

ARTICLE OPEN

Fine mapping the *CETP* region reveals a common intronic insertion associated to HDL-C

Elisabeth M van Leeuwen¹, Jennifer E Huffman^{2,3}, Joshua C Bis⁴, Aaron Isaacs¹, Monique Mulder⁵, Aniko Sabo⁶, Albert V Smith^{7,8}, Serkalem Demissie⁹, Ani Manichaikul¹⁰, Jennifer A Brody⁴, Mary F Feitosa¹¹, Qing Duan¹², Katharina E Schraut¹³, Pau Navarro², Jana V van Vliet-Ostaptchouk¹⁴, Gu Zhu¹⁵, Hamdi Mbarek¹⁶, Stella Trompet^{17,18}, Niek Verweij¹⁹, Leo-Pekka Lytyikäinen²⁰, Joris Deelen²¹, Ilja M Nolte²², Sander W van der Laan²³, Gail Davies^{24,25}, Andrea JM Vermeij-Verdoold¹, Andy ALJ van Oosterhout¹, Jeannette M Vergeer-Drop¹, Dan E Arking²⁶, Holly Trochet², Generation Scotland⁵⁸, Carolina Medina-Gomez^{1,5}, Fernando Rivadeneira^{1,5}, Andre G Uitterlinden^{1,5}, Abbas Dehghan¹, Oscar H Franco¹, Eric J Sijbrands⁵, Albert Hofman¹, Charles C White^{27,28,29}, Josyf C Mychaleckyj¹⁰, Gina M Peloso^{29,30,31,32}, Morris A Swertz³³, Lifelines Cohort Study⁵⁹, Gonneke Willemsen¹⁶, Eco J de Geus¹⁶, Yuri Milaneschi³⁴, Brenda WJH Penninx³⁴, Ian Ford³⁵, Brendan M Buckley^{36,37}, Anton JM de Craen¹⁸, John M Starr^{36,37}, Ian J Deary^{24,25}, Gerard Pasterkamp³⁸, Albertine J Oldehinkel³⁹, Harold Snieder²², P Eline Slagboom²¹, Kjell Nikus⁴⁰, Mika Kähönen⁴¹, Terho Lehtimäki²⁰, Jorma S Viikari⁴², Olli T Raitakari^{43,44}, Pim van der Harst¹⁹, J Wouter Jukema¹⁷, Jouke-Jan Hottenga¹⁶, Dorret I Boomsma¹⁶, John B Whitfield¹⁵, Grant Montgomery^{43,44}, Nicholas G Martin¹⁵, CHARGE Lipids Working Group⁵⁹, Ozren Polasek⁴⁵, Veronique Vitart², Caroline Hayward², Ivana Kolcic⁴⁵, Alan F Wright², Igor Rudan⁴⁶, Peter K Joshi¹³, James F Wilson¹³, Leslie A Lange¹⁰, James G Wilson⁴⁷, Vilmondur Gudnason^{7,8}, Tamar B Harris⁴⁸, Alanna C Morrison⁴⁹, Ingrid B Borecki¹¹, Stephen S Rich¹⁰, Sandosh Padmanabhan⁵⁰, Bruce M Psaty^{51,52}, Jerome I Rotter^{53,54,55}, Blair H Smith⁵⁶, Eric Boerwinkle⁴⁹, L Adrienne Cupples^{9,57} and Cornelia van Duijn¹

¹Department of Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands; ²MRC Human Genetics Unit, MRC IGMM, University of Edinburgh, Edinburgh, UK; ³National Heart, Lung, and Blood Institute (NHLBI) Cardiovascular Epidemiology and Human Genomics Branch, Framingham Heart Study, Framingham, MA, USA; ⁴Department of Medicine, University of Washington, Seattle, WA, USA; ⁵Department of Internal Medicine, Erasmus Medical Center, Rotterdam, The Netherlands; ⁶Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, USA; ⁷Icelandic Heart Association, Kopavogur, Iceland; ⁸Faculty of Medicine, University of Iceland, Reykjavik, Iceland; ⁹Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA; ¹⁰Center for Public Health Genomics, Department of Public Health Sciences, University of Virginia, Charlottesville, VA, USA; ¹¹Department of Genetics, Washington University School of Medicine, St Louis, MO, USA; ¹²Department of Genetics, University of North Carolina, Chapel Hill, NC, USA; ¹³Centre for Population Health Sciences, University of Edinburgh, Edinburgh, Scotland; ¹⁴Department of Endocrinology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands; ¹⁵Genetic Epidemiology, QIMR Berghofer Medical Research Institute, Brisbane, QLD, Australia; ¹⁶Department of Biological Psychology, VU University Amsterdam and EMGO Institute for Health and Care Research, Amsterdam, The Netherlands; ¹⁷Department of Cardiology, Leiden University Medical Center, Leiden, The Netherlands; ¹⁸Department of Gerontology and Geriatrics, Leiden University Medical Center, Leiden, The Netherlands; ¹⁹Department of Cardiology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands; ²⁰Department of Clinical Chemistry, Fimlab Laboratories and University of Tampere School of Medicine, Tampere, Finland; ²¹Department of Molecular Epidemiology, Leiden University Medical Center, Leiden, The Netherlands; ²²Department of Epidemiology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands; ²³Department of Experimental Cardiology, UMC Utrecht, Utrecht, The Netherlands; ²⁴Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, UK; ²⁵Department of Psychology, University of Edinburgh, Edinburgh, UK; ²⁶McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA; ²⁷Program in Translational NeuroPsychiatric Genomics, Institute for the Neurosciences, Departments of Neurology and Psychiatry, Brigham and Women's Hospital, Boston, MA, USA; ²⁸Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, USA; ²⁹Ann Romney Center for Neurologic Diseases, Brigham and Women's Hospital, Boston, MA, USA; ³⁰Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA, USA; ³¹Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA, USA; ³²Harvard Medical School, Boston, MA, USA; ³³Department of Genetics, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands; ³⁴Department of Psychiatry, VU University Medical Center Amsterdam/GGZinGeest and EMGO Institute for Health and Care Research and Neuroscience Campus Amsterdam, Amsterdam, The Netherlands; ³⁵Robertson Center for Biostatistics, University of Glasgow, Glasgow, UK; ³⁶Department of Pharmacology and Therapeutics, University College Cork, Cork, Ireland; ³⁷Alzheimer Scotland Dementia Research Centre, University of Edinburgh, Edinburgh, UK; ³⁸Laboratory of Clinical Chemistry and Hematology, Division Laboratories & Pharmacy, UMC Utrecht, Utrecht, the Netherlands; ³⁹Interdisciplinary Center Psychopathology and Emotion Regulation, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands; ⁴⁰Department of Cardiology, Heart Centre, Tampere University Hospital and University of Tampere School of Medicine, Tampere, Finland; ⁴¹Department of Clinical Physiology, Tampere University Hospital and University of Tampere School of Medicine, Tampere, Finland; ⁴²Division of Medicine, Turku University Hospital, and Department of Medicine, University of Turku, Turku, Finland; ⁴³Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, and Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland; ⁴⁴Molecular Epidemiology, QIMR Berghofer Medical Research Institute, Brisbane, QLD, Australia; ⁴⁵Department of Public Health, Faculty of Medicine, University of Split, Split, Croatia; ⁴⁶Centre for Population Health Sciences, Medical School, University of Edinburgh, Edinburgh, UK; ⁴⁷Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, MS, USA; ⁴⁸National Institute on Aging, National Institute of Health, Bethesda, MD, USA; ⁴⁹Human Genetics Center, The University of Texas School of Public Health, Houston, TX, USA; ⁵⁰Division of Cardiovascular and Medical Sciences, University of Glasgow, Glasgow, UK; ⁵¹Department of Medicine, Epidemiology & Health Services, University of Washington, Seattle, WA, USA; ⁵²Group Health Research Institute, Group Health cooperative, Seattle, WA, USA; ⁵³Institute for Translational Genomics and Population Sciences, Los Angeles BioMedical Research Institute at Harbor-UCLA Medical Center, Torrance, CA, USA; ⁵⁴Division of Genomic Outcomes, Departments of Pediatrics and Medicine, Harbor-UCLA Medical Center, Torrance, CA, USA; ⁵⁵Departments of Pediatrics, Medicine, and Human Genetics, UCLA, Los Angeles, CA, USA; ⁵⁶Medical Research Institute, University of Dundee, Dundee, UK and ⁵⁷Framingham Heart Study, Framingham, MA, USA.

Correspondence: CM van Duijn (c.vanduijn@erasmusmc.nl)

⁵⁸A Collaboration between the University Medical Schools and NHS, Aberdeen, Dundee, Edinburgh and Glasgow, UK

⁵⁹See Supplementary Information.

Received 8 March 2015; revised 24 July 2015; accepted 10 August 2015

BACKGROUND: Individuals with exceptional longevity and their offspring have significantly larger high-density lipoprotein concentrations (HDL-C) particle sizes due to the increased homozygosity for the I405V variant in the cholesteryl ester transfer protein (*CETP*) gene. In this study, we investigate the association of *CETP* and HDL-C further to identify novel, independent *CETP* variants associated with HDL-C in humans.

METHODS: We performed a meta-analysis of HDL-C within the *CETP* region using 59,432 individuals imputed with 1000 Genomes data. We performed replication in an independent sample of 47,866 individuals and validation was done by Sanger sequencing.

RESULTS: The meta-analysis of HDL-C within the *CETP* region identified five independent variants, including an exonic variant and a common intronic insertion. We replicated these 5 variants significantly in an independent sample of 47,866 individuals. Sanger sequencing of the insertion within a single family confirmed segregation of this variant. The strongest reported association between HDL-C and *CETP* variants, was rs3764261; however, after conditioning on the five novel variants we identified the support for rs3764261 was highly reduced ($\beta_{\text{unadjusted}} = 3.179$ mg/dl (P value = 5.25×10^{-509}), $\beta_{\text{adjusted}} = 0.859$ mg/dl (P value = 9.51×10^{-25})), and this finding suggests that these five novel variants may partly explain the association of *CETP* with HDL-C. Indeed, three of the five novel variants (rs34065661, rs5817082, rs7499892) are independent of rs3764261.

CONCLUSIONS: The causal variants in *CETP* that account for the association with HDL-C remain unknown. We used studies imputed to the 1000 Genomes reference panel for fine mapping of the *CETP* region. We identified and validated five variants within this region that may partly account for the association of the known variant (rs3764261), as well as other sources of genetic contribution to HDL-C.

npj Aging and Mechanisms of Disease (2015) **1**, 15011; doi:10.1038/npjamd.2015.11; published online 12 November 2015

INTRODUCTION

Aging is characterized by a deterioration in the maintenance of homeostatic processes over time, leading to functional decline and increased risk for disease and death.¹ One of the genes linked to healthy aging and longevity is the cholesteryl ester transfer protein (*CETP*) gene.^{1,2} Homozygosity in the 405VV variants of *CETP* is associated with lower concentrations of *CETP*, higher concentrations of high-density lipoprotein concentrations (HDL-C), and greater HDL-C particle size, all associated with both protection against cardiovascular disease³ and exceptional longevity.⁴

Functional analyses in mice,⁵ hamsters,⁶ and rabbits⁷ have revealed that the protein encoded by the *CETP* gene mediates the transfer of cholesteryl esters from HDL-C to other lipoproteins such as atherogenic (V)LDL particle and is a key participant in the reverse transport of cholesterol from the periphery to the liver.⁸ Due to the function of *CETP* and the association of the gene with HDL-C in humans,^{9,10} the *CETP* gene is one of the targets for drug development for dyslipidemia.^{6,11,12} *CETP*-inhibition leads to an increase of HDL-C from 30 up to 140% depending on the compound used. The first drug of its class, Torcetrapib was unfortunately associated with an increased mortality and morbidity in patients receiving the *CETP* inhibitor in addition to atorvastatin.^{13,14}

The estimated heritability of HDL-C levels is high in humans: 47–76%.^{15–23} Previously published whole-genome sequence data²³ reported that common variants (minor allele frequency (MAF) > 1%) explain up to 61.8% of the variance in HDL-C levels and that rare variants (MAF < 1%) explain an additional 7.8% of the variance. Genome-wide association studies revealed that numerous variants are associated with HDL-C, among which are various common^{9,10} and rare^{24,25} variants within the *CETP* gene in multiple ancestries.^{4,8,26–28} In this paper, we investigate the association between *CETP* and HDL-C in humans in further detail to identify variants that are likely to be causal.

To this end, we used a meta-analysis of association studies with imputed genotypes within the *CETP* region. Our study consisted of data from 59,432 samples, of which the genotypes were imputed to the 1000 Genomes project reference panel (version Phase 1 integrated release v3, April 2012, all populations). By using 1000 Genomes imputed data, we expected to find more rare or low-frequent variants, as well as novel insertions and deletions.

MATERIALS AND METHODS

Study descriptions

The descriptions of the participating cohorts can be found in the Supplementary Information. All studies were performed with the approval of the local medical ethics committees, and written informed consent was obtained from all participants.

Study samples and phenotypes

The total number of individuals in the discovery phase was 59,432 and in the replication phase 47,866. Of the discovery samples, 44,108 individuals (74.21%) were of European ancestry. Of the replication samples, 47,081 individuals (98.36%) were of European ancestry. A summary of the details of both the discovery and replication cohorts participating in this study can be found in Supplementary Table 1.

Genotyping and imputations

All cohorts were genotyped using commercially available Affymetrix or Illumina genotyping arrays, or custom Perlegen arrays. Quality control was performed independently for each study. To facilitate meta-analysis and replication, each discovery and replication cohort performed genotype imputation using IMPUTE2²⁹ or Minimac³⁰ with reference to the 1000 Genomes project reference panel. The details per cohort can be found in Supplementary Table 2.

Association analysis in discovery cohorts

The lipid measurements were adjusted for sex, age, and age² in all cohorts, and if necessary also for cohort-specific covariates (Supplementary Table 1). Some cohorts included samples using lipid-lowering medication; we did not adjust for lipid-lowering medication in our analysis because HDL-C levels are only minimally influenced by lipid-lowering medication. Each discovery cohort ran association analysis for all variants within the *CETP* region (chromosome 16, 56.99–57.02 Mbp) with HDL-C.

Meta-analysis of discovery cohorts

The association results of all discovery cohorts for all variants within the *CETP* region (chromosome 16, 56.99–57.02 Mbp) were combined using inverse-variance weighting as applied by METAL.³¹ This tool also applies genomic control by automatically correcting the test statistics to account for small amounts of population stratification or unaccounted relatedness and the tool also allows for heterogeneity. We used the following filters for the variants: $0.3 < R^2$ (measurement for the imputation quality) < 1.0 and expected minor allele count ($\text{expMAC} = 2 \times \text{MAF} \times R^2 \times \text{sample size}$) > 10 prior to meta-analysis. After meta-analysis of all available variants, we excluded the variants that were not present in at least three cohorts, to prevent false positive findings.

Selection of independent variants

To select only variants that were independently associated with HDL-C, we used the Genome-wide Complex Trait Analysis (GCTA) tool, version 1.13.³² Although this tool currently supports multiple functionalities, we only used the functions for conditional and joint genome-wide association analysis. This function performs a stepwise selection procedure to select independent single nucleotide polymorphisms (SNP) associations by a conditional and joint analysis approach. It utilizes summary-level statistics from the meta-analysis and linkage disequilibrium (LD) corrections between SNPs are estimated from the 1000 Genomes (1000G Phase I Integrated Release Version 22 Haplotypes (2010–11 data freeze, 14 February 2012 haplotypes)). GCTA estimates the effective sample size and determines the effect size, the *s.e.*, and the *P* value from a joint analysis of all the selected SNPs. In this way, we select the best associated variants in *CETP*. We subsequently checked whether these variants were in LD within the 1000 Genomes reference panel using PLINK³³ software (Supplementary Table 3).

Replication of independent CETP variants

Five variants were selected for replication in a sample of 12 independent cohorts: Athero-Express, CHS, FINCAVAS, LBC1936, Lifelines, LLS, NTR-NESDA, PREVEND, PROSPER, QIMR, TRAILS, and YFS. The lipid measurements were adjusted for sex, age, and age² in all cohorts, and if necessary also for cohort-specific covariates (Supplementary Table 1b). The details per cohort regarding variant genotyping and imputations can be found in Supplementary Table 2. The association results of all replication cohorts were combined and the *s.e.*-based weights were calculated by METAL.³¹ Since none of the five variants are in LD (Supplementary Table 3), the Bonferroni-corrected *P* value for multiple testing was 0.01.

Test previous published results

The meta-analysis of HDL-C as published by Teslovich *et al.*⁹ identified 38 genome-wide significant (*P* value < 5 × 10⁻⁸) variants within the *CETP* region (chromosome 16, 56.99–57.02 Mbp). Within all discovery and replication cohorts, we tested these 38 variants, adjusting for the 5 newly identified independent variants to explore whether the new variants explain previously published results. The association results of all cohorts were combined and the *s.e.*-based weights were calculated by METAL.³¹

We used the genotypes of all 1,092 individuals of the 1000 Genomes project to calculate the correlation between the 38 variants. This correlation matrix was used by matSpDLite³⁴ which examines the ratio of observed eigenvalue variance to its theoretical maximum to determine the number of independent variables. For these 38 genome-wide significant variants within the *CETP* region, the effective number of independent variables is 18 and therefore the experiment-wide significance threshold required to keep type I error rate at 5% is 2.85 × 10⁻³.

Conditional analysis of independent CETP variants

The replicated independent variants were selected for conditional analysis in both the discovery and the replication cohorts. In this analysis we adjusted for the lead SNP for this region as reported by Teslovich *et al.*⁹ (rs3764261, chromosome 16, position 56,993,324 bp). The association results of all discovery and replication cohorts were combined and the *s.e.*-based weights were calculated by METAL.³¹ The Bonferroni-corrected *P* value for multiple testing was 0.01, since none of the five variants is in LD (Supplementary Table 3).

Validation of the new CETP insertion within a family

Within the ERF study, 3,658 individuals have been genotyped on various Illumina (Illumina, San Diego, CA, USA) and Affymetrix chips (Affymetrix, Santa Clara, CA, USA), followed by imputations with MaCH (1.0.18c) and Minimac (minimac-β-14 March 2012) to the 1000 Genomes reference panel. Based on the best guess imputed genotypes, we selected one family in which we expected the insertion to segregate.

Validation of the insertion was performed by Sanger sequencing. Genomic DNA was isolated from peripheral blood using standard protocols (salting-out). The intron 2–3 of the *CETP* gene (Supplementary Table 4) was amplified using PCR and the following primer sequences were used to amplify: forward; 5'-tgggggactcaggctctcc-3'; reverse; 5'-aaagcacctggccca caacc-3'; size 409 bp.

PCR reactions was performed in 17.5 μl containing 37.5 ng DNA, 10 pmol/μl of each primer, 2.5 mM dNTPs, 10x PCR buffer with Mg⁺ (Roche) and 5U/μl FastStart Taq (Roche Nederland B.V., Woerden,

the Netherlands). Cycle conditions: 7 min at 94 °C; 10 cycles of 30-s denaturation at 94 °C, 30 s annealing at 70–1 °C per cycle and 90-s extension at 72 °C; followed by 20 cycles of 30-s denaturation at 94 °C, 30 s at 60 °C, and 90 s at 72 °C; final extension 10 min at 72 °C. Sephadex G50 (Amersham Biosciences) was used to purify the sequenced PCR products. Direct sequencing of both strands was performed using Big Dye Terminator chemistry version 4 (Applied Biosystems, Bleiswijk, the Netherlands). Fragments were loaded on an ABI3100 automated sequencer and analyzed with DNA Sequencing Analysis (version 5.3) and SeqScape (version 2.6) software (Applied Biosystems). All sequence variants are numbered at the nucleotide levels according to the following references: NC_000016.10:g.56963437_56963438insA (NCBI), NM_000078.2:c.233+313_233+314insA, Human Feb. 2009 (GRCh37/hg19) Assembly.

RESULTS

Meta-analysis in all discovery cohorts to select independent variants

The association of all variants within the *CETP* region (chromosome 16, 56.99–57.02 Mbp) to HDL-C was tested in all discovery cohorts. These results were combined using the inverse-variance weights as applied by METAL.³¹ After exclusion of the variants that were not present in at least 3 cohorts, 254 variants remained (Figure 1). A conditional and joint analysis of the 254 variants using GCTA identified 5 independent variants (Figure 2). Three variants were intronic (rs5817082, rs4587963, and rs7499892), one variant was intergenic (rs12920974) and one variant was exonic (rs34065661) (Table 1). Using PLINK software,³³ we calculated the LD between the five variants based on the 1000 Genomes reference panel, and found that none are in high LD with each other (Supplementary Table 3).

Replication of the independent CETP variants

The five independent variants within the *CETP* region were selected for replication within the following cohorts: Athero-Express, CHS, FINCAVAS, LBC1936, Lifelines, LLS, NTR-NESDA, PREVEND, PROSPER, QIMR, TRAILS, and YFS. Five variants were replicated at a *P* value of 2.99 × 10⁻³⁴ (Figure 3 and Table 2).

Test to explain the previously published results

In each discovery and replication cohort, we tested if the five independent variants explain the associations within the *CETP* region (chromosome 16, 56.99–57.02 Mbp) as reported in the study by Teslovich *et al.*⁹ We tested a total of 38 genome-wide significant (*P* value < 5 × 10⁻⁸) SNPs within this region identified by Teslovich *et al.*⁹ and conditioned for the five independent variants in all discovery and replication cohorts. All 38 variants were significantly (*P* value corrected for multiple

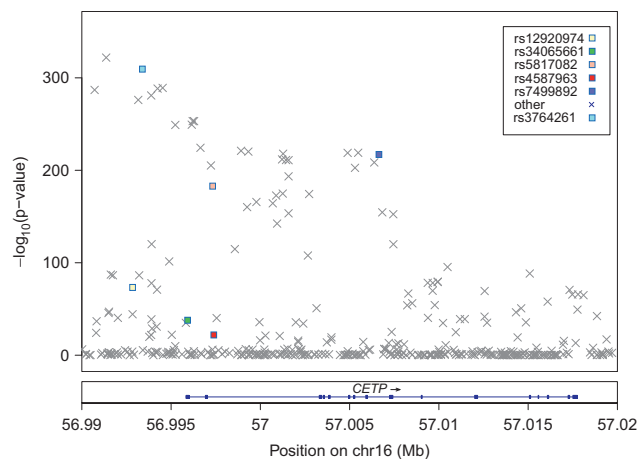


Figure 1. Results of the meta-analysis of all discovery cohorts within the *CETP* region. *CETP*, cholesteryl ester transfer protein.

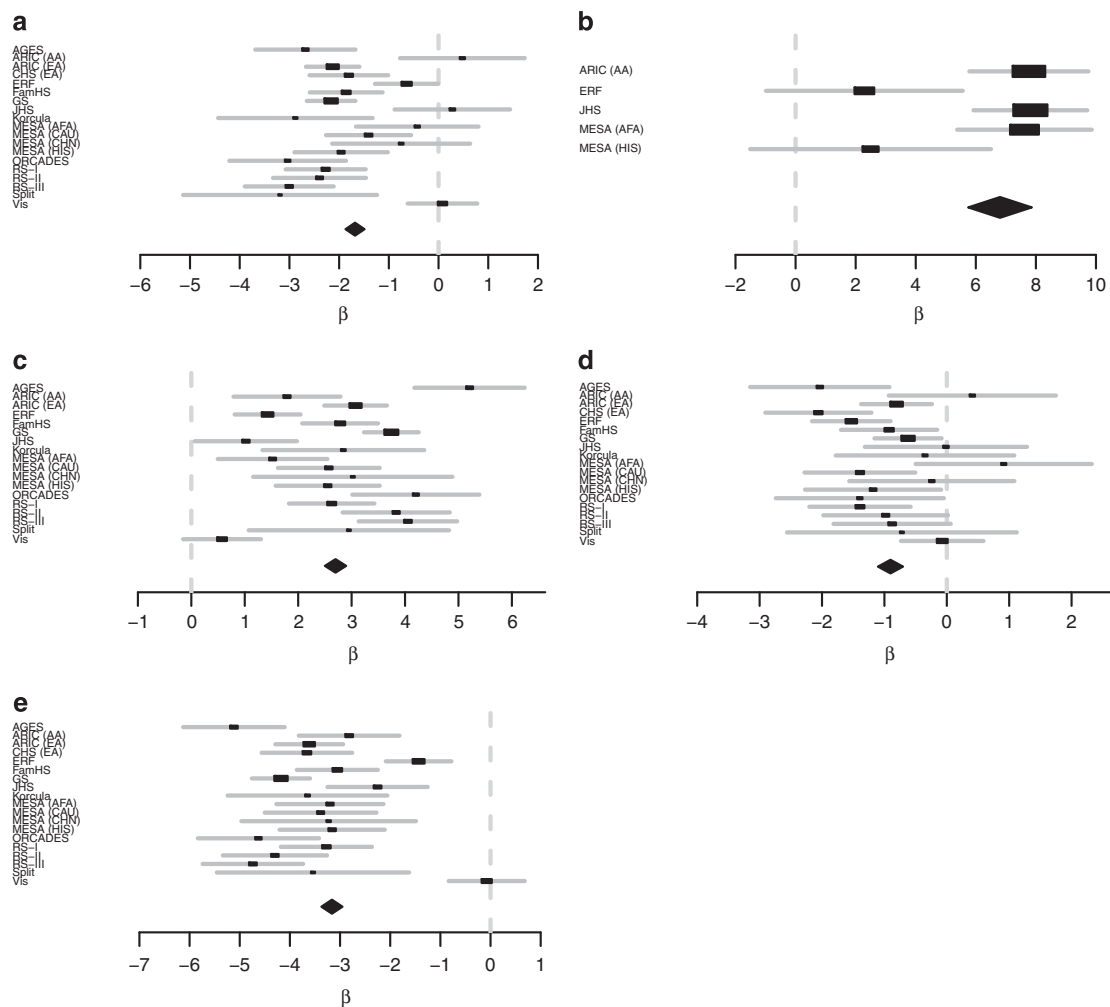


Figure 2. Forest plots from the discovery meta-analysis results for the five independent variants identified within the *CETP* region. Only cohorts in which the variants passed QC are included in the forest plot. **(a)** rs12920974 (chromosome 16, position 56,993,025), **(b)** rs34065661 (chromosome 16, position 56,995,935), **(c)** rs5817082 (chromosome 16, position 56,997,349), **(d)** rs4587963 (chromosome 16, position 56,997,369), and **(e)** rs7499892 (chromosome 16, position 57,006,590). *CETP*, cholesteryl ester transfer protein.

Table 1. The five independent variants after meta-analysis in the discovery cohorts

Marker name	Chr	Position	EA	Type	Freq	After meta-analysis			After GCTA analysis			
						β^a	S.e. $_{\beta}$	P value	Freq $_{geno}$	β_j^a	S.e. $_{\beta_j}$	P value $_j$
rs12920974	16	56,993,025	T	SNP	0.271	-1.748	0.096	1.41E-74	0.281	-1.806	0.139	2.40E-38
rs34065661	16	56,995,935	G	SNP	0.058	7.203	0.560	7.04E-38	0.020	6.782	0.582	2.23E-31
rs5817082	16	56,997,349	CA	INDEL	0.285	-2.869	0.098	8.95E-187	0.305	-4.286	0.172	1.55E-137
rs4587963	16	56,997,369	A	SNP	0.240	-0.972	0.101	5.25E-22	0.261	-2.014	0.165	2.11E-34
rs7499892	16	57,006,590	T	SNP	0.209	-3.384	0.107	2.94E-218	0.245	-2.083	0.150	1.31E-43

Abbreviations: EA, effect allele—the allele for which the effect on HDL-C is estimated; Freq, the frequency of reference allele in the discovery cohorts; Freq $_{geno}$, the frequency of the variant within the reference panel.

$^a\beta$ is the effect of the effect allele. β_j is the effect of the effect allele after joint analysis of all selected variants by GCTA.

testing $< 2.85 \times 10^{-3}$) associated with HDL-C in our joint analyses without adjusting for the 5 independent variants we identified in this work, and 37 (97.37%) were genome-wide significant (P value $< 5 \times 10^{-8}$) despite the fact that our sample size is about 65% of the study by Teslovich *et al.*⁹ (Table 3). When conditioning on the 5 variants identified in this work, 27 (71.05%) variants remained significant (P value $< 2.85 \times 10^{-3}$), though the P values were markedly reduced (Table 3). This finding suggests that the new variants we identified may explain in part the previously reported association. Remarkably, the P value of rs3764261 which

was reported as the lead SNP for this *CETP* region by Teslovich *et al.*⁹ was highly reduced from 5.25×10^{-509} to 9.51×10^{-25} while the β decreased from 3.179 mg/dl to 0.859 mg/dl. This variant is not in LD with any of the five new variants. Due to the lack of LD, the s.e. of rs3764261 does not change much (s.e. $_{unadj}$ = 0.066, s.e. $_{adj}$ = 0.084), but the effect of rs3764261 does (β_{unadj} = 3.179, β_{adj} = 0.859) and therefore the χ^2 decreases as well, and that results in a higher P value. This indicates that a part of the effect of rs3764261 can be explained by the effect of the five new variants.

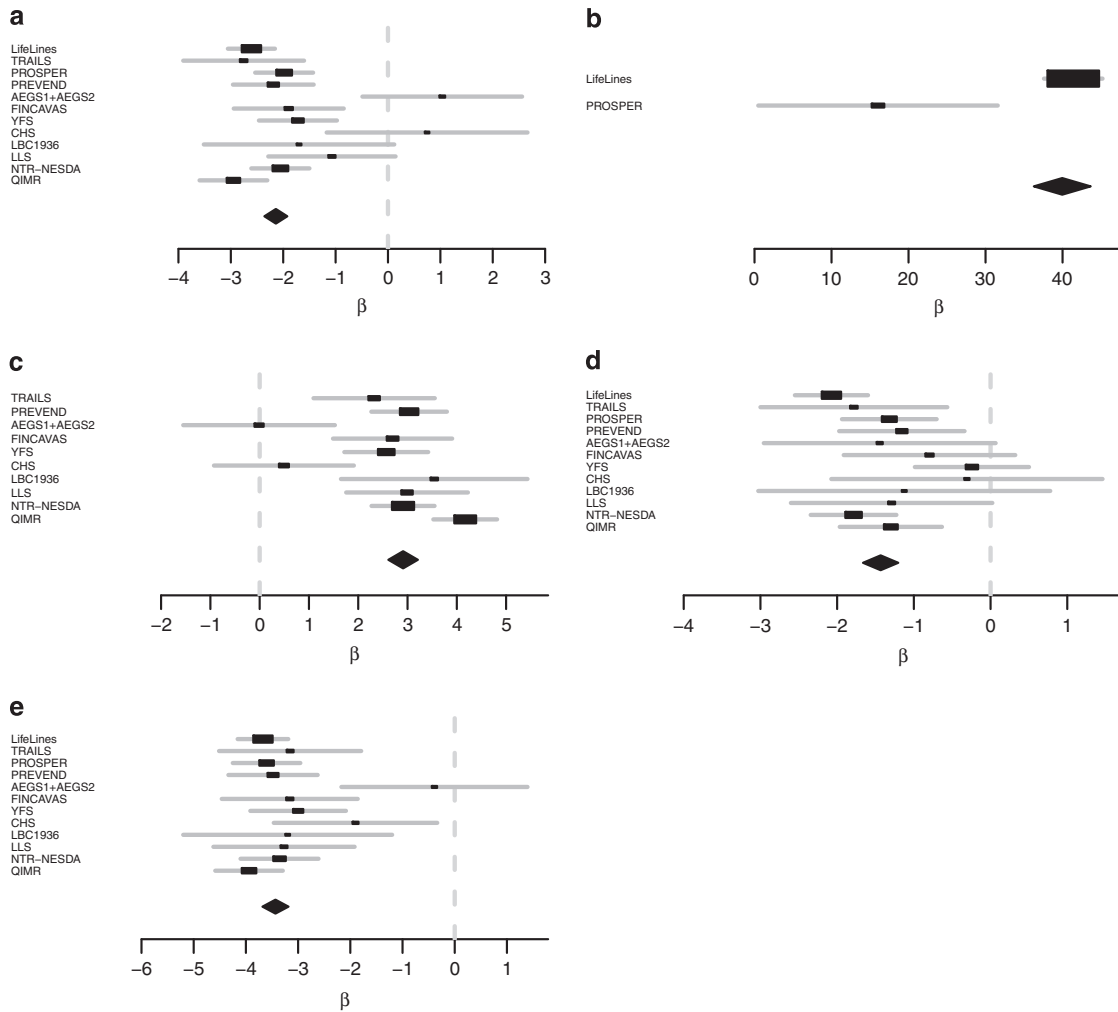


Figure 3. Forest plots of the replication meta-analysis for the five independent variants within the *CETP* region. Only cohorts in which the variants passed QC are included in the forest plot. (a) rs12920974 (chromosome 16, position 56,993,025), (b) rs34065661 (chromosome 16, position 56,995,935), (c) rs5817082 (chromosome 16, position 56,997,349), (d) rs4587963 (chromosome 16, position 56,997,369), and (e) rs7499892 (chromosome 16, position 57,006,590). *CETP*, cholesteryl ester transfer protein.

Table 2. Replication of the 5 independent variants within the *CETP* region

Marker name	Chr	Position	EA	Non effect allele	Freq	β^a	S.e. $_{\beta}$	P value	Direction of effect per cohort ^b												
rs12920974	16	56,993,025	T	G	0.288	-2.140	0.112	3.36E-81	-	-	-	-	-	-	-	-	-	-	-	-	-
rs34065661	16	56,995,935	G	C	0.018	39.958	1.884	8.46E-100	?	?	?	?	+	?	?	?	+	?	?	?	?
rs5817082	16	56,997,349	CA	C	0.229	-2.911	0.153	1.09E-80	+	-	-	-	?	-	-	-	-	-	-	-	-
rs4587963	16	56,997,369	A	T	0.325	-1.433	0.117	2.99E-34	-	-	-	-	-	-	-	-	-	-	-	-	-
rs7499892	16	57,006,590	T	C	0.257	-3.434	0.127	5.64E-160	-	-	-	-	-	-	-	-	-	-	-	-	-

Abbreviations: *CETP*, cholesteryl ester transfer protein; EA, effect allele—the allele for which the effect on HDL-C is estimated; Freq, the frequency of effect allele.

^a β is the effect of the effect allele.

^bDirection of the effect of the effect allele of the following cohorts: AEGS, CHS (AA), FINCAVAS, LBC1936, Lifelines, LLS, NTR-NESDA, PREVEND, PROSPER, QIMR, TRAILS, and YFS.

The question marks mean that the variant was removed prior to meta-analysis due to a low imputation quality and/or expMAC < 10.

Conditional analysis of the independent CETP variants

Next, we performed conditional analysis of the independent variants in both the discovery and replication cohorts. We conditioned on the lead SNP for the *CETP* region as reported by the study by Teslovich *et al.*⁹ (rs3764261, chromosome 16, position 56,993,324 bp), see Table 4 and Figure 4. This analysis showed that three out of the five variants (rs34065661, rs5817082, rs7499892) are independent of rs3764261. For all variants the *P* values and β 's decreased, but all *P* values remained significant. The effect of the

single variant rs34065661, of the insertion rs5817082, and of the single variant rs7499892 were reduced by 53.20%, 38.48%, and 32.67%, respectively.

Validation of the insertion within a family

We selected based on the best guess imputations of the ERF study, a large family of 30 individuals for Sanger sequencing of rs5817082. Using MERLIN³⁵ we estimated that the total heritability

Table 3. Unadjusted and conditional analysis of the Teslovich variants on the five independent variants in the combined analysis of all discovery and replication cohorts

Marker name	Chr	position	EA	NEA	Unadjusted analysis				Adjusted analysis			
					Freq	β^a	S.e. $_{\beta}$	P value	Freq	β^a	S.e. $_{\beta}$	P value
rs6499861	16	56,991,495	C	G	0.758	1.432	0.090	5.63E-57	0.781	1.083	0.106	1.47E-24
rs6499863	16	56,992,017	A	G	0.251	-1.420	0.093	1.02E-52	0.227	-1.162	0.112	2.59E-25
rs12708967	16	56,993,211	T	C	0.726	2.419	0.087	9.61E-170	0.768	-0.363	0.110	9.99E-04
rs3764261	16	56,993,324	A	C	0.409	3.179	0.066	5.25E-509	0.358	0.859	0.084	9.51E-25
rs12447839	16	56,993,935	T	C	0.665	1.215	0.077	1.87E-56	0.738	0.302	0.111	6.35E-03
rs12447924	16	56,994,192	T	C	0.683	1.218	0.077	8.54E-57	0.737	0.321	0.109	3.15E-03
rs4783961	16	56,994,894	A	G	0.496	1.680	0.064	9.60E-152	0.493	0.732	0.073	6.73E-24
rs4783962	16	56,995,038	T	C	0.318	-1.178	0.081	1.51E-48	0.255	-0.288	0.123	1.97E-02
rs1800775	16	56,995,236	A	C	0.471	2.788	0.064	2.12E-416	0.495	0.547	0.088	4.97E-10
rs711752	16	56,996,211	A	G	0.445	2.782	0.064	3.93E-414	0.435	0.396	0.083	1.56E-06
rs1864163	16	56,997,233	A	G	0.311	-2.991	0.076	1.33E-340	0.238	-0.307	0.115	7.75E-03
rs9929488	16	56,998,572	C	G	0.338	-2.189	0.075	7.55E-189	0.308	0.125	0.092	1.76E-01
rs7203984	16	56,999,258	A	C	0.693	2.903	0.080	2.44E-287	0.737	0.076	0.112	4.95E-01
rs11508026	16	56,999,328	T	C	0.417	2.703	0.065	1.27E-383	0.407	0.326	0.082	7.60E-05
rs820299	16	57,000,284	A	G	0.578	0.892	0.066	8.60E-42	0.595	0.336	0.084	6.07E-05
rs12597002	16	57,002,404	A	C	0.389	-1.228	0.071	2.02E-66	0.307	-0.481	0.103	3.25E-06
rs9926440	16	57,002,663	C	G	0.371	-2.141	0.072	1.18E-196	0.351	0.131	0.085	1.26E-01
rs9939224	16	57,002,732	T	G	0.288	-2.944	0.080	2.72E-300	0.229	0.051	0.109	6.41E-01
rs11076174	16	57,003,146	T	C	0.797	2.388	0.123	1.70E-83	0.825	0.496	0.133	1.99E-04
rs7205804	16	57,004,889	A	G	0.440	2.644	0.063	1.63E-386	0.422	0.291	0.082	3.51E-04
rs1532624	16	57,005,479	A	C	0.420	2.639	0.063	6.82E-386	0.412	0.291	0.082	3.48E-04
rs11076175	16	57,006,378	A	G	0.740	3.326	0.084	5.05E-342	0.815	-0.031	0.127	8.05E-01
rs7499892	16	57,006,590	T	C	0.323	-3.227	0.084	6.95E-323	0.241	-0.197	0.119	9.74E-02
rs289714	16	57,007,451	A	G	0.669	2.624	0.085	6.46E-208	0.708	0.540	0.101	1.01E-07
rs289715	16	57,008,508	A	T	0.256	2.047	0.106	5.38E-83	0.245	0.420	0.106	7.37E-05
rs289717	16	57,009,388	A	G	0.422	-1.357	0.068	1.39E-89	0.401	-0.353	0.077	4.15E-06
rs289719	16	57,009,941	T	C	0.383	1.701	0.070	2.85E-132	0.374	0.461	0.072	1.32E-10
rs4784744	16	57,011,185	A	G	0.396	-1.319	0.066	1.05E-87	0.386	-0.350	0.074	2.37E-06
rs4784745	16	57,014,875	A	G	0.614	1.327	0.068	5.66E-85	0.626	0.314	0.075	3.21E-05
rs5880	16	57,015,091	C	G	0.135	-4.495	0.175	4.42E-146	0.119	-1.331	0.181	1.92E-13
rs5882	16	57,016,092	A	G	0.613	-1.442	0.067	4.19E-102	0.614	-0.410	0.069	2.39E-09
rs9923854	16	57,017,002	T	G	0.802	-1.391	0.115	1.07E-33	0.805	-0.543	0.117	3.28E-06
rs289741	16	57,017,474	A	G	0.631	-1.547	0.068	3.37E-113	0.633	-0.476	0.070	1.02E-11
rs1801706	16	57,017,662	A	G	0.276	1.040	0.091	1.82E-30	0.270	0.493	0.095	1.92E-07
rs289742	16	57,017,762	C	G	0.295	1.811	0.098	1.21E-76	0.285	0.407	0.098	3.40E-05
rs289744	16	57,018,102	T	G	0.641	-1.544	0.069	4.99E-110	0.643	-0.469	0.071	3.33E-11
rs12720917	16	57,019,392	T	C	0.769	-1.474	0.110	1.15E-40	0.775	-0.377	0.109	5.43E-04
rs289745	16	57,019,532	A	C	0.579	0.276	0.081	6.82E-04	0.581	0.204	0.081	1.12E-02

Abbreviations: EA, effect allele for which the effect is estimated; Freq, the frequency of effect allele; NEA, non-effect allele.

^a β is the effect of effect allele.**Table 4.** Analysis of the independent variants within the *CETP* region conditioned on the lead SNP for the *CETP* region as reported by the study by Teslovich *et al.*⁹ (rs3764261) in the combined analysis of all discovery and replication cohorts

Marker name	Chr	Position	EA	NEA	Unadjusted analysis				Adjusted analysis			
					Freq	β^a	S.e. $_{\beta}$	P value	Freq	β^a	S.e. $_{\beta}$	P value
rs12920974	16	56,993,025	T	G	0.344	-1.880	0.074	9.91E-143	0.336	-0.278	0.076	2.82E-04
rs34065661	16	56,995,935	C	G	0.854	-9.333	0.520	6.02E-72	0.838	-4.368	0.550	1.94E-15
rs5817082	16	56,997,349	CA	C	0.360	-2.765	0.085	1.49E-231	0.351	-1.701	0.086	2.16E-86
rs4587963	16	56,997,369	A	T	0.351	-1.133	0.077	1.62E-48	0.339	0.309	0.079	8.81E-05
rs7499892	16	57,006,590	T	C	0.317	-3.275	0.082	2.90E-346	0.304	-2.205	0.083	5.14E-156

Abbreviations: *CETP*, cholesteryl ester transfer protein; EA, effect allele for which the effect on HDL-C is estimated; Freq, the frequency of effect allele; SNP, single nucleotide polymorphism.^a β is the effect of the effect allele.

of HDL-C within this family is 27.47%. DNA was available for 16 individuals. Figure 5 shows the results of the Sanger sequencing for rs5817082 for these 16 individuals within the family. The sequencing of the insertion confirmed the best guess results for 10 individuals (62.5%), of which 7 were heterozygous for the

insertion, 1 was homozygous for the insertion, and 2 did not carry the insertion. Three individuals that are homozygous for the insertion, were predicted to be heterozygous by the best guess imputations. Three individuals that are heterozygous for the insertion were not predicted to carry the insertion by the best

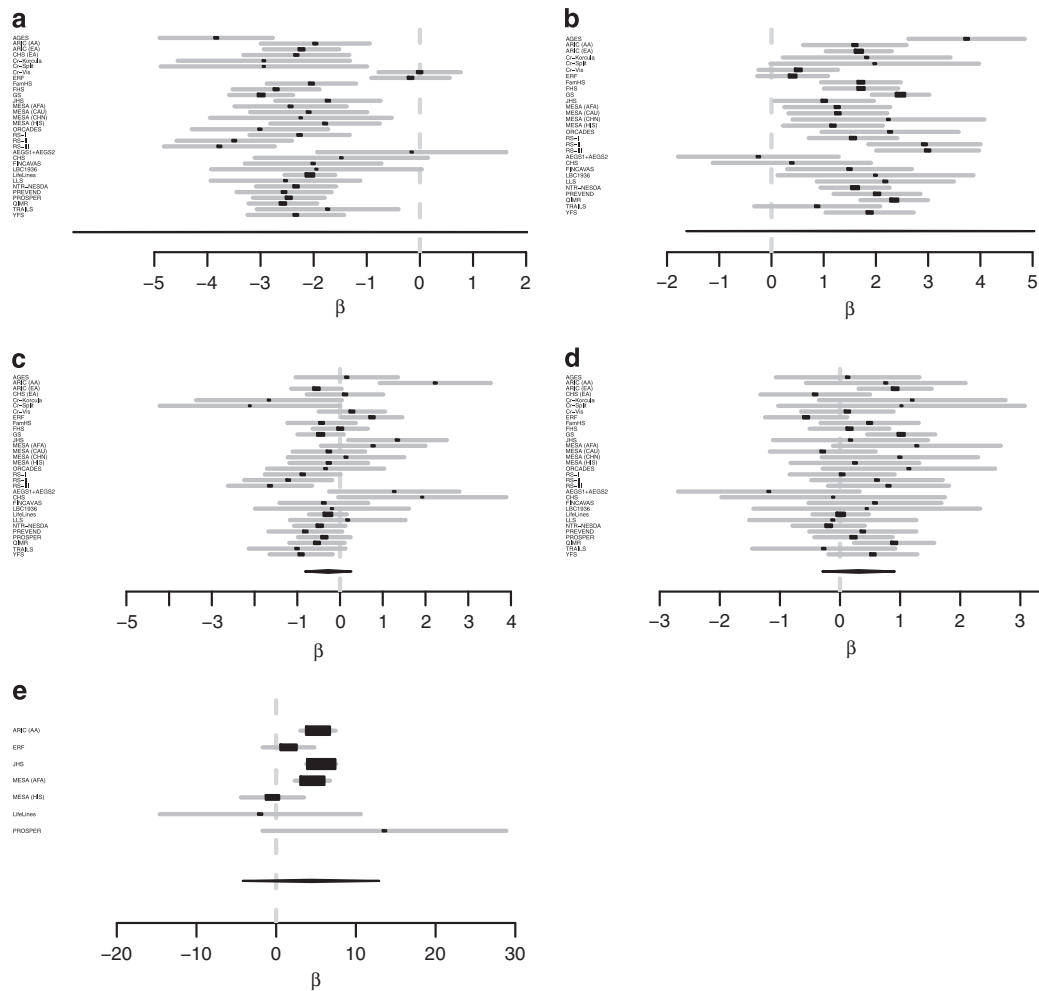


Figure 4. Forest plots of the conditional analysis in the combined discovery and replication cohorts for the five independent variants within the *CETP* region. Only cohorts in which the variants passed quality control (QC) are included in the forest plot. (a) rs12920974 (chromosome 16, position 56,993,025), (b) rs34065661 (chromosome 16, position 56,995,935), (c) rs5817082 (chromosome 16, position 56,997,349), (d) rs4587963 (chromosome 16, position 56,997,369), and (e) rs7499892 (chromosome 16, position 57,006,590). *CETP*, cholesteryl ester transfer protein.

guess imputations. Furthermore, the Sanger sequencing showed that the insertion segregates with the outcome within this family. The proportion of variance explained by the insertion within this family is 35.50%, while the proportion explained by rs3764261, the lead SNP within the *CETP* region as reported by the study by Teslovich *et al.*⁹ is 14.11%.

DISCUSSION

We conducted an analysis to fine map the association between *CETP* genetic variants and HDL-C. To this end, a total of 59,432 samples were imputed to the latest version of the 1000 Genomes (version Phase 1 integrated release v3, April 2012, all populations). We identified and replicated five independent variants within the *CETP* region (chromosome 16, 56.99–57.02 Mbp), of which four are SNPs and one is an insertion. We validated the insertion by Sanger sequencing within a large family, as the largest effect on HDL-C comes from this insertion.

The relationship between the *CETP* gene and HDL-C has been known for a long time⁹ and genome-wide association studies have revealed many common and rare variants in this region. Although the associated genetic variants are strongly correlated with HDL-C, the causal variants have not been determined. Our study showed that when using the latest 1000 Genomes reference

panel, we have more power to fine map this association. By conditional analysis of the five variants, we were able to reduce the *P* values of the genome-wide significant associations published before by Teslovich *et al.*⁹ Furthermore, conditional analysis showed that three out of the five variants are independent of the lead SNP for the *CETP* region as reported by the study by Teslovich *et al.*⁹ (rs3764261).

Several fine-mapping effort have been previously published^{36,37} and in all those efforts sequencing was used for the fine mapping. In our project we did not use sequencing, but imputations using the 1000 Genomes as a reference panel. This method has been widely used in the past and is much lower in cost. With new reference panels available, we were able to have a revised study of this region. The 1000 Genomes reference panel consists of 30 million variants including a million insertions and deletions. By using this reference panel for imputation, we were able to impute these insertions and deletions in 59,432 samples from various cohorts. This led to the significant association of an insertion within a known region with HDL-C. So far, no association between a structural variation and HDL-C has been found in such a large sample size. Validation of the insertion by Sanger sequencing confirms the correct imputations of this insertion in 62.5% of the individuals, of which seven heterozygous carriers, one homozygous carrier and two did not carry the insertion.

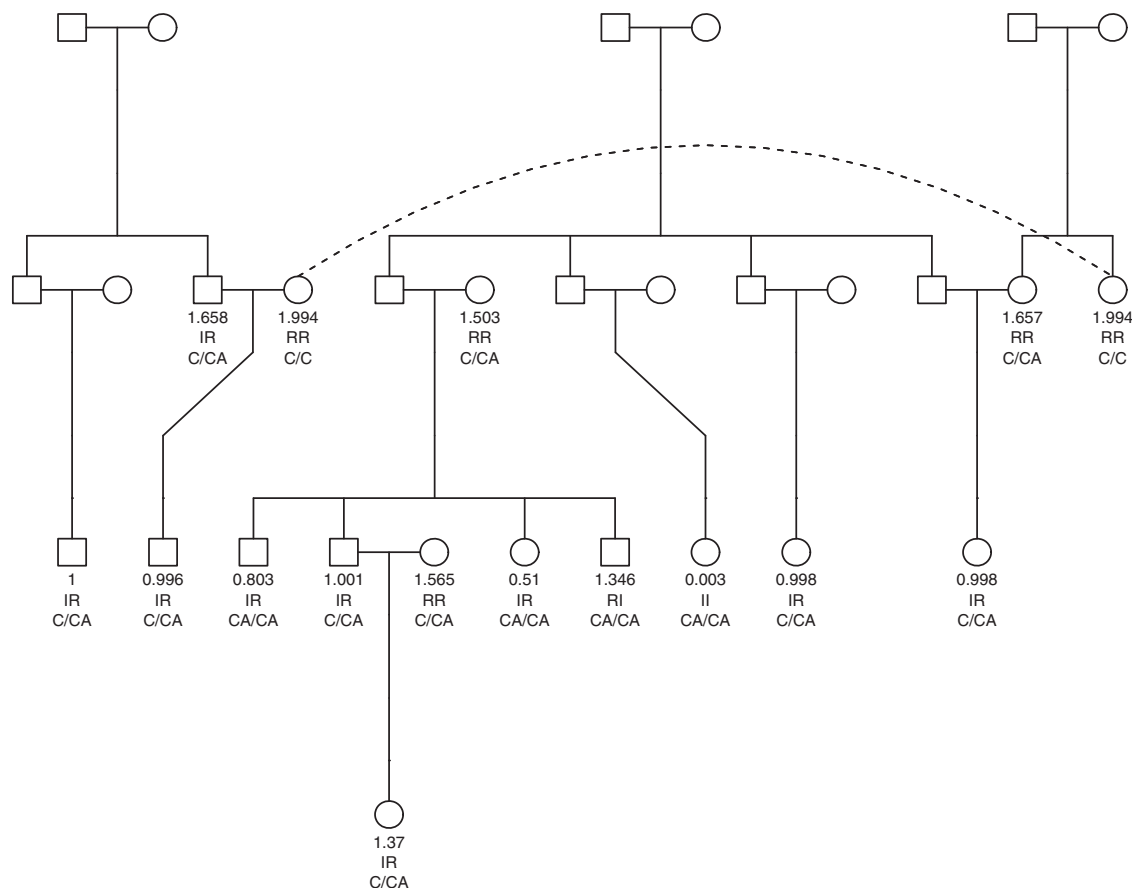


Figure 5. Validation of the insertion (rs5817082) with a large family. The numbers present the dosage for rs5817082 after imputations, second row the best guess result (I is insertion, R is reference) and the third row the genotypes of the insertion from Sanger sequencing.

The results of this study showed that by using the 1000 Genomes reference panel, the proportion of the variance explained can be increased and that multiple common variants in the same region may be implicated in a single family of the ERF study. The insertion we identified in this study explains 35.50% of variation in the HDL-C level in a single family of the ERF study; this is in concordance with the results of the whole-genome sequence data.²³ This is much higher than the proportion of the variance explained (14.11%) in the same family by rs3764261, which was reported before as the lead variant of this region. Fine mapping of various associations may help us to unravel the genetic background of various phenotypes.

Although rs3764261 was identified by Teslovich *et al.*⁹ to be the lead SNP of this region, other variants are used in clinical settings. Three of the classical variants are located in the promoter region of the CETP gene: $-1337C/T$ (rs708272 or Taq1B), $-971G/A$, and $-629C/A$ (rs1800775) polymorphisms.³⁸ Carriers of the B2 allele of the common Taq1B polymorphism exhibit lower plasma CETP levels and higher HDL-C. Furthermore, a recent meta-analysis showed that the B2 allele is associated with a reduced risk for coronary heart disease.³⁹ One more classical variant is rs5882A (405I/V), which is located outside the promoter region.⁴⁰ The $-1337C/T$ and $-629C/A$ are in strong LD, however, they are in very low LD (r^2 of 0.442 for rs708272 and 0.461 for rs1800775) with rs3764261, despite the fact that all three variant are within 3,000 bp of each other.

Large HDL-C particle sizes have been associated with exceptional longevity before and with an increased homozygosity for the I405V variant within the CETP gene.^{1–4} Many of the studies confirm this relationship, however, all are based on genotyping of the I405V variant. Our study, however, shows that more variants

within the CETP gene are associated with HDL-C levels in the blood circulation. Therefore we would suggest investigating more variants within the CETP gene for its association with longevity and healthy aging.

Some genetic variants identified in our study were published before,^{41,42} but so far no conditional analyses have been performed with these variants. Our study suggests that various CETP variants may be relevant for HDL-levels in the blood circulation and that these may have a substantial role in the heritability of HDL-C in specific families.

ACKNOWLEDGEMENTS

We especially thank all volunteers who participated in our study. Further detailed acknowledgements are provided in the Supplementary Information. The funding sources of this project can be found in the Supplementary Information.

CONTRIBUTIONS

EMvL organized the study and designed the study with substantial input from AI, LAC and CMvD. EMvL drafted the manuscript with substantial input from SSR, CvD, BMP, SWvdL, ST, JAB, JBW, GMP, AS, JvVv, DIB, GD, HS, L-PL, JEH and DEA. All authors had the opportunity to comment on the manuscript. Data collection, GWAS and statistical analysis were done by SWvdL, GP (AEGS); AVS, VG, TBH (AGES); AVS, DEA, ACM, EB (ARIC); JCB, JAB, BMP (CHS); AI, EMvL, CMvD (ERF); MFF, IBB (FamHS); SD, CCW, LAC (FHS); KN, L-PL, MK, TL (FINCAVAS and YFS); HT, SP, BHS (GS); QD, GMP, LAL, JGW (JHS); JEH, CH, IK (CROATIA Korcula); GD, JMS, IJD (LBC1936); JvVv, MAS (Lifelines); JD, AJMdc, PES (LLS); AM, JCM, SSR, JIR (MESA); HM, GW, EjdG, YM, BWJHP, J-JH, DIB (NTR-NESDA); KES, PKJ, JFW (ORCADES); NV, PvdH (PREVEND); ST, IF, BMB, JWJ (PROSPER); GZ, GW, NGM (QIMR); EMvL, MM, CM-G, FR, AGU, AD, OHF, EJS, AH, CMvD (RS); OTR, VV (CROATIA Split); IMN, AJO, HS (TRAILS); PN, AFW, IR (CROATIA Vis); JvVv-O and OTR (YFS). The Sanger sequencing was done by AJMV-V, AALJvO, JMV-D. EMvL

performed the meta-analysis and all follow-up steps. Biological association of loci and bioinformatics were carried out by EMvL and CMvD.

COMPETING INTERESTS

PSM serves on the DSMB of a clinical trial of a device funded by the manufacturer (Zoll LifeCor) and on the Steering Committee of the Yale Open Data Access Project funded by Johnson & Johnson. SWVL is a former employee of Cavadis B.V. GP is a founder and stockholder of Cavadis B.V.

REFERENCES

- Barzilai N, Gabrieli I, Atzmon G, Suh Y, Rothenberg D, Bergman A. Genetic studies reveal the role of the endocrine and metabolic systems in aging. *J Clin Endocrinol Metab* 2010; **95**: 4493–4500.
- Barzilai N, Huffman DM, Muzumdar RH, Bartke A. The critical role of metabolic pathways in aging. *Diabetes* 2012; **61**: 1315–1322.
- Vergani C, Lucchi T, Caloni M, Ceconi I, Calabresi C, Scurati S *et al*. I405V polymorphism of the cholesteryl ester transfer protein (CETP) gene in young and very old people. *Arch Gerontol Geriatr* 2006; **43**: 213–221.
- Barzilai N, Atzmon G, Schechter C, Schaefer EJ, Cupples AL, Lipton R *et al*. Unique lipoprotein phenotype and genotype associated with exceptional longevity. *JAMA* 2003; **290**: 2030–2040.
- Hayek T, Masucci-Magoulas L, Jiang X, Walsh A, Rubin E, Breslow JL *et al*. Decreased early atherosclerotic lesions in hypertriglyceridemic mice expressing cholesteryl ester transfer protein transgene. *J Clin Invest* 1995; **96**: 2071–2074.
- Briand F, Thieblemont Q, Muzotte E, Burr N, Urbain I, Sulpice T *et al*. Anacetrapib and dalcetrapib differentially alters HDL metabolism and macrophage-to-feces reverse cholesterol transport at similar levels of CETP inhibition in hamsters. *Eur J Pharmacol* 2014; **740**: 135–143.
- Kee P, Caiazza D, Rye KA, Barrett PHR, Morehouse LA, Barter PJ. Effect of inhibiting cholesteryl ester transfer protein on the kinetics of high-density lipoprotein cholesteryl ester transport in plasma: in vivo studies in rabbits. *Arterioscler Thromb Vasc Biol* 2006; **26**: 884–890.
- Zhong S, Sharp DS, Grove JS, Bruce C, Yano K, Curb JD *et al*. Increased coronary heart disease in Japanese-American men with mutation in the cholesteryl ester transfer protein gene despite increased HDL levels. *J Clin Invest* 1996; **97**: 2917–2923.
- Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M *et al*. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* 2010; **466**: 707–713.
- Global Lipids Genetics Consortium, Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S *et al*. Discovery and refinement of loci associated with lipid levels. *Nat Genet* 2013; **45**: 1274–1283.
- Siebel AL, Natoli AK, Yap FYT, Carey AL, Reddy-Luthmoodoo M, Sviridov D *et al*. Effects of high-density lipoprotein elevation with cholesteryl ester transfer protein inhibition on insulin secretion. *Circ Res* 2013; **113**: 167–175.
- Remaley AT, Norata GD, Catapano AL. Novel concepts in HDL pharmacology. *Cardiovasc Res* 2014; **103**: 423–428.
- Joy TR, Hegele RA. The failure of torcetrapib: what have we learned? *Br J Pharmacol* 2008; **154**: 1379–1381.
- Barter PJ, Caulfield M, Eriksson M, Grundy SM, Kastelein JJP, Komajda M *et al*. Effects of torcetrapib in patients at high risk for coronary events. *N Engl J Med* 2007; **357**: 2109–2122.
- Snieider H, van Doornen LJ, Boomsma DI. Dissecting the genetic architecture of lipids, lipoproteins, and apolipoproteins: lessons from twin studies. *Arterioscler Thromb Vasc Biol* 1999; **19**: 2826–2834.
- Friedlander Y, Kark JD, Stein Y. Biological and environmental sources of variation in plasma lipids and lipoproteins: the Jerusalem Lipid Research Clinic. *Hum Hered* 1986; **36**: 143–153.
- Souren NY, Paulussen ADC, Loos RJF, Gielen M, Beunen G, Fagard R *et al*. Anthropometry, carbohydrate and lipid metabolism in the East Flanders Prospective Twin Survey: heritabilities. *Diabetologia* 2007; **50**: 2107–2116.
- Sung J, Lee K, Song YM. Heritabilities of the metabolic syndrome phenotypes and related factors in Korean twins. *J Clin Endocrinol Metab* 2009; **94**: 4946–4952.
- Almgren P, Lehtovirta M, Isomaa B, Sarelin L, Taskinen MR, Lyssenko V *et al*. Heritability and familiarity of type 2 diabetes and related quantitative traits in the Botnia Study. *Diabetologia* 2011; **54**: 2811–2819.
- Vattikuti S, Guo J, Chow CC. Heritability and genetic correlations explained by common SNPs for metabolic syndrome traits. *PLoS Genet* 2012; **8**: e1002637.
- Zhou X, Carbonetto P, Stephens M. Polygenic modeling with bayesian sparse linear mixed models. *PLoS Genet* 2013; **9**: e1003264.

- Browning SR, Browning BL. Identity-by-descent-based heritability analysis in the Northern Finland Birth Cohort. *Hum Genet* 2013; **132**: 129–138.
- Morrison AC, Voorman A, Johnson AD, Liu X, Yu J, Li A *et al*. Whole-genome sequence-based analysis of high-density lipoprotein cholesterol. *Nat Genet* 2013; **45**: 899–901.
- Peloso GM, Auer PL, Bis JC, Voorman A, Morrison AC, Stitzel NO *et al*. Association of low-frequency and rare coding-sequence variants with blood lipids and coronary heart disease in 56,000 whites and blacks. *Am J Hum Genet* 2014; **94**: 223–232.
- Singaraja RR, Tietjen I, Hovingh GK, Franchini PL, Radomski C, Wong K *et al*. Identification of four novel genes contributing to familial elevated plasma HDL cholesterol in humans. *J Lipid Res* 2014; **55**: 1693–1701.
- Misra A, Shrivastava U. Obesity and dyslipidemia in South Asians. *Nutrients* 2013; **5**: 2708–2733.
- Sun L, Hu C, Zheng C, Huang Z, Lv Z, Huang J *et al*. Gene-gene interaction between CETP and APOE polymorphisms confers higher risk for hypertriglyceridemia in oldest-old Chinese women. *Exp Gerontol* 2014; **55**: 129–133.
- Walia GK, Gupta V, Aggarwal A, Asghar M, Dudbridge F, Timpson N *et al*. Association of common genetic variants with lipid traits in the Indian population. *PLoS ONE* 2014; **9**: e101688.
- Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* 2009; **5**: e1000529.
- Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet* 2012; **44**: 955–959.
- Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genome-wide association scans. *Bioinformatics* 2010; **26**: 2190–2191.
- Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* 2011; **88**: 76–82.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D *et al*. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007; **81**: 559–575.
- Li J, Ji L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity (Edinb)* 2005; **95**: 221–227.
- Abecasis GR, Cherny SS, Cookson WO, Cardon LR. Merlin-rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet* 2002; **30**: 97–101.
- Wu Y, Waite LL, Jackson AU, Sheu WHH, Buyske S, Absher D *et al*. Trans-ethnic fine-mapping of lipid loci identifies population-specific signals and allelic heterogeneity that increases the trait variance explained. *PLoS Genet* 2013; **9**: e1003379.
- Sanna S, Li B, Mulas A, Sidore C, Kang HM, Jackson AU *et al*. Fine mapping of five loci associated with low-density lipoprotein cholesterol detects variants that double the explained heritability. *PLoS Genet* 2011; **7**: e1002198.
- Le Goff W, Guerin M, Nicaud V, Dached C, Luc G, Arveiler D *et al*. A novel cholesteryl ester transfer protein promoter polymorphism (–971G/A) associated with plasma high-density lipoprotein cholesterol levels. Interaction with the TaqIB and –629C/A polymorphisms. *Atherosclerosis* 2002; **161**: 269–279.
- Boekholdt SM, Sacks FM, Jukema JW, Shepherd J, Freeman DJ, McMahon AD *et al*. Cholesteryl ester transfer protein TaqIB variant, high-density lipoprotein cholesterol levels, cardiovascular risk, and efficacy of pravastatin treatment: individual patient meta-analysis of 13,677 subjects. *Circulation* 2005; **111**: 278–287.
- Peloso GM, Demissie S, Collins D, Mirel DB, Gabriel SB, Cupples LA *et al*. Common genetic variation in multiple metabolic pathways influences susceptibility to low HDL-cholesterol and coronary heart disease. *J Lipid Res* 2010; **51**: 3524–3532.
- Ko A, Cantor RM, Weissglas-Volkov D, Nikkola E, Reddy PMVL, Sinsheimer JS *et al*. Amerindian-specific regions under positive selection harbour new lipid variants in Latinos. *Nat Commun* 2014; **5**: 3983.
- Feitosa MF, Wojczynski MK, Straka R, Kammerer CM, Lee JH, Kraja AT *et al*. Genetic analysis of long-lived families reveals novel variants influencing high density-lipoprotein cholesterol. *Front Genet* 2014; **5**: 159.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

Supplementary Information accompanies the paper on the *npj Aging and Mechanisms of Disease* website (<http://www.nature.com/npjamd>)