and slowly cooling down in ice water. Binding reactions for super shift assays were performed for 30 min at 4–8° C using the Binding Buffer B-1 (Active Motif 37480) together with Stabilizing Solution D (Active Motif 37488) containing either 5 μ g human liver or embryonic kidney nuclear extracts (Active Motif 36042; 36033), 2 μ g of either anti c/EBPbeta (*CEBPB*) (Santa Cruz Biotechnology; sc-150x), anti-Oct-1 (*POU2F1*) (Santa Cruz Biotechnology; sc232) plus anti-Oct-3/4 (*POU5F1*) (Santa Cruz Biotechnology; sc5279) or anti-Nfatc1 (*NFATC1*) as negative control (Santa Cruz Biotechnology; sc7294). Binding reactions with 3×10^5 cpm of labeled probe was performed for 20 min at 4°–8° C using Binding Buffer C-1 (Active Motif 37484) together with Stabilizing Solution D (Active Motif 37488). Samples were separated on a 4% native polyacrylamide gel in $0.5 \times$ TBE for 3 h at 250 V. For competition assays, 10-fold unlabeled cold oligonucleotides identical to the radioactively labeled non-methylated Oligo2 LPA_EMAS2 or the radioactively labeled methylated Oligo6 LPA_EMAS2 probes were added to the binding reaction in addition.

Supporting information

S1 Table. CpG sites in the LPA gene locus assayed by the illumina Infinium HumanMethylation450 BeadChip array with position, location and the p-value of the first stage epigenome-wide analysis on $\log(Lp(a))$. The location is given relative to the UCSC Genome browser hg38 annotation with a leading non-coding exon 1. See the [introduction](#) section of the main manuscript for an explanation of the issues with the LPA reference transcript annotation. (PDF)

S2 Table. Primer sequences. For Bisulfite Sanger sequencing primer the lower case letters denote tails that have been added by the design software to increase the annealing temperature of the primer. The PCR protocols are given in the Methods section of the manuscript. See [S2 Fig](#) for an overview of the target sequence. (PDF)

S3 Table. Study characteristics (Mean \pm sd [25%,50%,75% Percentile] for quantitative variables, n(%) for gender. (PDF)

S4 Table. Genotype frequencies of the de novo genotyped SNP rs76735376 in three cohorts. (PDF)

S5 Table. Allele frequencies of the PNR in SAPHIR and KORA F4, separated for genotypes of rs76735376.

(PDF)

S6 Table. Frequencies of the combined genotype distributions of rs76735376 with rs10455872 in all three cohorts together.

(PDF)

S1 Fig. Flow chart of the study design.

(PDF)

S2 Fig. Overview of the region around rs76735376 with location of the LPA pentanucleotide repeat, the location of the amplicons, the location of the CpG affected by rs76735376 and neighboring CpG sites. Upper case: sequence regions, which could be covered with amplicons for bisulfite sequencing. Lower case: regions which could not be covered by amplicons for bisulfite sequencing. Blue background (light and dark): primer binding sites for bisulfite sequencing. Note that the sequence given here is the unconverted sequence. The primer given in [S2 Table](#) bind on the converted sequence and the complement strand of the strand shown here. Yellow background: primer binding sites of the sequences used for SNP validation by sequencing. Bold character: LPA pentanucleotide repeat. Green: CpG affected by rs76735376 (enlarged cytosine base). Position of rs76735376 is underlined. Pink: CpG in the region, which were amenable to sequencing. Purple: CpG in the region, which were not amenable to sequencing (not represented in any amplicon). Sequence is the bisulfite-converted sequence given 5' to 3' on the minus strand of human genome GRCh37/hg19 (i.e. the same direction as LPA transcription direction).

(PDF)

S3 Fig. Oligos used for EMSA experiments. The bold case base gives the location of rs76735376. Green background: POU2F1/POU5F1 binding site, Green font: mutated POU2F1/POU5F1 binding site, Blue CEBPB binding site. The reverse primers (-r) are given in reverse orientation, as they are annealed to the forward oligos (-f).

(PDF)

S4 Fig. Scatterplot showing the lower of both apo(a) isoforms (x-axis) per person in KORA F3 (in red) and KORA F4 (in black) versus the beta-value of cg17028067 (y-axis).

(PDF)

S5 Fig. Panel A: Representative results of bisulfite sequencing in two homozygotes for the major allele and two heterozygotes in the SAPHIR study. The blue peak represents the unconverted C-allele, indicating that the major part is methylated in CC carriers, whereas only a minor part is methylated in CT carriers. **Panel B:** Boxplots of the methylation level (expressed as beta-value) of cg17028067, stratified for genotypes in the KORAF4 study (panel B).

(PDF)

S6 Fig. Distribution of the smaller of both isoforms, stratified for carriers (CT or TT) and non-carriers (CC) of the SNP rs76735376 rs76735376 (panel A) and carriers and non-carriers of rs10455872 (panel B)

(PDF)

S7 Fig. Possible paths and results of mediation analysis.

(PDF)

S8 Fig. Electrophoretic mobility shift assay for rs76735376. rs76735376 modifies binding of the transcription factor CEBPB. Left panel: Diagram of the double-stranded oligos used for EMSA analysis containing either cytosine (red, Oligo2 LPA_EMSA2-C), thymine (Oligo4 LPA_EMSA2-T) or methyl-cytosine (6 LPA_EMSA2-5mC) oligo. The transcription factor binding sites for CEBPB (blue) and POU2F1/POU5F1 (green) as predicted by Transfac analysis, 52 of the rs76735376 probes are shown. Position weight matrix of the TF is shown. The arrow indicates the orientation of the sequence matrix in the oligo. Right panel: EMSA analysis of human liver nuclear extract hybridized to Oligo2 LPA_EMSA2-C, Oligo4 LPA_EMSA2-T or Oligo6 LPA_EMSA2-5mC. Binding specificity to specific transcription factors was performed by using super-shift antibodies (s-shift AB) for the POU factors or CEBPB. An NFATC1 antibody and excess of cold unlabeled oligonucleotide (cold comp.) were used as negative control. One representative experiment out of three is shown. Note that the free probe has already left the gel because a long run time in order was required to distinguish the super-shift bands. (PDF)

S9 Fig. Scatterplot of p-values from a GWAS on Lp(a) versus p-values from eQTL analysis. The x-axis shows p-values from a GWAS on Lp(a) (using data from the KORA F4 study), the y-axis p-values from eQTL analysis (GTEx consortium V8). The index SNP (rs76735376) is marked as purple triangle. Only those SNPs are shown, which show a r^2 of ≥ 0.1 with the index SNP (LD information derived from 1000 genomes CEU). Color coding shows magnitude of LD. (PDF)

Author Contributions

Conceptualization: Stefan Coassin, Natascha Hermann-Kleiter, Florian Kronenberg, Claudia Lamina.

Data curation: Simone Wahl, Bernhard Paulweber, Sonja Kunze, Thomas Meitingner, Konstantin Strauch, Annette Peters, Melanie Waldenberger, Claudia Lamina.

Formal analysis: Stefan Coassin, Claudia Lamina.

Funding acquisition: Claudia Lamina.

Investigation: Stefan Coassin, Natascha Hermann-Kleiter, Margot Haun, Claudia Lamina.

Methodology: Stefan Coassin, Natascha Hermann-Kleiter, Simone Wahl, Rory Wilson, Sonja Kunze, Melanie Waldenberger.

Project administration: Claudia Lamina.

Supervision: Claudia Lamina.

Visualization: Stefan Coassin, Claudia Lamina.

Writing – original draft: Stefan Coassin, Natascha Hermann-Kleiter, Rory Wilson, Melanie Waldenberger, Claudia Lamina.

Writing – review & editing: Stefan Coassin, Natascha Hermann-Kleiter, Rory Wilson, Sonja Kunze, Melanie Waldenberger, Florian Kronenberg, Claudia Lamina.

References

1. Kronenberg F, Utermann G. Lipoprotein(a): resurrected by genetics. *J Intern Med.* 2013; 273: 6–30. <https://doi.org/10.1111/j.1365-2796.2012.02592.x> PMID: 22998429

2. Kronenberg F. Human Genetics and the Causal Role of Lipoprotein(a) for Various Diseases. *Cardio-vasc Drugs Ther.* 2016; 30: 87–100. <https://doi.org/10.1007/s10557-016-6648-3> PMID: 26896185
3. Nordestgaard BG, Langsted A. Lipoprotein(a) as a cause of cardiovascular disease: Insights from epidemiology, genetics, and biology. *J Lipid Res.* 2016; 57: 1953–1975. <https://doi.org/10.1194/jlr.R071233> PMID: 27677946
4. Kostner KM, Kostner GM. Lipoprotein (a): a historical appraisal. *J Lipid Res.* 2017; 58: 1–14. <https://doi.org/10.1194/jlr.R071571> PMID: 27821413
5. Nordestgaard BG, Chapman MJ, Ray K, Borén J, Andreotti F, Watts GF, et al. Lipoprotein(a) as a cardiovascular risk factor: current status. *Eur Heart J. Department of Clinical Biochemistry, Herlev Hospital, Copenhagen University Hospital, University of Copenhagen, DK-2730 Herlev, Denmark. brno@heh.regionh.dk;* 2010; 31: 2844–53. <https://doi.org/10.1093/eurheartj/ehq386> PMID: 20965889
6. Clarke R, Peden JF, Hopewell JC, Kyriakou T, Goel A, Heath SC, et al. Genetic variants associated with Lp(a) lipoprotein level and coronary disease. *N Engl J Med. Clinical Trial Service Unit, University of Oxford, Oxford, United Kingdom;* 2009; 361: 2518–28. <https://doi.org/10.1056/NEJMoa0902604> PMID: 20032323
7. Gudbjartsson DF, Thorgeirsson G, Sulem P, Helgadóttir A, Gylfason A, Saemundsdóttir J, et al. Lipoprotein(a) Concentration and Risks of Cardiovascular Disease and Diabetes. *J Am Coll Cardiol.* 2019; 74: 2982–2994. <https://doi.org/10.1016/j.jacc.2019.10.019> PMID: 31865966
8. Kamstrup PR, Tybjaerg-Hansen A, Steffensen R, Nordestgaard BG. Genetically elevated lipoprotein(a) and increased risk of myocardial infarction. *JAMA.* 2009; 301: 2331–9. <https://doi.org/10.1001/jama.2009.801> PMID: 19509380
9. Langsted A, Nordestgaard BG, Kamstrup PR. Elevated Lipoprotein(a) and Risk of Ischemic Stroke. *J Am Coll Cardiol.* 2019; 74: 54–66. <https://doi.org/10.1016/j.jacc.2019.03.524> PMID: 31272552
10. Laschkolnig A, Kollerits B, Lamina C, Meisinger C, Rantner B, Stadler M, et al. Lipoprotein (a) concentrations, apolipoprotein (a) phenotypes, and peripheral arterial disease in three independent cohorts. *Cardiovasc Res.* 2014; 103: 28–36. <https://doi.org/10.1093/cvr/cvu107> PMID: 24760552
11. Kamstrup PR, Tybjaerg-Hansen A, Nordestgaard BG. Elevated lipoprotein(a) and risk of aortic valve stenosis in the general population. *J Am Coll Cardiol.* 2014; 63: 470–7. <https://doi.org/10.1016/j.jacc.2013.09.038> PMID: 24161338
12. Thanassoulis G, Campbell CY, Owens DS, Smith JG, Smith A V., Peloso GM, et al. Genetic Associations with Valvular Calcification and Aortic Stenosis. *N Engl J Med.* 2013; 368: 503–512. <https://doi.org/10.1056/NEJMoa1109034> PMID: 23388002
13. Kamstrup PR, Nordestgaard BG. Elevated Lipoprotein(a) Levels, LPA Risk Genotypes, and Increased Risk of Heart Failure in the General Population. *JACC Hear Fail.* 2016; 4: 78–87. <https://doi.org/10.1016/j.jchf.2015.08.006> PMID: 26656145
14. Langsted A, Kamstrup PR, Nordestgaard BG. High lipoprotein(a) and high risk of mortality. *Eur Heart J.* 2019; 40: 2760–2770. <https://doi.org/10.1093/eurheartj/ehy902> PMID: 30608559
15. Mora S, Kamstrup PR, Rifai N, Nordestgaard BG, Buring JE, Ridker PM. Lipoprotein(a) and risk of type 2 diabetes. *Clin Chem.* 2010; 56: 1252–1260. <https://doi.org/10.1373/clinchem.2010.146779> PMID: 20511445
16. Kamstrup PR, Nordestgaard BG. Lipoprotein(a) concentrations, isoform size, and risk of type 2 diabetes: a Mendelian randomisation study. *Lancet Diabetes Endocrinol.* 2013; 1: 220–227. [https://doi.org/10.1016/S2213-8587\(13\)70064-0](https://doi.org/10.1016/S2213-8587(13)70064-0) PMID: 24622370
17. Sandholzer C, Saha N, Kark JD, Rees A, Jaross W, Dieplinger H, et al. Apo(a) isoforms predict risk for coronary heart disease. A study in six populations. *Arter Thromb.* 1992; 12: 1214–26. Available: <http://www.ncbi.nlm.nih.gov/pubmed/1390593>
18. Erqou S, Thompson A, Di Angelantonio E, Saleheen D, Kaptoge S, Marcovina S, et al. Apolipoprotein (a) isoforms and the risk of vascular disease: systematic review of 40 studies involving 58,000 participants. *J Am Coll Cardiol.* Elsevier Inc.; 2010; 55: 2160–7. <https://doi.org/10.1016/j.jacc.2009.10.080> PMID: 20447543
19. Saleheen D, Haycock PC, Zhao W, Rasheed A, Taleb A, Imran A, et al. Apolipoprotein(a) isoform size, lipoprotein(a) concentration, and coronary artery disease: a mendelian randomisation analysis. *lancet Diabetes Endocrinol.* 2017; 5: 524–533. [https://doi.org/10.1016/S2213-8587\(17\)30088-8](https://doi.org/10.1016/S2213-8587(17)30088-8) PMID: 28408323
20. 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; 90: 52–60. <https://doi.org/10.1172/JCI115855> PMID: 1386087

21. 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; 14: 190–201. <https://doi.org/10.1038/sj.ejhg.5201512> PMID: 16267501
22. Perombelon YFN, Soutar AK, Knight BL. Variation in lipoprotein(a) concentration associated with different apolipoprotein(a) alleles. *J Clin Invest.* 1994; 93: 1481–1492. <https://doi.org/10.1172/JCI117126> PMID: 8163653
23. Mack S, Coassin S, Rueedi R, Yousri NA, Seppälä I, Gieger C, et al. A genome-wide association meta-analysis on lipoprotein (a) concentrations adjusted for apolipoprotein (a) isoforms. *J Lipid Res.* 2017; 58: 1834–1844. <https://doi.org/10.1194/jlr.M076232> PMID: 28512139
24. 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; 8: 2087–96. <https://doi.org/10.1093/hmg/8.11.2087> PMID: 10484779
25. Parson W, Kraft HG, Niederstätter H, Lingenhel AW, Köchl S, Fresser F, et al. 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.* Institute of Forensic Medicine, Medical University of Innsbruck, Innsbruck, Austria; 2004; 24: 474–80. <https://doi.org/10.1002/humu.20101> PMID: 15523644
26. Lim ET, Würtz P, Havulinna AS, Palta P, Tukiainen T, Rehnström K, et al. Distribution and medical impact of loss-of-function variants in the Finnish founder population. *PLoS Genet.* 2014; 10: e1004494. <https://doi.org/10.1371/journal.pgen.1004494> PMID: 25078778
27. Emdin CA, Khera A V., Natarajan P, Klarin D, Won H-H, Peloso GM, et al. Phenotypic Characterization of Genetically Lowered Human Lipoprotein(a) Levels. *J Am Coll Cardiol.* 2016; 68: 2761–2772. <https://doi.org/10.1016/j.jacc.2016.10.033> PMID: 28007139
28. Coassin S, Erhart G, Weissensteiner H, Eca Guimarães de Araújo M, Lamina C, Schönherr S, et al. A novel but frequent variant in LPA KIV-2 is associated with a pronounced Lp(a) and cardiovascular risk reduction. *Eur Heart J.* 2017; 38: 1823–1831. <https://doi.org/10.1093/eurheartj/ehx174> PMID: 28444229
29. Chennamsetty I, Claudel T, Kostner KM, Baghdasaryan A, Kratky D, Levak-Frank S, et al. Farnesoid X receptor represses hepatic human APOA gene expression. *J Clin Invest.* Institute of Molecular Biology and Biochemistry, Center of Molecular Medicine, Medical University of Graz, Graz, Austria; 2011; 121: 3724–34. <https://doi.org/10.1172/JCI45277> PMID: 21804189
30. Wade DP, Clarke JG, Lindahl GE, Liu AC, Zysow BR, Meer K, et al. 5' control regions of the apolipoprotein(a) gene and members of the related plasminogen gene family. *Proc Natl Acad Sci U S A.* 1993; 90: 1369–73. Available: <https://doi.org/10.1073/pnas.90.4.1369> PMID: 7679504
31. Bopp S, Köchl S, Acquati F, Magnaghi P, Pethö-Schramm A, Kraft HG, et al. Ten allelic apolipoprotein [a] 5' flanking fragments exhibit comparable promoter activities in HepG2 cells. *J Lipid Res.* Department of Molecular Biology, Boehringer Mannheim GmbH, Germany; 1995; 36: 1721–8. Available: <http://www.jlr.org/content/36/8/1721.long> PMID: 7595093
32. Puckey LH, Knight BL. Sequence and functional changes in a putative enhancer region upstream of the apolipoprotein(a) gene. *Atherosclerosis.* 2003; 166: 119–27. [https://doi.org/10.1016/s0021-9150\(02\)00315-5](https://doi.org/10.1016/s0021-9150(02)00315-5) PMID: 12482558
33. Trégouët D-A, König IR, Erdmann J, Munteanu A, Braund PS, Hall AS, et al. Genome-wide haplotype association study identifies the SLC22A3-LPAL2-LPA gene cluster as a risk locus for coronary artery disease. *Nat Genet.* Institut National de la Sante Et de la Recherche Medicale (INSERM) Unite Mixte de Recherche (UMR_S) 525, Universite Pierre et Marie Curie (UPMC). Paris 06, Paris 75013, France. david.tregouet@upmc.fr; 2009; 41: 283–5. <https://doi.org/10.1038/ng.314> PMID: 19198611
34. Li J, Lange LA, Sabourin J, Duan Q, Valdar W, Willis MS, et al. Genome- and exome-wide association study of serum lipoprotein (a) in the Jackson Heart Study. *J Hum Genet.* The Japan Society of Human Genetics; 2015; 60: 755–61. <https://doi.org/10.1038/jhg.2015.107> PMID: 26377243
35. 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. *Arter Thromb Vasc Biol.* Falk Cardiovascular Research Center, Stanford University Medical School, CA 94305–5246, USA; 1995; 15: 58–64. Available: <http://www.ncbi.nlm.nih.gov/pubmed/7749816>
36. Kraft HG, Windegger M, Menzel HJ, Utermann G. Significant impact of the +93 C/T polymorphism in the apolipoprotein(a) gene on Lp(a) concentrations in Africans but not in Caucasians: confounding effect of linkage disequilibrium. *Hum Mol Genet.* 1998; 7: 257–64. <https://doi.org/10.1093/hmg/7.2.257> PMID: 9425232
37. Wade DP, Lindahl GE, Lawn RM. Apolipoprotein(a) gene transcription is regulated by liver-enriched trans-acting factor hepatocyte nuclear factor 1 alpha. *J Biol Chem.* Falk Cardiovascular Research Center, Stanford University Medical School, California 94305; 1994; 269: 19757–65. Available: <http://www.jbc.org/content/269/31/19757.long> PMID: 8051057

38. Mooser V, Mancini FP, Bopp S, Pethö-Schramm A, Guerra R, Boerwinkle E, et al. Sequence polymorphisms in the apo(a) gene associated with specific levels of Lp(a) in plasma. *Hum Mol Genet.* 1995; 4: 173–181. <https://doi.org/10.1093/hmg/4.2.173> PMID: 7757064
39. Trommsdorff M, Köchl S, Lingenhel A, Kronenberg F, Delpont R, Vermaak H, et al. 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. Institute of Medical Biology and Human Genetics, Innsbruck, Austria;* 1995; 96: 150–7. <https://doi.org/10.1172/JCI118015> PMID: 7615785
40. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature.* 2012; 489: 57–74. <https://doi.org/10.1038/nature11247> PMID: 22955616
41. Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet.* 2012; 13: 484–492. <https://doi.org/10.1038/nrg3230> PMID: 22641018
42. Schultz MD, He Y, Whitaker JW, Hariharan M, Mukamel EA, Leung D, et al. Human body epigenome maps reveal noncanonical DNA methylation variation. *Nature.* 2015; 523: 212–216. <https://doi.org/10.1038/nature14465> PMID: 26030523
43. Yin Y, Morgunova E, Jolma A, Kaasinen E, Sahu B, Khund-Sayeed S, et al. Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science (80-).* 2017; 356. <https://doi.org/10.1126/science.aaj2239> PMID: 28473536
44. Schalkwyk LC, Meaburn EL, Smith R, Dempster EL, Jeffries AR, Davies MN, et al. Allelic Skewing of DNA Methylation Is Widespread across the Genome. *Am J Hum Genet. MRC SGDP Research Centre, Institute of Psychiatry, King's College London, London SE5 8AF, UK;* 2010; 86: 196–212. Available: <https://doi.org/10.1016/j.ajhg.2010.01.014> PMID: 20159110
45. Shoemaker R, Deng J, Wang W, Zhang K. Allele-specific methylation is prevalent and is contributed by CpG-SNPs in the human genome. *Genome Res.* 2010; 20: 883–9. <https://doi.org/10.1101/gr.104695.109> PMID: 20418490
46. John G, Hegarty JP, Yu W, Berg A, Pastor DM, Kelly AA, et al. NKX2-3 variant rs11190140 is associated with IBD and alters binding of NFAT. *Mol Genet Metab.* 2011; <https://doi.org/10.1016/j.ymgme.2011.06.023> PMID: 21803625
47. Reynard LN, Bui C, Syddall CM, Loughlin J. CpG methylation regulates allelic expression of GDF5 by modulating binding of SP1 and SP3 repressor proteins to the osteoarthritis susceptibility SNP rs143383. *Hum Genet.* 2014; 133: 1059–73. <https://doi.org/10.1007/s00439-014-1447-z> PMID: 24861163
48. Kumar D, Puan KJ, Andiappan AK, Lee B, Westerlaken GHA, Haase D, et al. A functional SNP associated with atopic dermatitis controls cell type-specific methylation of the VSTM1 gene locus. *Genome Med.* 2017; 9: 18. <https://doi.org/10.1186/s13073-017-0404-6> PMID: 28219444
49. de Toro-Martín J, Guénard F, Tchernof A, Deshaies Y, Pérusse L, Biron S, et al. A CpG-SNP Located within the ARPC3 Gene Promoter Is Associated with Hypertriglyceridemia in Severely Obese Patients. *Ann Nutr Metab.* 2016; 68: 203–12. <https://doi.org/10.1159/000445358> PMID: 27055012
50. Dayeh TA, Olsson AH, Volkov P, Almgren P, Rönn T, Ling C. Identification of CpG-SNPs associated with type 2 diabetes and differential DNA methylation in human pancreatic islets. *Diabetologia.* 2013; 56: 1036–46. <https://doi.org/10.1007/s00125-012-2815-7> PMID: 23462794
51. Dos Santos C, Bougnères P, Fradin D. A single-nucleotide polymorphism in a methylatable Foxa2 binding site of the G6PC2 promoter is associated with insulin secretion in vivo and increased promoter activity in vitro. *Diabetes.* 2009; 58: 489–92. <https://doi.org/10.2337/db08-0587> PMID: 18984742
52. Mittelstraß K, Waldenberger M. DNA methylation in human lipid metabolism and related diseases. *Curr Opin Lipidol.* 2018; 29: 116–124. <https://doi.org/10.1097/MOL.0000000000000491> PMID: 29517982
53. Lamina C, Haun M, Coassin S, Kloss-Brandstätter A, Gieger C, Peters A, et al. A systematic evaluation of short tandem repeats in lipid candidate genes: riding on the SNP-wave. *Veitia RA, editor. PLoS One.* 2014; 9: e102113. <https://doi.org/10.1371/journal.pone.0102113> PMID: 25050552
54. Mele M, Ferreira PG, Reverter F, DeLuca DS, Monlong J, Sammeth M, et al. The human transcriptome across tissues and individuals. *Science (80-).* 2015; 348: 660–665. <https://doi.org/10.1126/science.aaa0355> PMID: 25954002
55. Matys V, Kel-Margoulis O V, Fricke E, Liebich I, Land S, Barre-Dirrie A, et al. TRANSFAC and its module TRANSCCompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res. BIOBASE GmbH, Halchtersche Strasse 33, D-38304 Wolfenbüttel, Germany. vma@biobase.de;* 2006; 34: D108–D110. <https://doi.org/10.1093/nar/gkj143> PMID: 16381825
56. Graham MJ, Viney N, Croke RM, Tsimikas S. Antisense inhibition of apolipoprotein (a) to lower plasma lipoprotein (a) levels in humans. *J Lipid Res.* 2016; 57: 340–51. <https://doi.org/10.1194/jlr.R052258> PMID: 26538546

57. Macchi C, Sirtori CR, Corsini A, Santos RD, Watts GF, Ruscica M. A new dawn for managing dyslipidemias: The era of rna-based therapies. *Pharmacol Res.* 2019; 150: 104413. <https://doi.org/10.1016/j.phrs.2019.104413> PMID: 31449975
58. Burgess S, Ference BA, Staley JR, Freitag DF, Mason AM, Nielsen SF, et al. Association of LPA Variants With Risk of Coronary Disease and the Implications for Lipoprotein(a)-Lowering Therapies: A Mendelian Randomization Analysis. *JAMA Cardiol.* 2018; 3: 619–627. <https://doi.org/10.1001/jamacardio.2018.1470> PMID: 29926099
59. Lamina C, Kronenberg F, Lp(a)-GWAS-Consortium. Estimation of the Required Lipoprotein(a)-Lowering Therapeutic Effect Size for Reduction in Coronary Heart Disease Outcomes. *JAMA Cardiol.* 2019; <https://doi.org/10.1001/jamacardio.2019.1041> PMID: 31017618
60. Madsen CM, Kamstrup PR, Langsted A, Varbo A, Nordestgaard BG. Lp(a) (Lipoprotein[a])-Lowering by 50 mg/dL (105 nmol/L) May Be Needed to Reduce Cardiovascular Disease 20% in Secondary Prevention: A Population-Based Study. *Arterioscler Thromb Vasc Biol.* 2019; ATVBaha119312951. <https://doi.org/10.1161/ATVBaha.119.312951> PMID: 31578080
61. Kronenberg F. Therapeutic lowering of lipoprotein(a): How much is enough? *Atherosclerosis.* 2019; <https://doi.org/10.1016/j.atherosclerosis.2019.07.003> PMID: 31327477
62. Tsimikas S. A Test in Context: Lipoprotein(a): Diagnosis, Prognosis, Controversies, and Emerging Therapies. *J Am Coll Cardiol.* Elsevier; 2017; 69: 692–711. <https://doi.org/10.1016/j.jacc.2016.11.042> PMID: 28183512
63. Erhart G, Lamina C, Lehtimäki T, Marques-Vidal P, Kähönen M, Vollenweider P, et al. Genetic Factors Explain a Major Fraction of the 50% Lower Lipoprotein(a) Concentrations in Finns. *Arterioscler Thromb Vasc Biol.* 2018; 38: 1230–1241. <https://doi.org/10.1161/ATVBaha.118.310865> PMID: 29567679
64. Puckey LH, Lawn RM, Knight BL. Polymorphisms in the apolipoprotein(a) gene and their relationship to allele size and plasma lipoprotein(a) concentration. *Hum Mol Genet.* 1997; 6: 1099–107. Available: <https://doi.org/10.1093/hmg/6.7.1099> PMID: 9215681
65. Moriarty PM, Varvel SA, Gordts PLSM, McConnell JP, Tsimikas S. Lipoprotein(a) Mass Levels Increase Significantly According to APOE Genotype: An Analysis of 431 239 Patients. *Arter Thromb Vasc Biol.* 2017; 37: 580–588. <https://doi.org/10.1161/ATVBaha.116.308704> PMID: 28062489
66. Vogel C, Marcotte EM. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat Rev Genet.* 2012; 13: 227–232. <https://doi.org/10.1038/nrg3185> PMID: 22411467
67. Taliun D, Harris DN, Kessler MD, Carlson J, Szpiech ZA, Torres R, et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *bioRxiv.* 2019; 563866. <https://doi.org/10.1101/563866>
68. Ma B, Wilker EH, Willis-Owen SAG, Byun H-M, Wong KCC, Motta V, et al. Predicting DNA methylation level across human tissues. *Nucleic Acids Res.* 2014; 42: 3515–28. <https://doi.org/10.1093/nar/gkt1380> PMID: 24445802
69. Fan S, Zhang X. CpG island methylation pattern in different human tissues and its correlation with gene expression. *Biochem Biophys Res Commun.* 2009; 383: 421–5. <https://doi.org/10.1016/j.bbrc.2009.04.023> PMID: 19364493
70. Lanktree MB, Anand SS, Yusuf S, Hegele RA. Comprehensive analysis of genomic variation in the LPA locus and its relationship to plasma lipoprotein(a) in south asians, chinese, and european caucasians. *Circ Cardiovasc Genet.* 2010; 3: 39–46. <https://doi.org/10.1161/CIRCGENETICS.109.907642> PMID: 20160194
71. Schmidt K, Noureen A, Kronenberg F, Utermann G. Structure, function, and genetics of lipoprotein (a). *J Lipid Res.* 2016; 57: 1339–1359. <https://doi.org/10.1194/jlr.R067314> PMID: 27074913
72. GTEx Consortium, Laboratory DA & Coordinating C (LDACC)—Analysis WG, Statistical Methods groups—Analysis Working Group, Enhancing GTEx (eGTEx) groups, NIH Common Fund, NIH/NCI, et al. Genetic effects on gene expression across human tissues. *Nature.* 2017; 550: 204–213. <https://doi.org/10.1038/nature24277> PMID: 29022597
73. White AL, Rainwater DL, Hixson JE, Estlack LE, Lanford RE. Intracellular processing of apo(a) in primary baboon hepatocytes. *Chem Phys Lipids.* 1994; 67–68: 123–33. Available: <http://www.ncbi.nlm.nih.gov/pubmed/8187206> [https://doi.org/10.1016/0009-3084\(94\)90131-7](https://doi.org/10.1016/0009-3084(94)90131-7) PMID: 8187206
74. Holle R, Happich M, Löwel H, Wichmann HE, MONICA/KORA Study Group. KORA—a research platform for population based health research. *Gesundheitswesen.* 2005; 67 Suppl 1: S19–25. <https://doi.org/10.1055/s-2005-858235> PMID: 16032513
75. Zeilinger S, Kühnel B, Klopp N, Baurecht H, Kleinschmidt A, Gieger C, et al. Tobacco smoking leads to extensive genome-wide changes in DNA methylation. *PLoS One.* 2013; 8: e63812. <https://doi.org/10.1371/journal.pone.0063812> PMID: 23691101
76. Chen LS, Storey JD. Eigen-R2 for dissecting variation in high-dimensional studies. *Bioinformatics.* 2008; 24: 2260–2. <https://doi.org/10.1093/bioinformatics/btn411> PMID: 18718946

77. Lehne B, Drong AW, Loh M, Zhang W, Scott WR, Tan S-T, et al. A coherent approach for analysis of the Illumina HumanMethylation450 BeadChip improves data quality and performance in epigenome-wide association studies. *Genome Biol.* 2015; 16: 37. <https://doi.org/10.1186/s13059-015-0600-x> PMID: 25853392
78. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics.* 2014; 30: 1363–9. <https://doi.org/10.1093/bioinformatics/btu049> PMID: 24478339
79. Smyth GK. limma: Linear Models for Microarray Data. In: Gentleman R, Carey VJ, Huber W, Irizarry RA, Dudoit S, editors. *Bioinformatics and Computational Biology Solutions Using R and Bioconductor.* New York, NY: Springer New York; 2005. pp. 397–420. https://doi.org/10.1007/0-387-29362-0_23
80. Touleimat N, Tost J. Complete pipeline for Infinium(®) Human Methylation 450K BeadChip data processing using subset quantile normalization for accurate DNA methylation estimation. *Epigenomics.* 2012; 4: 325–41. <https://doi.org/10.2217/epi.12.21> PMID: 22690668
81. Pfeiffer L, Wahl S, Pilling LC, Reischl E, Sandling JK, Kunze S, et al. DNA Methylation of Lipid-Related Genes Affects Blood Lipid Levels. *Circ Cardiovasc Genet.* 2015; 8: 334–342. <https://doi.org/10.1161/CIRCGENETICS.114.000804> PMID: 25583993
82. Kuhn RM, Karolchik D, Zweig AS, Wang T, Smith KE, Rosenbloom KR, et al. The UCSC Genome Browser Database: update 2009. *Nucleic Acids Res. Center for Biomolecular Science and Engineering, School of Engineering, University of California Santa Cruz, Santa Cruz, CA 95064, USA.* rhead@soe.ucsc.edu; 2009; 37: D613–D619. Available: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2686463>
83. Van Oosterhout C, Hutchinson, William F, Willis DPM, Shipley P. micro-checker: software for identifying and correcting genotyping errors in microsatellite data. *Mol Ecol Notes.* 2004; 4: 535–538. <https://doi.org/10.1111/j.1471-8286.2004.00684.x>
84. Excoffier L, Laval G, Schneider S. Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evol Bioinform Online.* 2005; 1: 47–50. <https://doi.org/10.1111/j.1755-0998.2010.02847.x> PMID: 21565059
85. Kronenberg F, Lobentanz EM, König 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; 35: 1318–28. Available: <http://www.ncbi.nlm.nih.gov/pubmed/7964193> PMID: 7964193
86. Dieplinger H, Gruber G, Krasznai K, Reschauer S, Seidel C, Burns G, et al. Kringle 4 of human apolipoprotein[a] shares a linear antigenic site with human catalase. *J Lipid Res. Institute of Medical Biology and Human Genetics, University of Innsbruck, Austria;* 1995; 36: 813–22. Available: <http://www.ncbi.nlm.nih.gov/pubmed/7542309> PMID: 7542309
87. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics.* 2012; 13: 86. <https://doi.org/10.1186/1471-2105-13-86> PMID: 22568884
88. Dekkers KF, van Iterson M, Sliker RC, Moed MH, Bonder MJ, van Galen M, et al. Blood lipids influence DNA methylation in circulating cells. *Genome Biol.* 2016; 17: 138. <https://doi.org/10.1186/s13059-016-1000-6> PMID: 27350042
89. GTEx Consortium Gte. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science.* 2015; 348: 648–60. <https://doi.org/10.1126/science.1262110> PMID: 25954001