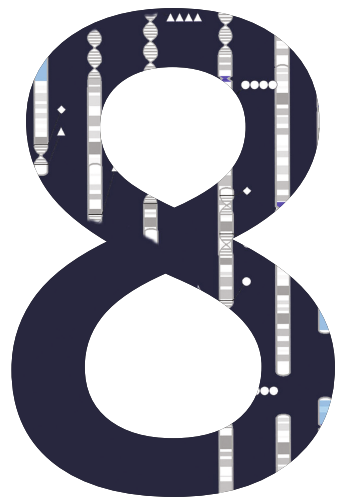
OMICS profiling of cardiometabolic diseases

Li, R.

Cholesteryl Ester Transfer Protein (CETP) Concentration: a Genome-wide Association Study followed by Mendelian Randomization on Coronary Artery Disease



Lisanne L. Blauw
Ruifang Li-Gao
Raymond Noordam
Renée de Mutsert
Stella Trompet
Jimmy F.P. Berbée
Yanan Wang
Jan B. van Klinken
Tim Christen
Diana van Heemst
Dennis O. Mook-Kanamori
Frits R. Rosendaal
J. Wouter Jukema
Patrick C.N. Rensen
Ko Willems van Dijki

Circ Genom Precis Med. 2018 May;11(5):e002034

ABSTRACT

Background—We aimed to identify independent genetic determinants of circulating cholesteryl ester transfer protein (CETP) to assess causal effects of variation in CETP concentration on circulating lipid concentrations and cardiovascular disease risk.

Methods and Results—A genome-wide association (GWA) discovery and replication study on serum CETP concentration were embedded in the Netherlands Epidemiology of Obesity (NEO) study. Based on the independent identified variants, Mendelian randomization was conducted on serum lipids (NEO study) and coronary artery disease (CAD) (CARDIoGRAMplusC4D consortium). In the discovery analysis (N=4,248), we identified three independent variants ($P<5\times10^{-8}$) that determine CETP concentration. These SNPs were mapped to *CETP*, and replicated in a separate subpopulation (N=1,458). Per-allele increase (SE) in serum CETP was 0.32 (0.02) µg/mL for rs247616-C, 0.35 (0.02) µg/mL for rs12720922-A, and 0.12 (0.02) µg/mL for rs1968905-G. Combined, these three variants explained 16.4% of the total variation in CETP concentration. One µg/mL increase in genetically-determined CETP concentration strongly decreased high-density lipoprotein (HDL) cholesterol (-0.23 mmol/L; 95% CI -0.26, -0.20), moderately increased low density lipoprotein (LDL) cholesterol (0.08 mmol/L; 0.00, 0.16), and was associated with an odds ratio of 1.08 (0.94, 1.23) for CAD risk.

Conclusions—This is the first GWAS study identifying independent variants that largely determine CETP concentration. While HDL-C is not a causal risk factor for CAD, it has been unequivocally demonstrated that LDL-C lowering is proportionally associated with a lower CAD risk. Therefore, the results of our study are fully consistent with the notion that CETP concentration is causally associated with CAD through LDL-C.

Keywords—CETP, Coronary artery disease, GWAS, Mendelian randomization

Non-standard abbreviations and acronyms

ApoB, apolipoprotein B; BMI, body mass index; C, cholesterol; CAD, coronary artery disease; CETP, cholesteryl ester transfer protein; CVD, cardiovascular disease; ELISA, enzyme-linked immune sorbent assay; eQTL, expression quantitative trait loci; GCTA, genome-wide complex trait analysis; GLGC, Global Lipids Genetics Consortium; GRS, genetic risk score; GTEx, genotype-tissue expression; GWAS, genome-wide association study; HDL, high-density lipoprotein; NEO, Netherlands Epidemiology of Obesity; PC, principal component; SNP, single nucleotide polymorphism; (V)LDL (very-) low-density lipoproteins

1. INTRODUCTION

Cholesteryl ester transfer protein (CETP) facilitates the net flux of cholesteryl esters from high-density lipoproteins (HDL) towards (very-) low-density lipoproteins ((V)LDL), coupled to a net flux of triglycerides from (V)LDL to HDL.[1] As such, CETP contributes to an atherogenic lipoprotein profile (i.e. high LDL-cholesterol/HDL-cholesterol ratio), as has been extensively studied in both humans and in mice transgenic for human CETP.[2, 3] Therefore, inhibition of CETP has long been regarded a promising therapeutic strategy to attenuate dyslipidaemia and ultimately prevent the development of cardiovascular disease (CVD).

Of the four clinical trials that have studied the effects of pharmacological CETP inhibition on CVD risk reduction, only the fourth and most recent REVEAL trial with anacetrapib did meet its primary endpoint, i.e. a reduction in major coronary events.[4] Contrary to expectations, the clinical trials with the CETP the inhibitors torcetrapib, dalcetrapib, and evacetrapib, were terminated: torcetrapib had off-target effects on blood pressure and caused an increase in cardiovascular events,[5] and both dalcetrapib and evacetrapib lacked efficacy in reducing cardiovascular events on top of statin therapy.[6, 7] All of these CETP inhibitors caused a large increase in HDL-C, and a low to moderate decrease in LDL-C.[5-8] Although high HDL-C concentration is associated with a decreased risk of CVD in epidemiological studies,[9] Voight and colleagues[10] showed in a Mendelian randomization study that genetically-determined higher HDL-C concentrations do not decrease the risk of myocardial infarction, indicating that the association between HDL-C and CVD is not causal. This may be one of the explanations for the lack of efficacy of the three initial CETP inhibitors. Although the underlying reason for success of the fourth CETP inhibitor is not yet elucidated, anacetrapib showed the largest reduction in LDL-C concentration compared to the three initial CETP inhibitors,[4] which may possibly explain its beneficial effects on CVD risk reduction.

Recent evidence shows that serum CETP is largely derived from hepatic macrophages[11], but the genetic basis of serum the CETP concentration in the general population remains to be elucidated. A large genome-wide association study (GWAS) on circulating CETP has not been performed to date. With identification of the main genetic determinants of circulating CETP, the causal effects of variation in serum CETP concentration on circulating lipid concentrations and CVD risk can be assessed using Mendelian randomization. In the past, several Mendelian randomization studies with a comparable aim have been performed,[12-14] including a recent meta-Mendelian randomization analysis by Ference *et al.* with data from over 100,000 participants.[15] However, these studies used candidate SNPs rather than GWAS-identified SNPs, which may be less powerful genetic instruments for assessing the causal role of CETP concentration in cardiovascular disease.[16, 17]

With the present study, we aim to identify independent genetic variants that determine circulating CETP concentration, using a genome-wide rather than a candidate gene approach. In addition, we aim to use these variants as genetic instruments in Mendelian randomization to assess the causal effects of variation in CETP on serum lipids and coronary artery disease (CAD) risk, which may assist in understanding the effectiveness of pharmaceutical CETP inhibition. To this end, we performed a GWAS on serum CETP concentration, using a discovery cohort (n=4,248) and a separate replication cohort (n=1,458) from the Netherlands Epidemiology of Obesity (NEO) study. Subsequently, we used the identified SNPs in Mendelian randomization analyses on serum lipid concentrations in the NEO study population and the Global Lipids Genetics Consortium (GLGC)[18], and on CAD using publically-available data from the CARDGIoGRAMplusC4D consortium.[19]

2. METHODS

2.1 Study design and population

The NEO study is a population-based prospective cohort study of men and women aged between 45 and 65 years. From the greater area of Leiden, The Netherlands, all inhabitants with a self-reported body mass index (BMI) of 27 kg/m^2 or higher were eligible to participate. In addition, inhabitants from one nearby municipality (Leiderdorp, The Netherlands) in the same age group were invited to participate regardless of their BMI, forming a reference population for BMI distribution. In total, 6,671 participants were included from September 2008 until September 2012. Participants visited the NEO study center for extensive physical examination. Venous blood samples were obtained from the antecubital vein after a 10 hour overnight fast. Research nurses recorded current medication use by means of a medication inventory. Prior to the study visit, participants completed questionnaires at home with respect to demographic, lifestyle, and clinical information.

The NEO study was approved by the medical ethics committee of the Leiden University Medical Center (LUMC), and all participants gave their written informed consent. Detailed information about the study design and data collection has been described elsewhere.[20]

Methods used for genotyping and biochemical analyses are described in detail in the Supplemental material.

Due to consent issues, the individual data of NEO study participants will not be made available to other researchers for purposes of reproducing the results or replicating the procedure. However, a file including GWAS summary statistics can be requested via k.willems_van_dijk@lumc.nl.

2.2 Genome-wide association study

We conducted the GWAS on the unstandardized serum CETP concentration for all autosomal chromosomes. We divided the total NEO study population based on the graphical area of recruitment into a discovery cohort (n=4,248; greater area of Leiden, The Netherlands) and a replication cohort (n=1,458; Leiderdorp, The Netherlands). This was considered to be valid as we recently showed, using data of the NEO study population, that serum CETP concentration was not associated with BMI nor with other measures of body fat.[21]

Additive (per-allele) linear regression analyses were conducted separately for the discovery and replication cohort in SNPTEST v2, adjusted for age, sex and the first four PCs. To identify variants that were independently associated with serum CETP concentration, we used conditional and joint analyses to perform a stepwise selection procedure using the genome-wide complex trait analysis (GCTA) tool version 1.24.4.[22] A conditioned P-value $<5\times10^{-8}$ was considered to be genome-wide significant, and a conditioned P-value $<1\times10^{-6}$ was considered a suggestive signal. Independent single nucleotide polymorphisms (SNPs) with a conditioned P-value $<1\times10^{-6}$ in the discovery analysis were validated in the replication sample. SNPs with a P-value <0.05 in the replication cohort were considered to be replicated. Upon identification of the lead SNPs, we determined whether the distribution of the coding alleles was similar in users and non-users of lipid-lowering drugs. More detailed information on this method is described in the Supplemental material.

As Taq1B (rs708272)[23] and -629C>A (rs1800775)[24] are the most studied variants in the *CETP* gene in literature, we specifically checked the GWAS results for their association with CETP concentration. In addition, we reported the linkage disequilibrium (NEO study) of the lead SNPs from the present GWAS with the eight *CETP* SNPs that were used in a recent meta-Mendelian randomization study on coronary heart disease and serum lipids by Ference *et al.*[15] (i.e. rs3764261, rs1800775, rs1864163, rs9929488, rs9989419, rs12708967, rs289714 and rs5880). This allows comparison between our genetic instrument composed of SNPs identified form a GWAS on serum CETP concentration and their genetic instrument composed of candidate SNPs.

The explained variance in serum CETP concentration for the independent variants was estimated in the replication cohort. For each individual SNP the explained variance was estimated as the partial R^2 from the linear regression model with the SNP as independent variable and serum CETP concentration as dependent variable. To estimate the total variance explained by all independent lead SNPs, a weighted genetic risk score (GRS) was calculated per individual. The GRS was constructed as the sum of the number of risk alleles on the lead SNPs weighted by their effect size on CETP concentration in the discovery cohort. The combined explained variance was the partial

R^2 from the linear regression model, with the weighted GRS as independent variable and serum CETP concentration as dependent variable.

To quantify the genome-wide cumulative effects of independent variants influencing various phenotypes, genetic correlations of serum CETP concentration with serum lipid concentrations (i.e. HDL-C, LDL-C, triglycerides and total cholesterol), and BMI were calculated (Supplemental material).

2.3 Expression quantitative trait loci (eQTLs) analysis

To investigate whether the identified lead SNPs could explain serum CETP concentration via transcriptional gene regulation, we checked if these SNPs were eQTLs for *CETP* using data from the genotype-tissue expression (GTEx) project portal (V7)[25] and the Blood eQTL browser.[26]

2.4 Mendelian randomization

Based on the identified independent and replicated SNPs for serum CETP concentration in our study population, we conducted Mendelian randomization analyses on serum lipid concentrations in the NEO study population, and on the risk of CAD based on publically-available summary statistics data from the CARDGIoGRAMplusC4D 1000 Genomes study.[19] A detailed description of the Mendelian randomization analyses on CAD risk using data from the CARDIoGRAMplusC4D consortium can be found in the Supplemental material. Effect estimates for CAD risk were reported as odds ratio with corresponding 95% CI. We used a publically available tool[27] to conduct a power analysis for the Mendelian randomization analysis on CAD, which was based on the findings from the GWAS on serum CETP concentration, the explained variance of the SNPs used to compose the GRS, and the sample size of the CARDIoGRAMplusC4D 1000 Genomes study (60,801 cases; 123,504 controls).

In the NEO study population, we calculated the individual weighted GRS based on the identified SNPs and determined the effect of 1 μg/mL increase in genetically-determined CETP concentration on the concentrations of total cholesterol, HDL-C, triglycerides, LDL-C, total cholesterol/HDL-C ratio, LDL-C/HDL-C ratio and apolipoprotein B (ApoB), using linear regression analysis adjusted for age and sex. Beta coefficients and 95% CIs were reported. We reported total cholesterol, triglycerides, HDL-C and LDL-C concentrations in mmol/L. We report total cholesterol, triglycerides, HDL-C and LDL-C concentrations in mmol/L. To yield concentrations in mg/dL, cholesterol values should be multiplied by 38.67 and triglyceride values by 88.57.[28]

In addition, for replication purposes, we extracted the independent leads SNPs from publically available datasets of the GLGC.[18] We also extracted the Taq1B (rs708272) and

-629C>A (rs1800775) polymorphisms from the CARDGIoGRAMplusC4D 1000 Genomes and GLGC datasets.[18]

3 RESULTS

3.1 Population characteristics

Characteristics of the discovery and replication cohorts are summarized in Table 1. Compared with the replication cohort, there were fewer women in the discovery cohort (50.6% versus 56.1%). Also, participants in the discovery cohort had a higher BMI (30.3 kg/m^2 versus 25.6 kg/m^2) and more often used lipid-lowering drugs (17.5% versus 10.4%) than participants in the replication cohort. Serum CETP and lipid concentrations were comparable between both cohorts.

TABLE 1 Characteristics of the discovery and replication cohort from the Netherlands Epidemiology of Obesity (NEO) study.

Characteristics	Discovery cohort	Replication cohort
Number of participants	4,248*	1,458†
Age (year)	56 (51, 61)	57 (51, 61)
Women	2,148 (50.6%)	818 (56.1%)
Body mass index (kg/m^2)	30.3 (28.4, 33.0)	25.6 (23.2, 28.2)
Lipid-lowering drug users	745 (17.5%)	151 (10.4%)
Fasting serum concentrations		
CETP (µg/mL)	2.50 (0.67)	2.43 (0.64)
Total cholesterol (mmol/L)	5.66 (1.08)	5.69 (1.07)
LDL-cholesterol (mmol/L)	3.58 (0.99)	3.56 (0.98)
HDL-cholesterol (mmol/L)	1.38 (0.38)	1.58 (0.46)
Triglycerides (mmol/L)	1.34 (0.95, 1.87)	1.00 (0.71, 1.45)

Results are presented as median (inter quartile range) for not normally distributed data, mean (SD) or number (percentage).
* Missing data: n=13 for total cholesterol concentration, n=14 for HDL-cholesterol concentration, triglycerides and LDL-cholesterol concentration
† No missing data

3.2 Genome-wide association analysis

The -Log(P-value) plot for the GWAS is shown in Figure 1. The accompanying list of SNPs that reached genome-wide significance ($P<5\times10^{-8}$) is presented in Supplementary table 1. After conditioning on the lead SNPs, three independent variants reached genome-wide significance (conditioned $P<5\times10^{-8}$; Supplementary Figure 2) and seven suggestive signals were identified (conditioned $P<1\times10^{-6}$) in the discovery cohort (Table 2). The

three genome-wide significant variants were all mapped to the *CETP* gene. Notably, these independent variants were rs247616 ($P=1.86\times10^{-64}$), rs12720922 ($P=6.68\times10^{-13}$) and rs1968905 ($P=1.66\times10^{-12}$), which had a per-allele increase (SE) in serum CETP of 0.32 (0.02) µg/mL (rs247616-C), 0.35 (0.02) µg/mL (rs12720922-A) and 0.12 (0.02) µg/mL (rs1968905-G). These three variants were all replicated in the replication analysis ($P<0.05$). In the NEO study, these variants together explained 16.4% of the serum CETP concentration. The distributions of the effect alleles of the three lead SNPs were similar for individuals taking lipid-lowering drugs and not taking lipid-lowering drugs (Supplementary table 2). A number of SNPs were suggestively associated with serum CETP concentration, including SNPs mapped to *ADAMTS3*, *PPARG* and *LPL*.

The unconditioned per-allele effect size of the well-known Taq1B (rs708272) and -629C>A (rs1800775) variants was 0.27 µg/mL for both SNPs (Supplementary table 1). Of the three lead SNPs, Taq1B was in high linkage disequilibrium with rs247616 (LD=0.55), as was -629C>A (LD=0.51). Taq1B and -629C>A were also in high linkage with each other (LD=0.83).

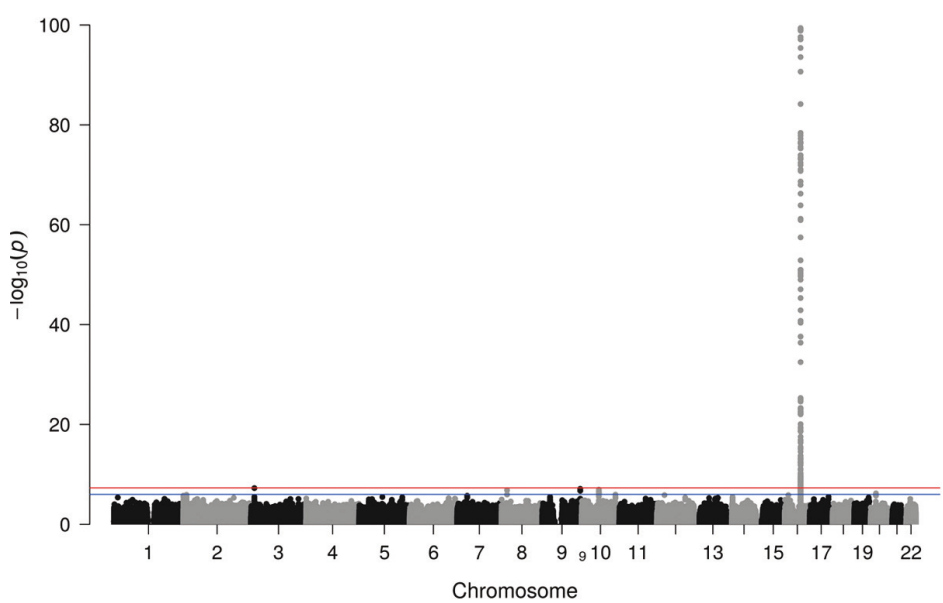


FIGURE 1 −Log(P-value) plot for the genome-wide association study in the discovery cohort (n=4,248).
The *CETP* gene is located on chromosome 16. The red line represents the threshold for genome-wide significance ($P<5\times10^{-8}$). The blue line represents the threshold for suggestive signals ($P<1\times10^{-6}$).

TABLE 2 Summary statistics of the associations of the three independent lead SNPs that reached genome-wide significance and seven suggestive signals with serum CETP concentration.

								Discovery cohort (n=4,248)				Replication cohort (n=1,458)		
Chr	**SNP**	**Position**	**Location**	**Closest gene**	**Coding/ Non-coding allele**	**Coding allele frequency***	**Imputation quality**	**Effect size per allele† (µg/mL)**	**SE**	**P-value**	**Conditioned P-value‡**	**Effect size per allele† (µg/mL)**	**SE**	**P-value**
Lead SNPs														
16	rs247616	56989590	Intergenic	*CETP*	C/T	0.67	1	0.32	0.015	3.98×10^{-100}	1.86×10^{-64}	0.31	0.024	1.24×10^{-37}
16	rs12720922	57000885	Intron	*CETP*	A/G	0.17	0.98	0.35	0.019	3.48×10^{-74}	6.68×10^{-13}	0.36	0.030	3.27×10^{-33}
16	rs1968905	57010948	Intron	*CETP*	G/T	0.82	0.85	0.12	0.02	4.12×10^{-9}	1.66×10^{-12}	0.098	0.031	1.80×10^{-3}
Suggestive SNPs														
2	rs185550357	50249349	Intron	*NRXN1*	C/T	0.007	0.70	0.63	0.12	3.94×10^{-7}	4.57×10^{-7}	0.064	0.15	0.67
3	rs6442310	12358230	Intron	*PPARG*	A/T	0.54	0.92	0.08	0.015	5.52×10^{-8}	6.13×10^{-8}	0.0044	0.024	0.85
8	chr8:19811023:I	19811023	Intron	*LPL*	ATG/A	0.12	0.87	0.12	0.023	1.45×10^{-7}	1.59×10^{-7}	-0.018	0.038	0.63
9	rs3094377	136312119	Intron	*ADAMTS13*	T/C	0.03	0.41	0.35	0.064	7.65×10^{-8}	9.58×10^{-8}	-0.18	0.11	0.098
10	rs12253367	62003462	Intron	*ANK3*	G/A	0.17	0.94	0.10	0.019	1.08×10^{-7}	1.18×10^{-7}	0.0051	0.031	0.87
16	rs117427818	57010486	Intron	*CETP*	T/C	0.050	0.75	0.46	0.038	3.26×10^{-33}	1.02×10^{-7}	0.54	0.061	1.93×10^{-18}
20	rs150904289	18845428	Intergenic		G/A	0.11	0.79	0.13	0.026	6.31×10^{-7}	6.87×10^{-7}	-0.0084	0.042	0.84

Threshold for genome-wide significance is 5×10^{-8}, based on the conditioned P-value‡. Threshold for the suggestive signals is 1×10^{-6}. Threshold for replication is 0.05.
* In the discovery cohort; † Beta coefficient expressed as the difference in serum CETP concentration; ‡ P-value conditioned on the top lead SNPs using step-wise conditional analysis
Chr, chromosome

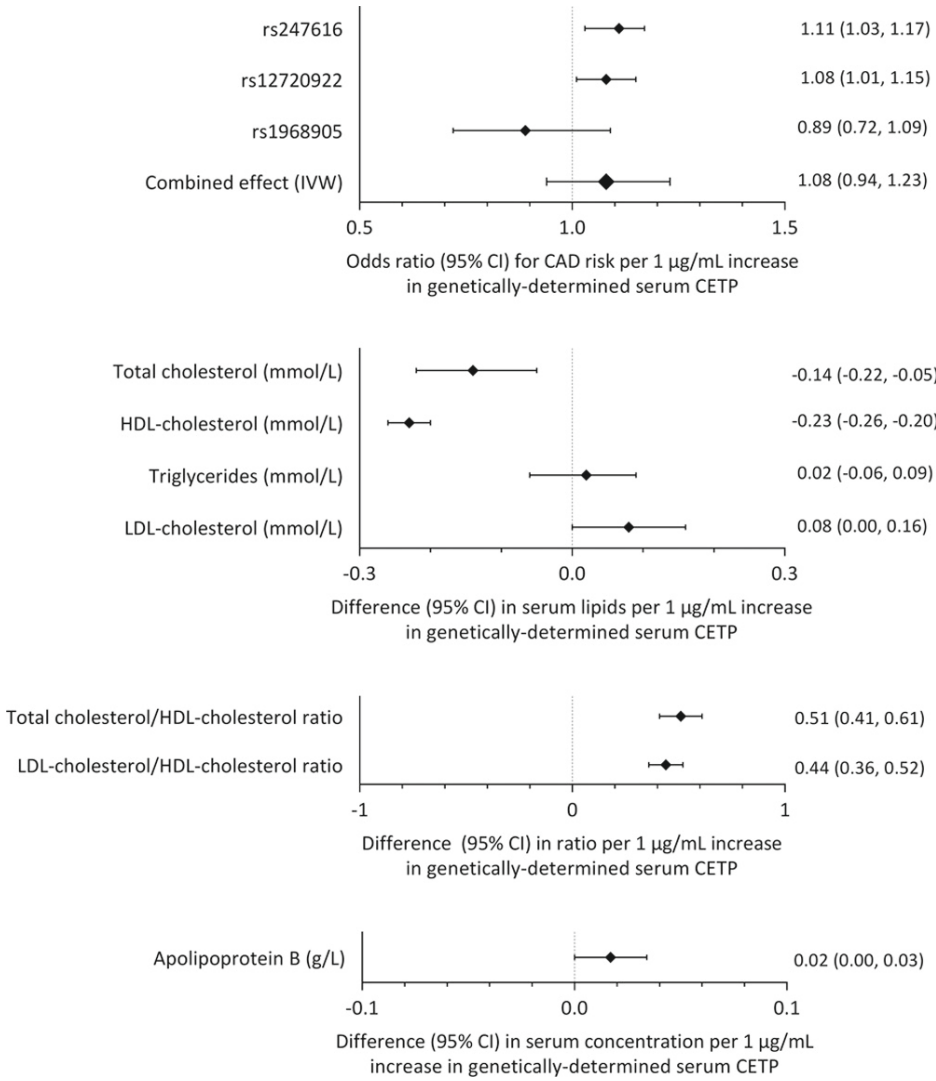


FIGURE 2 Results from the Mendelian randomization study on coronary artery disease in the CARDIoGRAMplusC4D 1000 Genomes Consortium (60,801 cases; 123,504 controls), and on serum lipid and lipoprotein B concentrations in the total Netherlands Epidemiology of Obesity (NEO) study population (n=5,706*).

* Missing data: n=13 for total cholesterol concentration, n=14 for HDL-cholesterol concentration, triglycerides and LDL-cholesterol concentration, n=41 for apolipoprotein B concentration. Results were adjusted for age and sex.

CAD, coronary artery disease; CETP, cholesteryl ester transfer protein; IVW, inverse-variance weighted.

The linkage disequilibrium between the three lead SNPs and eight SNPs that were used as genetic instruments for a *CETP* GRS in a recent meta-Mendelian randomization analysis on coronary heart disease and serum lipids[15] are shown in Supplementary figure 3. The two strongest lead SNPs from the present GWAS, i.e. rs247616 and rs12720922, were in high linkage disequilibrium with the eight candidate SNPs of the GRS that was used in the study of Ference *et al.*,[15] with the highest linkage disequilibrium between rs247616 and rs3764261 (0.996), and between rs12720922 and rs1864163 (0.646).

The genetic correlation of serum CETP concentration with serum lipid concentrations and BMI is reported in Supplementary table 3. The genetic correlation of serum CETP concentration was highest with serum HDL-C concentration (0.17) and serum triglyceride concentration (-0.29), and lowest with serum total cholesterol concentration (-0.020), serum LDL-C concentration (0.074) and BMI (0.032).

TABLE 3 Expression quantitative trait loci (eQTL) for the three GWAS-identified lead SNPs.

SNP	Assessed Allele	Gene	Chr	P-value	Effect size	Tissue	Database
rs247616	C	*NLRC5*	16	9.5×10^{-14}	0.34	Transformed fibroblasts	GTEx
rs247616	C	*CETP*	16	6.3×10^{-10}	0.30	Lung	GTEx
rs247616	C	*CETP*	16	1.1×10^{-7}	0.41	Transverse colon	GTEx
rs247616	C	*CETP*	16	1.4×10^{-6}	0.45	Terminal Ileum	GTEx
rs247616	C	*CETP*	16	3.6×10^{-6}	0.32	Liver	GTEx
rs247616	C	*CETP*	16	7.9×10^{-6}	0.30	Esophagus (mucosa)	GTEx
rs247616	C	*CETP*	16	9.3×10^{-6}	0.41	Pancreas	GTEx
rs247616	C	*BBS2*	16	1.7×10^{-5}	0.37	Cerebellar Hemisphere	GTEx
rs247616	C	*CETP*	16	4.1×10^{-5}	4.10	Whole blood	Blood eQTL browser
rs12720922	A	*NLRC5*	16	1.9×10^{-6}	0.24	Transformed fibroblasts	GTEx
rs1968905	G	-	16	-	-	-	

3.3 eQTL analysis of the lead SNPs

Table 3 shows the eQTLs for the genetic variants rs247616, rs12720922 and rs1968905. The SNP that was most strongly associated with serum CETP concentration in the GWAS, i.e. rs247616, was identified as an eQTL for the *CETP* gene in several tissues (P-value range 6.3×10^{-10} to 4.1×10^{-5}). Rs12720922 was an eQTL for *NLRC5*, but not for *CETP*. However, we found rs1864163, which is in moderate linkage disequilibrium with rs12720922 (LD=0.65), to be an eQTL for *CETP* in whole blood ($P=8.5\times10^{-4}$, effect size A-allele 3.33). The third lead SNP, i.e. rs1968905, was not identified as an eQTL for *CETP* in the studied tissues, neither were any variants in strong linkage disequilibrium with this SNP.

3.4 Mendelian randomization

Figure 2 shows the results from the Mendelian randomization analyses on CAD risk and serum lipid and ApoB concentrations. We had a power of 0.90 to detect an odds ratio of 1.04 with conventional Mendelian randomization analyses (which makes use of a formal weighted genetic risk score), when taking into account an alpha of 0.05, the explained variance of the SNPs that compose the GRS (i.e. 16.4%), and the sample size of the CARDIoGRAMplusC4D 1000Genomes study. Per 1 µg/mL increase in genetically-determined serum CETP concentration the odds ratio for CAD risk was 1.08 (95% CI: 0.94, 1.23). For the lead SNPs separately, odds ratios were 1.11 (95% CI: 1.03, 1.17) for rs247616, 1.08 (95% CI: 1.01, 1.15) for rs12720922, and 0.89 (95% CI: 0.72, 1.09) for rs1968905. For Taq1B and -629C>A, odds ratios for CAD risk were 1.02 (95% CI 1.01, 1.04) and 1.03 (95% CI 1.01, 1.05), respectively (Supplementary table 4).

A 1 µg/mL increase in genetically-determined serum CETP concentration was associated with decreased total cholesterol concentration, i.e. -0.14 (95% CI: -0.22, -0.05) mmol/L, and HDL-C concentration, i.e. -0.23 (95% CI: -0.26, -0.20) mmol/L, while it was associated with increased serum LDL-C concentration, i.e. 0.08 (95% CI: 0.00, 0.16) mmol/L, and ApoB concentration, i.e. 0.02 (95% CI: 0.00, 0.03) g/L. Genetically-determined serum CETP concentration was not associated with serum triglycerides concentration, i.e. 0.02 (95% CI -0.05, 0.09) mmol/L. Supplementary table 5 shows the results from the Mendelian randomization analysis with data from GLGC. The results for total cholesterol, triglycerides and LDL-C concentrations from GLGC were comparable with the results from the NEO study. Effect sizes for HDL-C were larger in GLGC than in the NEO study.

4 DISCUSSION

With this first large GWAS on serum CETP concentration, we identified and replicated three independent SNPs, all mapping to the *CETP* region. These three variants, notably rs12720922, rs247616 and rs1968905, explained 16.4% of the total variation in serum CETP concentration. Effect sizes of all lead SNPs were large, with rs12720922-A having the largest effect on serum CETP: +0.35 µg/mL per additional risk allele. Also, we showed that genetically-determined variation in circulating CETP associates with a stepwise substantial decrease in HDL-C concentration, a moderate increase in LDL-C and ApoB concentration, and a concordant 8% increase in CAD risk.

We found three independent SNPs in the *CETP* region that largely explained CETP concentration. The association of the rs12720922 variant with circulating CETP, blood lipids or risk of CAD has, to the best of our knowledge, never been described before. In the eQTL studies that we considered,[25, 26] rs12720922 was not reported as an eQTL for

CETP. However, rs1864163, which is in moderate linkage disequilibrium with rs12720922 (LD=0.65), was found to be an eQTL for *CETP* in whole blood,[26] although not in the liver.[25] Despite the absence of a direct association between rs12720922 and CETP mRNA levels, it is not ruled out that rs12720922 affects CETP levels via transcriptional regulation. Possibly, we were not able to identify eQTLs of rs12720922 for *CETP* due to the low sample size of eQTL studies in liver tissue, and the dilution that is introduced by considering whole liver expression, since *CETP* is specifically expressed by hepatic macrophages (i.e. Kupffer cells).[29] The second independent lead SNP, rs247616, is located in the promotor region of the *CETP* gene.[30] This SNP has also not been associated with serum CETP concentration before, but it has previously been shown that the minor allele of this variant (rs247616-T) is associated with decreased *CETP* mRNA expression in human liver and increased HDL-C concentrations[30-33], which is in line with our findings.

The third identified variant, rs1968905, was reported by one study to associate with HDL-C concentration specifically in Africans, but has not been linked to serum CETP concentration previously.[34] Of note, rs1968905 is in strong linkage disequilibrium (LD=0.89) with rs1801706 (i.e. G84A), which has been reported as a risk factor for CAD in South Indians.[35] Interestingly, rs1801706 is located in the 3' untranslated region (3' UTR) of the CETP gene, suggesting involvement in posttranscriptional regulation.[36] In addition to these lead SNPs, we found a suggestive signal (i.e. rs117427818) that could be a potential fourth hit in the *CETP* gene. Its statistical significance was, however, largely reduced upon conditional analyses. Indeed, this SNP is in linkage disequilibrium with (one of) the lead SNPs and not completely independently associated with serum CETP concentration (highest linkage disequilibrium with rs12720922; LD=0.20).

To obtain insight in the role of LDL-C and HDL-C in the causal association between serum CETP and CAD risk, we performed Mendelian randomization analyses on serum lipid concentrations. A higher CETP GRS was associated with a large decrease in HDL-C concentration and a moderate increase in LDL-C concentration. The strong association of genetically-determined serum CETP concentration with HDL-C concentration is probably partially explained by a shared genetic background of these two phenotypes, as the genetic correlation between serum CETP concentration and HDL-C concentration was relatively high, which indicates pleiotropy. It should be noted, however, that HDL-C has been observationally, but not causally associated with CVD risk.[10] Thus, although a genetically-determined increase in serum CETP is causally associated with a decrease in HDL-C concentration, this likely does not explain the association between serum CETP and CAD risk. On the other hand, a causal, proportional, log-linear association between LDL-C concentration and CAD risk has been firmly established.[37] In a previously performed Mendelian randomization study using an LDL-C GRS, it was shown that the odds ratio for CAD risk was 1.68 (1.51-1.87) per 1 SD increase in LDL-C (i.e.0.98

mmol/L).[38] To compare, we showed that per 1 µg/mL increase in serum CETP, LDL-C concentration increased with 0.08 mmol/L and the odds ratio for CAD risk was 1.08. Thus, expressed per 0.98 mmol/L increase in LDL-C concentration, we observed 1.98 times increase in CAD risk using the CETP GRS, which is comparable with the effect estimate found with the LDL-C GRS.[38] Taken together, our study suggests that the causal association between CETP concentration and CAD risk may be explained by effects on LDL-C concentration. Interestingly, a recent large meta-Mendelian randomization analysis by Ference *et al.*,[15] indicated that ApoB concentration is an even more important causal link between CETP and CAD risk than LDL-C concentration. This implies that an increase in the absolute number of VLDL, IDL and LDL (i.e. non-HDL) particles, as reflected by ApoB concentration, due to increased circulating CETP may explain the association with CAD risk, rather than the amount of cholesterol in LDL particles.

Our findings are in line with this recent meta-Mendelian randomization analysis, including over a 100,000 participants that showed comparable effects for a CETP GRS on CAD risk, LDL-C and ApoB concentration.[15] In that meta-analysis, a CETP GRS was composed of eight candidate SNPs selected from the *CETP* gene with a forward conditional regression analysis on HDL-C concentration. In the present study, we identified three different *CETP* SNPs that independently determine circulating CETP concentration by using a hypothesis-free approach (i.e. GWAS). These GWAS-identified SNPs are therefore direct genetic instruments to study the causal effects of CETP in Mendelian randomization. Although none of the lead SNPs that we identified with GWAS were included in the GRS composed by Ference *et al.*[15], we observed that the two strongest lead SNPs from our GWAS were in high (rs247616) to moderate (rs12720922) linkage disequilibrium with the eight candidate SNPs of that GRS. This indicates that the GRS composed of candidate *CETP* SNPs is a reliable genetic instrument to study the causal effects of CETP, and our results therefore extend this recent meta-Mendelian randomization analysis.[15]

Although we showed a causal association between CETP concentration and CAD risk, the three initial CETP inhibitors did not reduce the risk of cardiovascular events when given in addition to statin treatment. In fact, clinical trials with those CETP inhibitors were even terminated due to off-target effects (torcetrapib) or a lack of efficacy (dalcetrapib and evacetrapib).[5-7] Dalcetrapib had minimal effects in LDL-C concentration possibly explaining its futility.[6] Evacetrapib did significantly reduce LDL-C concentration, but did not evoke a concordant decrease in ApoB,[39] indicating unfavourable LDL particle remodelling rather than removal from the circulation.[37, 40] This explanation for the failure of the evacetrapib trial is in line with a recent meta-Mendelian randomization study by Ference *et al.*[15] Data from that study indicate that the success of CETP inhibitors when prescribed on top of statin treatment is dependent

on their capability to reduce the absolute number non-HDL particles as reflected by a reduction in ApoB concentration. A reduction in the LDL-C concentration through CETP inhibition may thus only be beneficial when a concordant reduction in ApoB concentration is achieved.[15] Indeed, anacetrapib did show a concordant reduction in non-HDL and ApoB concentration of -18%, which was accompanied by a reduced rate ratio for major coronary events of 0.91 (95% CI 0.85, 0.97).[4]

Our study may have had insufficient statistical power to identify additional variants with small effects on serum CETP concentration or with low allele frequencies. As we did not replicate our GWAS findings in additional heterogeneous populations, caution should be taken when extrapolating the results to other populations. Also, despite the similar associations between the lead SNPs and LDL-C concentrations in GLGC and the NEO study, effect sizes for HDL-C concentrations were higher in GLGC. A possible explanation might involve differences in the composition of the study populations, as GLGC is a meta-analysis of a wide variety of cohorts.

In conclusion, with a GWAS, we identified and replicated three independent SNPs mapping to the *CETP* gene that together explained 16.4% of the total variation in serum CETP concentration, which shows that serum CETP concentration is strongly genetically determined. Using Mendelian randomization, we showed that 1 µg/mL increase in serum CETP causally associates with a large decrease in HDL-C cholesterol of -0.23 mmol/L, moderate increases in LDL-C concentration of 0.08 mmol/L and ApoB concentration of 0.02 g/L, and an odds ratio of 1.08 for CAD risk. While HDL-C is not a causal risk factor for CAD, it has been unequivocally demonstrated that LDL-C lowering is proportionally associated with a lower CAD risk. Therefore, the results of our study are fully consistent with the notion that CETP concentration is causally associated with CAD through LDL-C.

Acknowledgments

We express our gratitude to all individuals who participate in the NEO study. We are grateful to all participating general practitioners for inviting eligible participants. We furthermore thank Pat van Beelen and all research nurses for collecting the data, Petra Noordijk and her team for laboratory management and DNA isolation, and Ingeborg de Jonge for all data management of the NEO study. We sincerely thank Chris van der Bent for performing the serum CETP concentration measurements. The genotyping in the NEO study was supported by the Centre National de Génotypage (Paris, France), headed by Jean-Francois Deleuze. Data on coronary artery disease have been contributed by CARDIoGRAMplusC4D investigators and have been downloaded from www.CARDIOGRAMPLUSC4D.ORG.

Funding sources

The NEO study was supported by the participating Departments, the Division and the Board of Directors of the Leiden University Medical Center, and by the Leiden University, Research Profile Area 'Vascular and Regenerative Medicine'. L.L. Blauw was supported by a grant from the Board of Directors of the Leiden University Medical Center. Y. Wang is supported by the Dutch Science Organization [The Netherlands Organisation for Health Research and Development (ZonMW) VENI Grant 91617027]. D. van Heemst was supported by the European Commission funded project HUMAN [Health-2013-INNOVATION-1-602757]. D. Mook-Kanamori is supported by the Dutch Science Organization [ZonMW VENI Grant 91614023]. P.C.N. Rensen is an Established Investigator of the Dutch Heart Foundation [2009T038].

Disclosures

None

REFERENCES

1. Tall AR. Plasma cholesteryl ester transfer protein. *Journal of lipid research*. 1993;34:1255-1274
2. Chapman MJ, Le Goff W, Guerin M, Kontush A. Cholesteryl ester transfer protein: At the heart of the action of lipid-modulating therapy with statins, fibrates, niacin, and cholesteryl ester transfer protein inhibitors. *European heart journal*. 2010;31:149-164
3. Westerterp M, van der Hoogt CC, de Haan W, Offerman EH, Dallinga-Thie GM, Jukema JW, Havekes LM, Rensen PC. Cholesteryl ester transfer protein decreases high-density lipoprotein and severely aggravates atherosclerosis in apoe*3-leiden mice. *Arteriosclerosis, thrombosis, and vascular biology*. 2006;26:2552-2559
4. Bowman L, Hopewell JC, Chen F, Wallendszus K, Stevens W, Collins R, et al. Effects of anacetrapib in patients with atherosclerotic vascular disease. *The New England journal of medicine*. 2017;377:1217-1227
5. Barter PJ, Caulfield M, Eriksson M, Grundy SM, Kastelein JJ, Komajda M, et al. Effects of torcetrapib in patients at high risk for coronary events. *The New England journal of medicine*. 2007;357:2109-2122
6. Schwartz GG, Olsson AG, Abt M, Ballantyne CM, Barter PJ, Brumm J, et al. Effects of dalcetrapib in patients with a recent acute coronary syndrome. *The New England journal of medicine*. 2012;367:2089-2099
7. Lincoff AM, Nicholls SJ, Riesmeyer JS, Barter PJ, Brewer HB, Fox KAA, et al. Evacetrapib and cardiovascular outcomes in high-risk vascular disease. *The New England journal of medicine*. 2017;376:1933-1942
8. Eyvazian VA, Frishman WH. Evacetrapib: Another cetp inhibitor for dyslipidemia with no clinical benefit. *Cardiology in review*. 2017;25:43-52
9. Gordon DJ, Probstfield JL, Garrison RJ, Neaton JD, Castelli WP, Knoke JD, et al. High-density lipoprotein cholesterol and cardiovascular disease. Four prospective american studies. *Circulation*. 1989;79:8-15
10. Voight BF, Peloso GM, Orho-Melander M, Frikke-Schmidt R, Barbalic M, Jensen MK, et al. Plasma hdl cholesterol and risk of myocardial infarction: A mendelian randomisation study. *Lancet (London, England)*. 2012;380:572-580
11. Wang Y, van der Tuin S, Tjeerdema N, van Dam AD, Rensen SS, Hendrikx T, et al. Plasma cholesteryl ester transfer protein is predominantly derived from kupffer cells. *Hepatology (Baltimore, Md.)*. 2015;62:1710-1722
12. Thompson A, Di Angelantonio E, Sarwar N, Erqou S, Saleheen D, Dullaart RP, et al. Association of cholesteryl ester transfer protein genotypes with cetp mass and activity, lipid levels, and coronary risk. *Jama*. 2008;299:2777-2788
13. Niu W, Qi Y. Circulating cholesteryl ester transfer protein and coronary heart disease: Mendelian randomization meta-analysis. *Circulation. Cardiovascular genetics*. 2015;8:114-121
14. Johannsen TH, Frikke-Schmidt R, Schou J, Nordestgaard BG, Tybjaerg-Hansen A. Genetic inhibition of cetp, ischemic vascular disease and mortality, and possible adverse effects. *Journal of the American College of Cardiology*. 2012;60:2041-2048
15. Ference BA, Kastelein JJP, Ginsberg HN, Chapman MJ, Nicholls SJ, Ray KK, et al. Association of genetic variants related to cetp inhibitors and statins with lipoprotein levels and cardiovascular risk. *Jama*. 2017;318:947-956
16. Thompson JF, Lira ME, Durham LK, Clark RW, Bamberger MJ, Milos PM. Polymorphisms in the cetp gene and association with cetp mass and hdl levels. *Atherosclerosis*. 2003;167:195-204
17. Thompson JF, Wood LS, Pickering EH, Dechairo B, Hyde CL. High-density genotyping and functional snp localization in the cetp gene. *Journal of lipid research*. 2007;48:434-443
18. Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, et al. Discovery and refinement of loci associated with lipid levels. *Nature genetics*. 2013;45:1274-1283
19. Nikpay M, Goel A, Won HH, Hall LM, Willenborg C, Kanoni S, et al. A comprehensive 1,000 genomes-based genome-wide association meta-analysis of coronary artery disease. *Nature genetics*. 2015;47:1121-1130

20. de Mutsert R, den Heijer M, Rabelink TJ, Smit JW, Romijn JA, Jukema JW, et al. The netherlands epidemiology of obesity (neo) study: Study design and data collection. *European journal of epidemiology*. 2013;28:513-523
21. Blauw LL, de Mutsert R, Lamb HJ, de Roos A, Rosendaal FR, Jukema JW, et al. Serum cetp concentration is not associated with measures of body fat: The neo study. *Atherosclerosis*. 2016;246:267-273
22. Yang J, Lee SH, Goddard ME, Visscher PM. Gcta: A tool for genome-wide complex trait analysis. *American journal of human genetics*. 2011;88:76-82
23. Drayna D, Lawn R. Multiple rflps at the human cholesteryl ester transfer protein (cetp) locus. *Nucleic acids research*. 1987;15:4698
24. Dachet C, Poirier O, Cambien F, Chapman J, Rouis M. New functional promoter polymorphism, cetp/−629, in cholesteryl ester transfer protein (cetp) gene related to cetp mass and high density lipoprotein cholesterol levels. *Role of Sp1/Sp3 in Transcriptional Regulation*. 2000;20:507-515
25. The GC. The genotype-tissue expression (gtex) project. *Nature genetics*. 2013;45:580-585
26. Westra HJ, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J, et al. Systematic identification of trans eqtls as putative drivers of known disease associations. *Nature genetics*. 2013;45:1238-1243
27. Brion MJ, Shakhbazov K, Visscher PM. Calculating statistical power in mendelian randomization studies. *International journal of epidemiology*. 2013;42:1497-1501
28. Rugge B, Balshem H, Sehgal R, al. e. Screening and treatment of subclinical hypothyroidism or hyperthyroidism [internet]. Rockville (md): Agency for healthcare research and quality (us); (comparative effectiveness reviews, no. 24.) available from: Https://www.Ncbi.Nlm.Nih.Gov/books/nbk83496/. 2011 Oct
29. Wang Y, Tuin Svd, Tjeerdema N, Dam ADv, Rensen SS, Hendrikx T, Berbée JFP, Atanasovska B, Fu J, Hoekstra M, Bekkering S, Riksen NP, Buurman WA, Greve JW, Hofker MH, Shiri-Sverdlov R, Meijer OC, Smit JWA, Havekes LM, Dijk KWv, Rensen PCN. Plasma cholesteryl ester transfer protein is predominantly derived from kupffer cells. *Hepatology*. 2015;62:1710-1722
30. Papp AC, Pinsonneault JK, Wang D, Newman LC, Gong Y, Johnson JA, et al. Cholesteryl ester transfer protein (cetp) polymorphisms affect mrna splicing, hdl levels, and sex-dependent cardiovascular risk. *PloS one*. 2012;7:e31930
31. Suhy A, Hartmann K, Newman L, Papp A, Toneff T, Hook V, et al. Genetic variants affecting alternative splicing of human cholesteryl ester transfer protein. *Biochemical and biophysical research communications*. 2014;443:1270-1274
32. Smith EN, Chen W, Kahonen M, Kettunen J, Lehtimaki T, Peltonen L, et al. Longitudinal genome-wide association of cardiovascular disease risk factors in the bogalusa heart study. *PLoS genetics*. 2010;6:e1001094
33. Suhy A, Hartmann K, Papp AC, Wang D, Sadee W. Regulation of cholesteryl ester transfer protein expression by upstream polymorphisms: Reduced expression associated with rs247616. *Pharmacogenetics and genomics*. 2015;25:394-401
34. Pirim D, Wang X, Niemsiri V, Radwan ZH, Bunker CH, Hokanson JE, et al. Resequencing of the cetp gene in american whites and african blacks: Association of rare and common variants with hdl-cholesterol levels. *Metabolism: clinical and experimental*. 2016;65:36-47
35. Ganesan M, Nizamuddin S, Katkam SK, Kumaraswami K, Hosad UK, Lobo LL, et al. C.*84g>a mutation in cetp is associated with coronary artery disease in south indians. *PloS one*. 2016;11:e0164151
36. Matoulkova E, Michalova E, Vojtesek B, Hrstka R. The role of the 3' untranslated region in post-transcriptional regulation of protein expression in mammalian cells. *RNA biology*. 2012;9:563-576
37. Ference BA, Ginsberg HN, Graham I, Ray KK, Packard CJ, Bruckert E, et al. Low-density lipoproteins cause atherosclerotic cardiovascular disease. 1. Evidence from genetic, epidemiologic, and clinical studies. A consensus statement from the european atherosclerosis society consensus panel. *European heart journal*. 2017;38:2459-2472

38. White J, Swerdlow DI, Preiss D, Fairhurst-Hunter Z, Keating BJ, Asselbergs FW, et al. Association of lipid fractions with risks for coronary artery disease and diabetes. *JAMA cardiology*. 2016;1:692-699
39. Eli lilly and company. Lilly to discontinue development of evacetrapib for high-risk atherosclerotic cardiovascular disease. *https://investor.lilly.com/releasedetail.cfm?ReleaseID=936130, Accessed 15 October 2015*. 2015
40. Sniderman AD, Islam S, Yusuf S, McQueen MJ. Discordance analysis of apolipoprotein b and non-high density lipoprotein cholesterol as markers of cardiovascular risk in the interheart study. *Atherosclerosis*. 2012;225:444-449