Analysis with the exome array identifies multiple new independent variants in lipid loci

1. William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK.

2. Department of Cardiovascular Sciences, University of Leicester, BHF Cardiovascular Research Centre, Glenfield Hospital, Leicester, UK.

3. NIHR Leicester Cardiovascular Biomedical Research Unit, Glenfield Hospital, Leicester, UK.

4. Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK.

5. The Broad Institute of MIT and Harvard, Boston, MA, USA.

6. Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA, USA.

7. Department of Clinical Sciences, Genetic and Molecular Epidemiology Unit, Lund University, Skåne University Hospital Malmö, Malmö, Sweden.

8. Clinical Pharmacology, William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK.

9. NIHR Barts Cardiovascular Biomedical Research Unit, William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, UK.

10. Medical Research Council (MRC) Epidemiology Unit, Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge, UK.

11. Wellcome Trust Centre for Human Genetics, Oxford, OX3 7BN, UK

12. Department of Epidemiology and Biostatistics, Imperial College London, London, UK.

13. Ealing Hospital NHS Trust, Middlesex, UK.
14. Genetics of Complex Traits, University of Exeter Medical School, University of Exeter, Exeter, UK.

15. Department of Medicine I, University Hospital Grosshadern, Ludwig-Maximilians-Universität, Munich, Germany.

16. Institute of Genetic Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany.

17. Institute of Medical Informatics, Biometry and Epidemiology, Ludwig-Maximilians-Universität, Munich, Germany.

18. DZHK (German Centre for Cardiovascular Research), partner site Munich Heart Alliance, Munich, Germany

19. Department of Epidemiology and Biostatistics, MRC Health Protection Agency (HPE) Centre for Environment and Health, School of Public Health, Imperial College London, London, UK.

20. Atherosclerosis Research Unit, Department of Medicine Solna, Karolinska Institutet, Stockholm, Sweden.

21. Section of Investigative Medicine, Imperial College London, London, UK.

22. CEA, Institut de Génomique, Centre National de Génotypage, Evry, France.

23. Medical Research Institute, University of Dundee, Ninewells Hospital and Medical School, Dundee, UK.

24. BHF Glasgow Cardiovascular Research Centre, Division of Cardiovascular and Medical Sciences, University of Glasgow, Western Infirmary, Glasgow, UK

25. Department of Nutrition and Dietetics, School of Health Science and Education, Harokopio University Athens, Athens, Greece.

27. Research Unit of Molecular Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany.

28. Institute of Epidemiology II, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany.

29. German Center for Diabetes Research, Neuherberg, Germany

30. Department of Public Health and Clinical Medicine, Skellefteå Research Unit, Umeå University, Umeå.

31. Echinos Medical Centre, Echinos, Greece.

32. Institute of Health Sciences, University of Oulu, Oulu, Finland

33. International Centre for Circulatory Health, Imperial College London, London, UK.


35. Department of Biobank Research, Umeå University, Umeå, Sweden.

36. Department of Public Health & Clinical Medicine, Section for Family Medicine, Umeå University, Umeå, Sweden.


38. Complex and Adaptive Systems Laboratory, University College Dublin, Belfield, Dublin, Ireland.


40. Anogia Medical Centre, Anogia, Greece.
41. Department of Life Sciences, College of Health and Life Sciences, Brunel University London, UK.
42. Imperial College Healthcare NHS Trust, London, UK.
43. National Heart and Lung Institute, Imperial College London, London, UK.
44. Oxford National Institute for Health Research (NIHR) Biomedical Research Centre, Churchill Hospital, Oxford, UK.
45. Department of Health Sciences, University of Leicester, Leicester, UK.
46. Biocenter Oulu, University of Oulu, Finland.
47. Unit of Primary Care, Oulu University Hospital, Oulu, Finland.
48. Department of Children and Young People and Families, National Institute for Health and Welfare, Oulu, Finland.
49. Department of Nutrition, Harvard School of Public Health, Boston, MA, USA.
50. Department of Public Health & Clinical Medicine, Umeå University Hospital, Umeå, Sweden.
51. Princess Al-Jawhara Al-Brahim Centre of Excellence in Research of Hereditary Disorders (PACER-HD), King Abdulaziz University, Jeddah, Saudi Arabia.

¶ These authors contributed equally to this work

* Corresponding author: Panos Deloukas, William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK, Tel: + 44 (0)20 78822103, Fax: + 44 (0) 207 882 3408, Email: (p.deloukas@qmul.ac.uk)
ABSTRACT

It has been hypothesised that low frequency (1-5% MAF) and rare (<1% MAF) variants with large effect sizes may contribute to the missing heritability in complex traits. Here we report an association analysis of lipid traits (total cholesterol, LDL-cholesterol, HDL-cholesterol triglycerides) in up to 27,312 individuals with a comprehensive set of low frequency coding variants (ExomeChip), combined with conditional analysis in the known lipid loci. No new locus reached genome-wide significance. However, we found a new lead variant in 26 known lipid association regions of which 16 were >1000 fold more significant than the previous sentinel variant and not in close LD (6 had MAF < 5%). Furthermore, conditional analysis revealed multiple independent signals (ranging from 1-5) in a third of the 98 lipid loci tested, including rare variants. Addition of our novel associations resulted in between 1.5-2.5 fold increase in the proportion of heritability explained for the different lipid traits. Our findings suggest that rare coding variants contribute to the genetic architecture of lipid traits.
INTRODUCTION

Genome-wide association studies (GWAS) have identified hundreds of mainly common variants that are robustly associated with cardiometabolic traits (1-4). For lipid levels, a series of large-scale meta-analyses (N > 100,000) identified a total of 164 independent single nucleotide polymorphisms (SNPs) in 159 loci contributing to variation in plasma concentrations of total cholesterol (TC), triglycerides (TG), low-density lipoprotein cholesterol (LDL-C) and high-density lipoprotein cholesterol (HDL-C) (2, 3, 5-8). Blood lipid levels have an estimated heritability of 40-70% (9); however, variants reaching genome-wide significance explain only ~15% of the heritable fraction for these traits (2, 3). The clinical relevance of these 164 SNPs of which 71 associate with more than one lipid trait, is underscored by an overall excess of significant association signals for coronary artery disease (CAD), fasting glucose, type 2 diabetes, blood pressure traits and body mass index (BMI) among them (2).

It has been hypothesised that low frequency (1-5% minor allele frequency (MAF)) and rare (<1% MAF) variants with larger effects may account in part for the missing heritability in complex traits (10, 11). To test this hypothesis in relation to lipid traits we used the Illumina HumanExome Beadchip (ExomeChip), an array which provides comprehensive coverage of low frequency coding variants (nonsynonymous, splice-site and stop altering), to profile 27,312 individuals. The Exome array also includes most of the lead (or a proxy) GWAS variants in the 159 known lipid loci which allowed to assess the independent contribution of additional, mainly low frequency, coding variants in these loci by performing conditional analysis with the GCTA software.
RESULTS

Single-marker analysis

In our single-marker meta-analysis of ExomeChip (Illumina) data in 27,312 individuals (an overview of the study design is given in Figure 1), we did not find any new variant associated with a lipid trait at either a genome-wide threshold of significance \( P < 5 \times 10^{-8} \) or an array-wide threshold of significance \( P < 2 \times 10^{-7} \) outside the 159 previously reported loci (considering a 1Mbp window centered on the sentinel SNP).

The 159 unique loci known to be associated with one or more lipid traits represent 247 association signals (73 for HDL-C, 58 for LDL-C, 74 for TC and 42 for TG) (6, 8, 12-14). The ExomeChip array does not have a good proxy of the reported sentinel SNP (2, 3) in 21 of the 159 lipid loci (see Materials and Methods). In our study, we detected 209 association signals with a lipid trait (55 for HDL-C, 50 for LDL-C, 67 for TC and 37 for TG) at \( P < 0.01 \) (nominal significance; direction of effect same as published (2)) in 135 of the 159 unique lipid loci (Supplementary Table 1). Of the remaining 24 loci, 9 had no lead SNP or a proxy on the ExomeChip, 4 had the lead SNP or proxy fail QC, and 11 did not show a nominal association in our study.

Further assessing the results of our single-marker analysis, we found that in about half (n = 98) of the 209 association signals (26 for HDL-C, 17 for LDL-C, 35 for TC and 20 for TG), the lead SNP was either the published one or a highly linked proxy \( (r^2 > 0.8; \text{Supplementary Table 1}) \); in two instances the proxy is a putative functional variant (rs2792751 in \textit{GPAM} for LDL-C and rs35332062 in \textit{MLXIPL} for TG) (Supplementary Table 1). In many loci our top hit was different
than the previously published lead SNP (this study) and not in close LD ($r^2 < 0.8$) (Supplementary Table 1). Table 1 lists the 27 most significant of these associations ($P < 10^{-4}$) of which 8 are due to low frequency or rare coding variants. Interestingly, for 16 of these 27 association signals, the new sentinel variant was >1000 fold more significant than the previously published one. These results were also corroborated with one exception by conditional and joint analyses (see below); SNP rs5015480, a downstream variant in \textit{CYP26A1} previously reported for T2D (15), was not significant in the joint analysis. For the remaining 11 signals the new sentinel variant was only marginally more significant than the previously published one (includes two low frequency variants in \textit{LPA} and \textit{KDM2B}).

Among the 27 association signals (Table 1), four variants had not been previously associated with a lipid trait. Two of them had a 1000-fold more significant association for TG levels than the previous sentinel SNP: rs72836561 a missense variant in \textit{CD300LG} (p.R82C) and rs3094216 a synonymous variant (p.C448=) in \textit{CDSN}. The other two variants were rs3751813 (TG association) an intronic variant in the \textit{FTO} gene andrs34606562 (TC association) a synonymous change (p.L1174=) in \textit{KDM2B} both of which overlap a strong peak for H3K27Ac which marks active regulatory elements. In the \textit{LPA} locus, the missense variant rs3798220 (p.I1891M), previously associated with Lp(a) lipoprotein levels and CAD (16) had the strongest signal for LDL-C.

In 15 of the 135 lipid loci with an association signal in our data ($P < 0.01$), we lacked the published lead SNP or a proxy (12 not on the ExomeChip and 3 QC failures; Supplementary Table 1, column AJ) and therefore we were unable to undertake a direct comparison of the strength of association between our top hit and the published one. However, in one such locus, \textit{ABCA8}, which is associated with HDL-C (2, 3), we found a missense variant (rs77542162; Cys1319Arg; 1.57%
MAF) associated with both LDL-C ($P=6.40 \times 10^{-13}$) and TG ($P=6.23 \times 10^{-11}$) but not HDL-C ($P=0.46$) located in the *ATP-binding cassette, sub-family A (ABC1), member 6* (*ABCA6*) gene (Supplementary Table 1). *ABCA6* encodes a membrane-associated protein and is located together with *ABCA8* and three other *ABC1* family members on 17q24. *ABCA6* may play a role in macrophage lipid homeostasis (17) and in intercellular lipid transport processes in vascular endothelial cells (18).

*Conditional analysis*

We next undertook conditional analysis which looks for association signals that are independent of the lead SNP from the unconditional analysis, in the 135 unique loci harbouring 209 lipid association signals at a nominal significance level of $P<0.01$ (boundaries are listed in Supplementary Table 2) using the GCTA software (19) and meta-analysis summary statistics from all 27,312 samples. We considered a signal from the conditional analysis to be significant if it passed a Bonferroni correction threshold based on the number of SNPs tested across the locus examined (Supplementary Table 2). Therefore only loci which had a lead SNP with $P$ unconditional $<$ the locus-wide Bonferroni threshold for multiple testing (i.e. based on the number of tested SNPs per locus), were amenable to conditional analysis. Based on the threshold calculated for each locus (Supplementary Table 2), it was possible to examine 98 of the 209 lipid association regions for a secondary signal (see Materials and Methods; Supplementary Table 1). We found 31 (31.6%) of these association regions to have at least one additional independent signal. In total, we identified 89 independent signals (29 for HDL-C, 16 for LDL-C, 19 for TC and 25 for TG) in the 31 association regions (Table 2 and Supplementary Table 3) corresponding to the 31 sentinel SNPs
from the unconditional analysis and 58 SNPs from the subsequent rounds of conditional analysis. The largest number of independent signals per locus was five, in the \textit{APOA1} locus for HDL-C as well as in the \textit{APOB} and \textit{APOE} loci for LDL-C (the latter illustrated in Supplementary Figure 1). Approximately 30\% of the 89 independent variants were either low frequency (13; 14.6\%) or rare variants (14; 15.7\%), (Supplementary Table 1).

Out of the 58 additional signals identified from the conditional analysis, 42 have previously been associated with a lipid trait (34 reported for the investigated lipid trait and 8 for a lipid trait other than the investigated one; Supplementary Table 3) and 16 have not been previously associated with any lipid trait (Table 2). Among the 34 lipid signals previously reported, two variants, rs439401 in \textit{APOE} (TC) and rs35120633 in \textit{APOA1} (TG and HDL-C), showed an increase in their effect size after conditioning for the top hit in the corresponding region. In both instances, the very strong association signal of the lead SNP in the region (e.g. rs7412 $P=7.43\times10^{-145}$ in the \textit{APOE} locus) appears to partially mask the weaker secondary signal. In the unconditional analysis, SNP rs439401 had an effect size ($\beta$) of -0.051 per T allele ($P=1.77\times10^{-8}$) whereas after conditioning on rs7412 the $\beta$ doubled to -0.103 ($P=3.70\times10^{-31}$); this variant has been associated with HDL-C and TG as a bivariate phenotype (20). Similarly, the association of rs35120633 with TG and HDL-C (unconditional $P=2.67\times10^{-40}$ and $P=2.02\times10^{-9}$ respectively) became stronger after conditioning on rs2266788 ($P=5.21\times10^{-46}$ and $P=1.38\times10^{-10}$ respectively).

Of the 16 conditional signals not previously associated with a lipid trait, 10 are rare variants (Table 2). These variants are missense except rs76353203 (MAF 0.04\%; $\beta=-1.258$ per T allele) which introduces a stop codon in \textit{APOC3} (Arg19TER) and is known to cause hyperalphalipoproteinemia 2 (HALP2). At several loci, conditional analysis identified variants
with much larger effect sizes than the sentinel SNP e.g. missense variant rs116329129 (rs116329129:T>C, p.V280A; MAF 0.03%) in BANK1 which was associated with HDL-C, had a \( \beta \) of -1.835 per C allele compared to -0.104 for the lead variant rs13107325 (MAF 6.02%) which is located in SLC39A8. The B-cell scaffold protein with ankyrin repeats 1 (BANK1) gene encodes a B-cell-specific scaffold protein involved in B-cell receptor-induced calcium mobilization from intracellular stores. Variants in BANK1 have been associated with susceptibility to systemic lupus erythematosus (21). Similarly, the missense variant rs139788907 (MAF 0.03%) in the phospholipase A2, group IVF (PLA2G4F) gene (rs139788907:A>G, p.L326P; deleterious change per SIFT) which was associated with TG levels, had a \( \beta = 1.665; \sim 1.98 \text{ mmol/l per G allele} \) 10-fold higher than that of the intronic lead variant rs2412710 (MAF 1.8%; \( \beta = 0.165; 0.20 \text{ mmol/l per G allele} \) which is located in CAPN3. PLA2G4F encodes a calcium-dependent phospholipase A2 that selectively hydrolyzes glycerophospholipids in the sn-2 position.

**Joint analysis**

Regional analyses can determine the specific contribution that each locus makes to the trait heritability. The iterative rounds of conditional analyses within GCTA described above enabled the identification of independently associated variants at each locus. Joint analyses can simultaneously estimate the effects of each of these significant variants adjusted for all other effects.

The joint analyses were performed in both the discovery studies with appropriate ethical approval for sharing individual level data (16 of the 19 cohorts; N=24,894) and in the replication studies (4 cohorts; N=9,029), in order to (i) validate the original conditional analyses based on
GCTA regarding the handling of rare and low-frequency variants as well as check for any impact of sample size difference and (ii) confirm consistency between the discovery and replication studies, to ensure that the replication data is sufficiently concordant to be used for a risk score analyses (see below).

Over 95% of the variants that were significant in the conditional analyses also had $P<0.05$ in the joint analyses (29/29 for HDL-C; 17/18 for LDL-C; 18/19 for TC; and 18/20 for TG) (Supplementary Table 4). A comparison of the $\beta$s between conditional and joint analyses (Figure 2) revealed close agreement between the two analyses, with almost perfect directional consistency (55 out of 56 variants). Comparison of the p-values from each analysis (Figure 2) also showed close agreement for most variants despite a difference in sample size between the two analyses and the more conservative nature of the joint tests. Importantly, the relationship between p-values did not appear to depend on MAF, and none of the variants with noticeably discordant p-values was rare or of low frequency. Of the four variants with $P>0.05$ we excluded rs5015480 and rs2068888 (TG signals in CYP26A1) from further analyses but retained rs3208856 (APOE-LDL) and rs920915 (LIPC-TC) given that they are established associations. In summary, we found good concordance between the joint and conditional analyses within the discovery studies.

Joint analysis in the replication studies detected an effect in the same direction for 91.8% of the variants (Supplementary Table 4), showing good concordance between the discovery and replication data sets.

**Locus specific genetic score analysis**
Next we calculated an overall genetic risk score association for each region except \textit{CYP26A1} (see joint analysis above), assessing the combined effects of all independent variants within a locus. In such analyses, the score is weighted by the effect sizes of each included variant. We used the beta estimates from the conditional analyses as risk score weights and performed the analyses in the replication set which comprised four independent studies (N=9,029). It is paramount to use an independent data set in order to minimise any bias.

The genetic score analyses identified several strong effects (Table 3; effect estimates are from the unweighted model and are expressed as per 1-allele increment); for example, in the \textit{CETP} locus each trait-increasing allele associated with 12.4\% of an SD (~0.06 mmol/l) increase in HDL-C accounting for 3.1\% of the overall trait variation. We also note the \textit{PCSK9} locus in which each trait-increasing allele was associated with 19.2\% of an SD (~0.18 mmol/l) increase in LDL-C, but this region accounted for only 0.4\% of the variation. \textit{PCSK9} was also associated with a large effect on TC (16.6\% of an SD; 0.18 mmol/l per trait-increasing allele), and explained 0.3\% of the variation. For TG, the strongest effect was found at the \textit{APOA1} locus (17.2\% of an SD; ~0.20 mmol/l per trait-increasing allele accounting for 1.7\% of the variation). Cumulatively per trait, all regions tested accounted for 6.3\% (HDL-C), 2.9\% (LDL-C), 2\% (TC), and 3.8\% (TG) of the variation (Table 3).

\textit{Heritability}

First, we assessed heritability in the 135 unique known lipid loci which reached P < 0.01 in our study, considering only the published lead SNP (or proxy) and estimated a 7.12\% heritability
for HDL-C, 6.52% for LDL-C, 7.03% for TC and 6.31% for TG. When we considered for the same loci all independent sentinel SNPs from our study (lead and secondary signals as per Supplementary Table 1) we observed a between 1.5 and 2.5-fold increase in the heritability estimates (14.73% for HDL-C, 15.06% for LDL-C, 13.49% for TC and 9.62% for TG).

Finally, after exclusion of the CYP26A1 locus (2 variants) we assessed the incremental contribution of the multiple independent signals we detected by conditional analysis in the remaining 30 loci (87 in total) to heritability. Accounting for all signals per locus increased their contribution to heritability estimates for all lipid traits; 4.78% vs. 11.07% (HDL-C), 1.26% vs. 8.89% (LDL-C), 2.29% vs. 7.00% (TC), and 5.70% vs. 6.55% (TG) when comparing heritability estimates based on the known sentinel SNPs alone.
DISCUSSION

We undertook an association study in 27,312 individuals to test the hypothesis that low-frequency and rare coding variants contribute to the genetic architecture of the four main lipid traits, TC, TG, HDL-C and LDL-C explaining some of the missing heritability in large-scale genetic studies of common variants (2, 3). Of the 203,350 non-synonymous (missense, nonsense, splice-site, and frameshift) variants present on the ExomeChip, ~64,000 had a minor allele frequency above 0.1% to allow for single variant association testing. We did not find any new loci to be significant at the genome-wide level of significance in addition to the 159 loci known to be robustly associated \( (P<5\times10^{-8}) \) with plasma concentrations of these lipid traits. Our findings are in agreement with other recent studies that have used exome sequencing, exome arrays or 1000 Genome Project imputed GWAS studies to investigate circulating blood lipid levels or related traits (23-25) that have also not found new loci harbouring low frequency / rare coding variants with large effect sizes (23-26).

To extend our assessment of the impact of low frequency / rare coding variation on lipid levels, we also examined the 159 known lipid (2, 3) by assessing the results of both the single-marker and conditional analyses at these loci; for the latter we took advantage of the presence of previously reported index lipid-associated variant (or a good proxy) at 135 of these loci on the ExomeChip. We note that a recent study by the ENGAGE consortium (26) has identified an additional 10 unique loci associated with lipid traits but the Exome-chip does not harbour the sentinel SNP or a good proxy to allow conditional analysis (7 loci have a variant on the array reaching nominal significance; Supplementary Table 5). Interestingly, in 16 of the loci tested in our study we detected lead variants having an index signal at least 1000-fold more significant than the previously-reported sentinel SNP (Table 1); these included variants previously reported in the
literature for either the investigated and other lipid traits (10) or for lipid traits other than the investigated one (4) as well as two variants not previously associated with a lipid trait (rs3094216 and rs72836561). SNP rs3094216 (MAF 77.6%) is located in the CDSN gene, corneodesmosin, which encodes a protein found in human epidermis and other cornified squamous epithelia. Furthermore, rs3094216, is in strong LD with rs3095318 a missense variant in CDSN (p. M18L). Mutations in CDSN are known to cause peeling skin syndrome (PSS) type B disease, a rare recessive genodermatosis whereas a common synonymous SNP (rs1062470) has been associated with psoriasis (27). The other variant, rs72836561, is a low frequency missense variant in CD300LG (p.R82C; MAF 2.69%). CD300LG encodes the CD300 molecule–like family member G protein; a type I cell surface glycoprotein that contains a single immunoglobulin V–like domain and has a role in lymphocyte binding and transmigration.

In addition to rs72836561 (CD300LG) described above, two more low frequency or rare coding variants had not been previously associated with a lipid trait: the missense variant rs3798220 (p. I1891M) in LPA which was associated with LDL-C levels and rs34606562 a synonymous change (p.L1174) in KDM2B associated with TC levels. KDM2B encodes a member of the F-box protein family which is characterized by an approximately 40 amino acid motif, the F-box. The F-box proteins constitute one of the four subunits of ubiquitin protein ligase complex called SCFs (SKP1-cullin-F-box), which function in phosphorylation-dependent ubiquitination. KDM2B gene has been recently associated with methylation in adipose tissue and our lead variant (rs34606562) which overlaps a strong peak for H3K27Ac, is located 95.6 kb away of the methylation probe (cg13708645) used to detect it (28). In total, after taking in to consideration the results of the conditional analyses, we found 27 low frequency (13) or rare coding variants (14) to be associated with lipid levels. Interestingly, we observed higher effect sizes for variants with minor allele
frequency below 3% compared to more common SNPs (Supplementary Figure 2; power calculations based on total cholesterol showed 80% power to detect a minimum effect size of 0.125 at 1% MAF). However, this may reflect the fact that our study only had power to detect rare variants with higher effect sizes.

Overall, our study identified 14 missense variants not previously reported to be associated with a lipid trait. Two of them, rs3798220 and rs72836561, were new sentinel SNPs (Table 1) and the remaining 12 were identified as distinct additional signals through conditional analysis (rs116329129, rs138407155, rs140029729, rs200684324, rs117623631, rs5167, rs36053277, rs145814749, rs1132899, rs5742904, rs139788907, rs5880; Table 2). In total, 21 unique lipid loci (28 regions of association; all lipid traits) where we had the known sentinel SNP on the ExomeChip array had a missense variant as lead SNP (P < 10^{-4}) (Supplementary Table 1).

Our joint analyses validated the results from the conditional analyses showing that the GCTA software is also suitable for handling low-frequency variants, despite being designed for common variant analysis. We removed from further analyses one locus for TG, CYP26A1, based on the results of the joint analysis as both variants had Pj>0.05. This was partly due to using a random effect model for rs2068888 which showed significant heterogeneity. In the genetic score analyses, estimating the combined effect of genetic variants within each locus, we observed substantial overall effects on lipid traits in several loci (CETP, PCSK1, APOA1); for example, in the CETP locus each trait-increasing allele associated with 12.4% of an SD (~0.06 mmol/l) increase in HDL-C accounting for 3.1% of the overall trait variation.

As shown by others (26, 29-32), we found a substantial increase in the explained variance when we assessed heritability estimates based on the 209 variants (all traits) identified by both the unconditional and conditional analysis compared to the published lead SNPs in the corresponding
135 loci. In some loci the inclusion of new independent secondary signals contributes only marginally to heritability estimates. For example, in the APOE locus, 96.26% of the locus-specific heritability for TC was explained by rs7412 and rs769449 (72.41% and 23.84% respectively) which capture the \textit{APOE} 2/3/4 alleles. SNP rs7412 was the most significant variant for TC and LDL-C ($P=7.43\times10^{-145}$ and $P=1.43\times10^{-80}$ respectively) in our study whereas rs769449, a proxy of rs429358 ($r^2 0.82$), was the lead variant for HDL-C ($P=7.29\times10^{-13}$) and a secondary independent signal for TC and LDL-C. For LDL-C, these two variants explain 91.51% of the locus-specific heritability (70.95% and 20.56% respectively). Among the three additional secondary signals in the APOE locus for LDL-C, the low-frequency missense variant rs3208856 explained most of the remaining variance (7.55%). Overall the inclusion of low frequency / rare variants appear to significantly impact heritability estimates, for example, we observed a seven-fold increase in LDL-C variance explained cumulative when comparing only the loci that harboured secondary signals. But rare coding variants with large effect sizes are not likely to explain the overall missing fraction of the genetic component of lipid traits.

Some important limitations of our study merit to be highlighted. First, the list of tested coding variants is by no means exhaustive especially at the rare end of the frequency spectrum. Hamond and colleagues, estimated the Exome array to capture 72.5% and 66.2% of loss-of-function (LoF) and missense variation with MAF 0.5 and 0.1%, respectively (33). Second, our study does not have sufficiently high power to detect very low-frequency and / or rare variants with small effect sizes. Power calculations for our study (based on total cholesterol) showed that we had 80% power to detect a minimum effect size of 0.07 at a 3% MAF, 0.125 at 1% MAF, 0.4 at 0.1% and 1.25 at 0.01% MAF. Therefore, even larger sample sizes will be required to identify new rare variants with small effect sizes.
In conclusion, we demonstrate that low frequency / rare coding variants contribute to the genetic architecture and heritability of lipid traits despite a paucity of low frequency coding variants with large effect sizes.

MATERIALS AND METHODS

Samples and Phenotypes
We collected summary statistics for ExomeChip SNPs from 19 studies (N~26,000). Among these, 17 studies consisted primarily of individuals of European ancestry, and two studies consisted of individuals of South Asian descent (see Supplementary Note and Supplementary Table 6 for details). Both population-based studies and case-control studies were included; for case-control studies, cases and control samples were analysed separately. Results for blood lipid levels were provided in mmol/l units and trait residuals within each cohort were adjusted for age, age$^2$, and sex, and then inverse-rank normalized. Individuals known to be on lipid-lowering medication were excluded from the analysis (Supplementary Table 6).

Genotyping
A total of 247,870 genetic variants were genotyped using the Illumina ExomeChip array. The ExomeChip variants comprise 203,350 non-synonymous, 10,690 splice and 5,641 stop variants as well as 4,761 SNPs from the GWAS NHGRI catalogue. Genotypes were called with GenCall, subjected to QC (Supplementary Table 7) to remove poor quality samples and finally recalled using zCall, an algorithm optimised for rare variant detection (34). Average standard errors for
association statistics from each study were plotted against study sample size to identify outlier studies. Allele frequencies were inspected to ensure all analyses used the same strand assignment.

**Primary linear regression analysis**

Analyses were performed for each trait (HDL-C, LDL-C, TC and TG) using the assumption of an additive genetic model. Individual SNP association tests were performed using linear regression with the inverse normal transformed trait values as the dependent variable and the expected allele count for each individual as the independent variable. Explicit adjustments for population sub-structure using principal components (35) were carried out. These analyses were performed using a range of analytical software (Supplementary Table 7).

**Meta-analysis**

An inverse-variance weighted meta-analysis using a fixed effect model was performed, using both GWAMA (36) and METAL (37) and results were compared and checked for consistency. SNPs were excluded from the meta-analysis if they had MAF>5% and were absent in >90% of the samples or had MAF<5% and were absent in >25% of the samples and present in at least two studies and/or failed cluster plot evaluation. Heterogeneity was evaluated using Cochran’s Q- and I² statistic. For SNPs with non-significant heterogeneity (\(P\) for Q>0.01), we report the results from the fixed effect model whereas in the presence of significant heterogeneity (\(P\) for Q<0.01) we used a random effect model. Signals were considered to be novel if they reached a genome-wide significance (\(P<5\times10^{-8}\)) in the meta-analysis and were > +/- 500 kB away from the nearest previously described lipid locus. For the previously published lipid loci we considered replication
at nominal significance level of $P<0.01$. We note that for 28 loci the published lead SNP was not present on the ExomeChip (21) or were removed during QC (Supplementary Table 8).

**Approximate conditional analysis**

Conditional analysis was implemented in GCTA (19) using meta-analysis summary statistics from all 27,312 samples. A subset of 11,396 samples (part of the contributing studies: BC1958, BRIGHT, FIA3, EPIC and GoDARTS) of European origin was used as a reference panel for LD calculations. We considered in total 159 published lipid loci (2, 3) and 247 lipid association signals (73 for HDL-C, 58 for LDL-C, 74 for TC and 42 for TG). SNPs failing the cluster plot inspection were replaced by the next most significant SNP in the locus. Subsequent rounds of stepwise conditional analysis were performed in each locus until no significant SNP could be identified. The level of significance for each round of the conditional analysis was defined as $0.05/(\text{locus SNP content} - \text{conditional SNPs})$ to account for multiple testing (Supplementary Table 2).

**Joint analyses**

Joint analyses were performed for any loci identified in the conditional analyses as containing more than one statistically significant SNP. The joint tests estimated the associations between the phenotype and all statistically significant independent SNPs within a region simultaneously (by fitting one linear regression model per region).

**Locus specific genetic score analyses**

The genetic score analyses estimated the combined effect of all statistically significant SNPs within a region (Supplementary Table 2) by regressing a genetic score against the phenotype.
Genetic scores were derived in two ways: 1) by summing the number of trait-increasing alleles (as defined by the estimated directions of the SNPs effects in the conditional analyses) carried by each individual; and 2) by producing a weighted sum of the number of trait-increasing alleles against the phenotype (38). In this latter scenario, the genetic scores were weighted by multiplying genotypes by the corresponding estimated SNP effect (i.e. the “β”) from the conditional analysis. Joint tests and genetic score analyses were performed on the inverse-rank normalised trait values, which had been adjusted for age, age² and sex. Adjustments for principal components were also made, where applicable, to control for any potential population stratification within each study.

The joint analyses were run in a total of 20 cohorts (N_max=33,923). Of these, 16 (N=24,894) contributed to the individual SNP meta- and conditional analyses and were considered “discovery” cohorts, while a further 4 cohorts (N=9,029) that did not contribute to the preceding analyses were also included as “replication” cohorts. The genetic score analyses were only run in the replication cohorts in order to minimise bias, due to using weights estimated from the discovery meta-analyses. Only studies with unrelated individuals were included in these analyses. Studies with any missing data (i.e. where a SNP had been dropped during QC) within a particular region did not contribute to the overall result for that region.

Linear regression tests and genetic score analyses were conducted separately by each study. Meta-analyses were performed using the *metafor* package in R (39). Overall estimates of the proportion of variation explained by each region (“R²”) were derived by taking a weighted average over contributing studies (with weights based on sample size).

**Heritability**
Heritability estimates were calculated using the multifactorial liability threshold model (40). The calculations are performed using the inverse normal transformed traits meta-analysis results, based on a population SD of 1 and under the additive genetic model assumption. All variants included in the heritability calculations per trait were not in LD ($r^2<0.3$).

**URLs**

http://genome.sph.umich.edu/wiki/Exome_Chip_Design

http://www.metafor-project.org/ http://www.wvbauer.com

The results of the meta-analysis are available upon request and will be made available at

http://www.qmul.ac.uk/ExomeChip.Lipids.SummaryStatistics.zip
ACKNOWLEDGMENTS

1958BC: We are grateful for being able to use the British 1958 Birth Cohort DNA collection. Sample collection funded by the Medical Research Council grant G0000934 and the Wellcome Trust grant 068545/Z/02. Genotyping was funded by the Wellcome Trust. Professor Deloukas’ work forms part of the research themes contributing to the translational research portfolio of Barts Cardiovascular Biomedical Research Unit which is supported and funded by the National Institute for Health Research. Analysis was supported by BHF grant (Deloukas) RG/14/5/30893.

ASCOT: We thank all ASCOT trial participants, physicians, nurses, and practices in the participating countries for their important contribution to the study. The NIHR Barts Cardiovascular Biomedical Research Unit funded the genotyping and this forms part of the research themes contributing to the translational research portfolio. This work was supported by Pfizer, New York, NY, USA, for the ASCOT study and the collection of the ASCOT DNA repository; by Servier Research Group, Paris, France; and by Leo Laboratories, Copenhagen, Denmark. The study was investigator-led and was conducted, analyzed, and reported independently of Pfizer who funded the trial.

BRIGHT: The BRIGHT study is extremely grateful to all the patients who participated in the study and the BRIGHT nursing team. This work was facilitated by the National Institute for Research Barts Cardiovascular Biomedical Research Unit. This work was supported by the Medical Research Council of Great Britain (grant number G9521010D); and by the British Heart Foundation (grant number PG/02/128). Exome genotyping was supported by Wellcome Trust Strategic Awards 083948A and 085475. A. F.D. was supported by the British Heart Foundation (grant numbers RG/07/005/23633, SP/08/005/25115); and by the European Union Ingenious
HyperCare Consortium: Integrated Genomics, Clinical Research, and Care in Hypertension (grant number LSHM-C7-2006-037093).

**DIABNORD:** We are grateful to the study participants who dedicated their time and samples to these studies. We also thank the VHS, the Swedish Diabetes Registry and Umeå Medical Biobank staff for biomedical data and DNA extraction. We also thank M Sterner, M Juhas and P Storm for their expert technical assistance with genotyping and genotype data preparation. The current study was funded by Novo Nordisk, the Swedish Research Council, Påhlssons Foundation, the Swedish Heart Lung Foundation, and the Skåne Regional Health Authority (all to PWF).

**EFSOCH:** We are extremely grateful to the EFSOCH study participants and the EFSOCH study team. The opinions given in this paper do not necessarily represent those of NIHR, the NHS or the Department of Health. The EFSOCH study was supported by South West NHS Research and Development, Exeter NHS Research and Development, the Darlington Trust, and the Peninsula NIHR Clinical Research Facility at the University of Exeter. Timothy Frayling is supported by the European Research Council grant: SZ-245 50371-GLUCOSEGENES-FP7-IDEAS-ERC.

**EPIC:** The authors would like to acknowledge the contribution of the staff and participants of the EPIC Study. Sample collection was supported by the Medical Research Council UK and Cancer Research UK. Genotyping was funded by the Wellcome Trust.

**EPIC_incidentT2D:** The authors would like to acknowledge the contribution of the staff and participants of the EPIC-Norfolk Study. EPIC-Norfolk is supported by the Medical Research Council programme grants (G0401527, G1000143) and Cancer Research UK programme grant (C864/A8257).

**Fenland:** We are grateful to all the volunteers for their time and help, and to the General Practitioners and practice staff for assistance with recruitment. We thank the Fenland Study
Investigators, Fenland Study Co-ordination team and the Epidemiology Field, Data and Laboratory teams. Biochemical assays were performed by the National Institute for Health Research, Cambridge Biomedical Research Centre, Core Biochemistry Assay Laboratory, and the Cambridge University Hospitals NHS Foundation Trust, Department of Clinical Biochemistry. The Fenland Study is funded by the Medical Research Council (MC_U106179471).

**FIA3:**
We are indebted to the study participants who dedicated their time and samples to these studies. We also thank J Hutiainen and Å Ågren (Umeå Medical Biobank) for data organization and K Enquist and T Johansson (Västerbottens County Council) for technical assistance with DNA extraction. The study was supported in part by a grant from the Swedish Heart-Lung Foundation (grant no. 2020389 to PW Franks). J-H Jansson was responsible for the identification of MI cases. Genotyping was funded by the Wellcome Trust.

**GLACIER:** We are indebted to the study participants who dedicated their time and samples to these studies. We J Hutiainen and Å Ågren (Umeå Medical Biobank) for data organization and K Enquist and T Johansson (Västerbottens County Council) for technical assistance with DNA extraction. We also thank M Sterner, M Juhas and P Storm for their expert technical assistance with genotyping and genotype data preparation. The current study was funded by Novo Nordisk, the Swedish Research Council, Pålssons Foundation, the Swedish Heart Lung Foundation, and the Skåne Regional Health Authority (all to PWF).

**GoDARTS:** We are grateful to all the participants who took part in this study, to the general practitioners, to the Scottish School of Primary Care for their help in recruiting the participants, and to the whole team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists, and nurses. We acknowledge the
support of the Health Informatics Centre, University of Dundee for managing and supplying the anonymised data and NHS Tayside, the original data owner. The Wellcome Trust supported the Wellcome Trust UK Type 2 Diabetes Case Control Collection (Go-DARTS) and the Scottish Health Informatics Programme. The Chief Scientist Office supported informatics. Project funded by the UK Medical Research Council (G0601261). Genotyping was funded by the Wellcome Trust.

**GRAPHIC:** We are grateful to the participants of the GRAPHIC Study and to the nurses who recruited the patients. Recruitment for the GRAPHIC Study was funded by the British Heart Foundation. Exome array genotyping was funded by the NIHR and the Wellcome Trust. NJS holds a Chair funded by the British Heart Foundation and is a NIHR Senior Investigator. NM is funded by the NIHR Leicester Cardiovascular Biomedical Research Unit. The GRAPHIC Study is part of the portfolio of studies supported by the NIHR Leicester Cardiovascular BRU.

**HELICMANOLIS:** The MANOLIS cohort is named in honour of Manolis Giannakakis, 1978-2010. We thank the residents of Anogia and surrounding Mylopotamos villages for taking part. The HELIC study has been supported by many individuals who have contributed to sample collection (including Olina Balafouti, Christina Batzaki, Georgios Daskalakis, Eleni Emmanouil, Chrisoula Giannakaki, Margarita Giannakopoulou, Anastasia Kaparou, Vasiliki Kariakli, Stella Koinaki, Dimitra Kokori, Maria Konidari, Hara Koundouraki, Dimitris Koutoukidis, Eirini Mamalaki, Eirini Mpamiaki, Maria Tsoukana, Dimitra Tzakou, Katerina Vosdogianni, Niovi Xenaki), data entry (Thanos Antonos, Dimitra Papagrigoriou, Betty Spiliopoulou), sample logistics (Sarah Edkins, Emma Gray), genotyping (Robert Andrews, Hannah Blackburn, Doug Simpkin, Siobhan Whitehead), research administration (Anja Kolb-Kokocinski, Carol Smee) and
informatics (Martin Pollard, Josh Randall). HELICMANOLIS was funded by the Wellcome Trust (098051) and the European Research Council (ERC-2011-StG 280559-SEPI).

**HELICPomak:** We thank the residents of the Pomak villages for taking part. The HELIC study has been supported by many individuals who have contributed to sample collection (including Olina Balafouti, Christina Batzaki, Georgios Daskalakis, Eleni Emmanouil, Chrisoula Giannakaki, Margarita Giannakopoulou, Anastasia Kaparou, Vasiliki Kariakli, Stella Koinaki, Dimitra Kokori, Maria Konidari, Hara Koundouraki, Dimitris Koutoukidis, Eirini Mamalaki, Eirini Mpamiaki, Maria Tsoukana, Dimitra Tzakou, Katerina Vosdogianni, Niovi Xenaki), data entry (Thanos Antonos, Dimitra Papagrigoiour, Betty Spiliopoulou), sample logistics (Sarah Edkins, Emma Gray), genotyping (Robert Andrews, Hannah Blackburn, Doug Simpkin, Siobhan Whitehead), research administration (Anja Kolb-Kokocinski, Carol Smee) and informatics (Martin Pollard, Josh Randall). HELICPomak was funded by the Wellcome Trust (098051) and the European Research Council (ERC-2011-StG 280559-SEPI).

**KORA:** The authors thank Nadine Lindemann, Viola Maag and Franziska Scharl for excellent technical support. The KORA research platform (KORA, Cooperative Research in the Region of Augsburg) was initiated and financed by the Helmholtz Zentrum München - German Research Center for Environmental Health, which is funded by the German Federal Ministry of Education and Research and by the State of Bavaria. Furthermore, KORA research was supported within the Munich Center of Health Sciences (MC Health), Ludwig-Maximilians-Universität, as part of LMUinnovativ, by DZHK (German Centre for Cardiovascular Research) and the BMBF (German Ministry of Education and Research).

**LOLIPOP:** The LOLIPOP study is supported by the National Institute for Health Research (NIHR) Comprehensive Biomedical Research Centre Imperial College Healthcare NHS Trust, the
British Heart Foundation (SP/04/002), the Medical Research Council (G0601966,G0700931), the Wellcome Trust (084723/Z/08/Z) the NIHR (RP-PG-0407-10371), European Union FP7 (EpiMigrant, 279143) and Action on Hearing Loss (G51). We thank the participants and research staff who made the study possible. LOLIPOP was funded by the British Heart Foundation (SP/04/002), the Medical Research Council (G0601966,G0700931), the Wellcome Trust (084723/Z/08/Z), the NIHR (RP-PG-0407-10371), European Union FP7 (EpiMigrant, 279143).

**NFBC1986**: The authors are grateful to the NFBC1986 participants and their families and to the NFBC research staff for data collection. We also thank late Professor Paula Rantakallio (launch of NFBC1966 and 1986), Ms Outi Tornwall and Ms Minttu Jussila (DNA biobanking). Financial support was received from the Academy of Finland (project grants 104781, 120315, 129269, 1114194, Center of Excellence in Complex Disease Genetics and SALVE), University Hospital Oulu, Biocenter, University of Oulu, Finland (75617), the European Commission (EURO-BLCS, Framework 5 award QLG1-CT-2000-01643), NHLBI grant 5R01HL087679-02 through the STAMPEED program (1RL1MH083268-01), NIH/NIMH (5R01MH63706:02), the Medical Research Council, UK (G0500539, G0600705, PrevMetSyn/SALVE) and the Wellcome Trust (project grant GR069224), UK, ENGAGE project and grant agreement HEALTH-F4-2007-201413, and the EU Framework Programme 7 small-scale focused research collaborative project EurHEALTHAgeing 277849. The DNA extractions, sample quality controls, biobank up-keeping and aliquotting were performed in the National Public Health Institute, Biomedicum Helsinki, Finland and supported financially by the Academy of Finland and Biocentrum Helsinki.

**SCARF**: The authors would like to thank all participants in this study. The study was funded by the Foundation for Strategic Research, the Swedish Heart-Lung Foundation, the Swedish Research Council (8691, 12660, 20653), the European Commission (LSHM-CT-2007-037273), the Knut
and Alice Wallenberg Foundation, the Torsten and Ragnar Söderberg Foundation, the Strategic Cardiovascular and Diabetes Programmes of Karolinska Institutet and the Stockholm County Council, and the Stockholm County Council (560183). BS acknowledge funding from the Magnus Bergvall Foundation and the Foundation for Old Servants. MF acknowledge funding from the Swedish e-Science Research Center (SeRC).

CONFLICT OF INTEREST STATEMENT

The authors have no conflict to report.
REFERENCES


7 Sandhu, M.S., Waterworth, D.M., Debenham, S.L., Wheeler, E., Papadakis, K., Zhao, J.H.,
Song, K., Yuan, X., Johnson, T., Ashford, S. et al. (2008) LDL-cholesterol concentrations: a

8 Willer, C.J., Sanna, S., Jackson, A.U., Scuteri, A., Bonnycastle, L.L., Clarke, R., Heath,

9 Kaess, B., Fischer, M., Baessler, A., Stark, K., Huber, F., Kremer, W., Kalbitzer, H.R.,


12 Sandhu, M.S., Waterworth, D.M., Debenham, S.L., Wheeler, E., Papadakis, K., Zhao, J.H.,
Song, K., Yuan, X., Johnson, T., Ashford, S. et al. (2008) LDL-cholesterol concentrations: a

13 Teslovich, T.M., Musunuru, K., Smith, A.V., Edmondson, A.C., Stylianou, I.M., Koseki,
M., Pirruccello, J.P., Ripatti, S., Chasman, D.I., Willer, C.J. et al. Biological, clinical and
population relevance of 95 loci for blood lipids. *Nature*, **466**, 707-713.

14 Willer, C.J., Schmidt, E.M., Sengupta, S., Peloso, G.M., Gustafsson, S., Kanoni, S., Ganna,
A., Chen, J., Buchkovich, M.L., Mora, S. et al. Discovery and refinement of loci associated with


LEGENDS TO FIGURES

**Figure 1 legend:** Flow diagram of the study

**Figure 2 legend:** Comparison of results between conditional and joint analyses. Note that these figures do not include the lead SNP in each region (i.e. round > 0 means that variants from the conditional analysis are shown only), as the conditional analyses do not produce adjusted estimates of their effects in contrary to the joint analyses.
<table>
<thead>
<tr>
<th>Trait</th>
<th>Locus(Gene; Amino Acid change)</th>
<th>rsID</th>
<th>Trait-raising/Other allele</th>
<th>%EAF</th>
<th>N</th>
<th>β</th>
<th>SE</th>
<th>p-value</th>
<th>Published lead SNP</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>HDL</strong></td>
<td>LIPG (LIPG;N/S)</td>
<td>rs779603474,a</td>
<td>G/A</td>
<td>a</td>
<td>1.29</td>
<td>25373</td>
<td>0.289</td>
<td>0.040</td>
<td>6.42E-13</td>
</tr>
<tr>
<td></td>
<td>ANGPTL4(ANGPTL4;E/K)</td>
<td>rs116843064,b,c,d,e</td>
<td>A/G</td>
<td>b</td>
<td>1.88</td>
<td>24406</td>
<td>0.254</td>
<td>0.034</td>
<td>6.79E-14</td>
</tr>
<tr>
<td></td>
<td>LIPC</td>
<td>rs1800588</td>
<td>T/C</td>
<td>a</td>
<td>21.89</td>
<td>25375</td>
<td>0.138</td>
<td>0.011</td>
<td>1.12E-36</td>
</tr>
<tr>
<td></td>
<td>APOA1</td>
<td>rs2266788</td>
<td>A/G</td>
<td>a</td>
<td>91.66</td>
<td>25375</td>
<td>0.115</td>
<td>0.016</td>
<td>3.29E-12</td>
</tr>
<tr>
<td></td>
<td>TRIB1</td>
<td>rs2954033</td>
<td>G/A</td>
<td>a</td>
<td>70.47</td>
<td>25374</td>
<td>0.041</td>
<td>0.010</td>
<td>3.29E-05</td>
</tr>
<tr>
<td></td>
<td>APOE</td>
<td>rs769449</td>
<td>G/A</td>
<td>a</td>
<td>88.08</td>
<td>25375</td>
<td>0.100</td>
<td>0.014</td>
<td>7.39E-13</td>
</tr>
<tr>
<td><strong>LDL</strong></td>
<td>APOE (APOE;R/C)</td>
<td>rs7412</td>
<td>C/T</td>
<td>a</td>
<td>92.79</td>
<td>24361</td>
<td>0.561</td>
<td>0.029</td>
<td>1.43E-80</td>
</tr>
<tr>
<td></td>
<td>PCSK9(PCSK9;R/L)</td>
<td>rs11591147</td>
<td>G/T</td>
<td>a</td>
<td>98.35</td>
<td>24361</td>
<td>0.461</td>
<td>0.036</td>
<td>1.93E-37</td>
</tr>
<tr>
<td></td>
<td>CILP2</td>
<td>rs2304130</td>
<td>A/G</td>
<td>a</td>
<td>91.06</td>
<td>24360</td>
<td>0.099</td>
<td>0.016</td>
<td>9.01E-10</td>
</tr>
<tr>
<td></td>
<td>APOA1</td>
<td>rs2075290</td>
<td>C/T</td>
<td>a</td>
<td>8.26</td>
<td>27312</td>
<td>0.120</td>
<td>0.016</td>
<td>3.52E-14</td>
</tr>
<tr>
<td><strong>TC</strong></td>
<td>FRMD5(MAP1A;P/L)</td>
<td>rs55707100</td>
<td>T/C</td>
<td>a</td>
<td>2.60</td>
<td>26483</td>
<td>0.186</td>
<td>0.028</td>
<td>6.87E-11</td>
</tr>
<tr>
<td></td>
<td>APOE(APOE;R/C)</td>
<td>rs7412</td>
<td>C/T</td>
<td>b</td>
<td>92.73</td>
<td>27312</td>
<td>0.423</td>
<td>0.017</td>
<td>7.43E-14</td>
</tr>
<tr>
<td></td>
<td>PCSK9(PCSK9;R/L)</td>
<td>rs11591147</td>
<td>G/T</td>
<td>b</td>
<td>98.39</td>
<td>27312</td>
<td>0.406</td>
<td>0.034</td>
<td>5.42E-32</td>
</tr>
<tr>
<td></td>
<td>MPP3(CD300LG;R/C)</td>
<td>rs72836561</td>
<td>T/C</td>
<td>c</td>
<td>2.69</td>
<td>26484</td>
<td>0.133</td>
<td>0.027</td>
<td>1.21E-06</td>
</tr>
<tr>
<td></td>
<td>APOA1</td>
<td>rs2266788</td>
<td>G/A</td>
<td>a</td>
<td>8.34</td>
<td>26484</td>
<td>0.265</td>
<td>0.016</td>
<td>5.97E-61</td>
</tr>
<tr>
<td></td>
<td>WASF5_HLAB(CD SN)</td>
<td>rs3094216</td>
<td>A/G</td>
<td>c</td>
<td>77.60</td>
<td>26484</td>
<td>0.052</td>
<td>0.011</td>
<td>3.66E-06</td>
</tr>
<tr>
<td><strong>HDL</strong></td>
<td>ABCA1</td>
<td>rs3905000</td>
<td>G/A</td>
<td>a</td>
<td>86.23</td>
<td>25374</td>
<td>0.095</td>
<td>0.013</td>
<td>4.85E-13</td>
</tr>
<tr>
<td></td>
<td>MLXIPL</td>
<td>rs1178979</td>
<td>C/T</td>
<td>b</td>
<td>18.21</td>
<td>25375</td>
<td>0.055</td>
<td>0.012</td>
<td>3.10E-06</td>
</tr>
<tr>
<td></td>
<td>COBLL1</td>
<td>rs13389219</td>
<td>T/C</td>
<td>b</td>
<td>38.05</td>
<td>25374</td>
<td>0.037</td>
<td>0.009</td>
<td>9.09E-05</td>
</tr>
</tbody>
</table>

*Note: ρ is a measure of association and indicates the strength of the relationship between the two variables.*

**Table 1. New sentinel SNPs in known lipid loci (P<10⁻⁴) showing stronger association than the previously reported variant**

The table provides information on new sentinel SNPs and their associations with lipid traits, comparing them to previously reported lead SNPs. The table includes information on the trait, locus (gene and amino acid change), rsID, trait-raising/other allele, %EAF, N, β, SE, p-value, and status of publication. For HDL, LDL, and TC, specific loci and genes are highlighted, and the table compares the new lead SNPs with published lead SNPs in terms of their %EA, F, N, β, SE, and p-value. The table also indicates whether the new lead SNP is more significant than the published lead SNP.
<table>
<thead>
<tr>
<th>Gene</th>
<th>rs Number</th>
<th>VCF Chromosome</th>
<th>VCF Position</th>
<th>VCF Depth</th>
<th>VCF Frequency</th>
<th>Minor Allele</th>
<th>Major Allele</th>
<th>Z-Score</th>
<th>Minor Allele</th>
<th>Major Allele</th>
<th>Minor Allele</th>
<th>Major Allele</th>
<th>Z-Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>LDL</td>
<td>BRAP(SH2B3;W/R)</td>
<td>rs3184504</td>
<td>1q25</td>
<td>24361</td>
<td>0.040</td>
<td>0.010</td>
<td>3.28E-05</td>
<td>rs11065987</td>
<td>1q25</td>
<td>24361</td>
<td>0.036</td>
<td>0.010</td>
<td>2.04E-04</td>
</tr>
<tr>
<td>LDL</td>
<td>MYLIP(MYLIP;N/S)</td>
<td>rs9370867</td>
<td>1p35</td>
<td>24158</td>
<td>0.038</td>
<td>0.009</td>
<td>3.52E-05</td>
<td>rs3757354</td>
<td>1p35</td>
<td>24359</td>
<td>0.035</td>
<td>0.011</td>
<td>1.18E-03</td>
</tr>
<tr>
<td>LDL</td>
<td>LPA(LPA;LM)</td>
<td>rs3798220</td>
<td>1q32</td>
<td>24361</td>
<td>0.037</td>
<td>0.014</td>
<td>3.42E-05</td>
<td>rs1564348</td>
<td>1q32</td>
<td>24361</td>
<td>0.032</td>
<td>0.012</td>
<td>9.31E-03</td>
</tr>
<tr>
<td>TC</td>
<td>APOB</td>
<td>rs541041</td>
<td>19p13</td>
<td>27312</td>
<td>0.141</td>
<td>0.012</td>
<td>2.29E-33</td>
<td>rs1367117</td>
<td>19p13</td>
<td>27312</td>
<td>0.112</td>
<td>0.009</td>
<td>6.19E-33</td>
</tr>
<tr>
<td>TC</td>
<td>CETP</td>
<td>rs1532624</td>
<td>12q14</td>
<td>27311</td>
<td>0.049</td>
<td>0.009</td>
<td>5.88E-06</td>
<td>rs3764261</td>
<td>12q14</td>
<td>27312</td>
<td>0.041</td>
<td>0.009</td>
<td>1.03E-05</td>
</tr>
<tr>
<td>TG</td>
<td>FTO</td>
<td>rs3751813</td>
<td>19p13</td>
<td>24410</td>
<td>0.037</td>
<td>0.009</td>
<td>5.77E-05</td>
<td>rs1121980</td>
<td>19p13</td>
<td>26484</td>
<td>0.027</td>
<td>0.009</td>
<td>2.56E-03</td>
</tr>
<tr>
<td>TG</td>
<td>CYP26A1</td>
<td>rs5015480</td>
<td>11p15</td>
<td>26480</td>
<td>0.038</td>
<td>0.009</td>
<td>2.62E-05</td>
<td>rs2068888</td>
<td>11p15</td>
<td>26483</td>
<td>0.037</td>
<td>0.009</td>
<td>3.97E-05</td>
</tr>
</tbody>
</table>

**Publication status a:** previously published variants for the investigated and other lipid traits, **publication status b:** previously published variants for lipid traits other than the investigated one, **publication status c:** new variants (not published for any lipids trait)

- Pair wise LD estimation between the ExomeChip lead SNP and the published lead SNP was based on the reference panel used for the conditional analysis (n=11,396 samples).

1NHGRI catalogue, 2intron variant, 3independent secondary hit in the conditional analysis, 4nonsynonymous variant, 5missense variant, 6intergenic variant, 7synonymous variant, 83' prime UTR variant, 9common variant, 10non-coding transcript variant, 11downstream gene variant, 12splice region variant, 13upstream gene variant, 14nonsense-mediated mRNA decay transcript variant, 15regulatory region variant, 16non-coding exon variant, 17did not reach significance threshold in joint analysis (Pj > 0.05; see text)
Table 2. New variants identified as independent signals in the approximate conditional analysis

<table>
<thead>
<tr>
<th>Trait</th>
<th>Chromosome: Position</th>
<th>rsID</th>
<th>Trait-raising/Other allele</th>
<th>%EA</th>
<th>Single SNP meta-analysis</th>
<th>Approximate conditional meta-analysis</th>
<th>Locus</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>β</td>
<td>SE p-value</td>
<td>Conditioned on</td>
</tr>
<tr>
<td>HDL</td>
<td>4:102816487</td>
<td>rs116329129-4,4</td>
<td>T/C</td>
<td>99.97</td>
<td>1.835</td>
<td>0.448 4.23E-05</td>
<td>rs13107325</td>
</tr>
<tr>
<td></td>
<td>11:117112526</td>
<td>rs79554110-2,3</td>
<td>T/G</td>
<td>37.94</td>
<td>0.038</td>
<td>0.009 5.69E-05</td>
<td>rs140029729</td>
</tr>
<tr>
<td></td>
<td>11:116707044</td>
<td>rs138407155-4,4</td>
<td>A/T</td>
<td>99.88</td>
<td>0.109</td>
<td>0.273 1.88E-04</td>
<td>rs1805688, rs10468017</td>
</tr>
<tr>
<td></td>
<td>15:58830639</td>
<td>rs140029729-1,4</td>
<td>A/G</td>
<td>99.92</td>
<td>0.760</td>
<td>0.223 6.62E-04</td>
<td>rs1805688, rs10468017</td>
</tr>
<tr>
<td></td>
<td>15:58837989</td>
<td>rs200684324-1,4</td>
<td>A/G</td>
<td>0.13</td>
<td>0.389</td>
<td>0.116 8.36E-04</td>
<td>rs1805688, rs10468017</td>
</tr>
<tr>
<td></td>
<td>18:47113165</td>
<td>rs117623631-1,4</td>
<td>T/C</td>
<td>0.22</td>
<td>0.042</td>
<td>0.009 6.19E-06</td>
<td>rs77960347, rs4939883</td>
</tr>
<tr>
<td></td>
<td>19:45484865</td>
<td>rs5167-4,4,5</td>
<td>G/T</td>
<td>36.02</td>
<td>0.030</td>
<td>0.009 1.38E-03</td>
<td>rs769449</td>
</tr>
<tr>
<td></td>
<td>19:45028231</td>
<td>rs36053277-1,4</td>
<td>G/A</td>
<td>99.62</td>
<td>0.030</td>
<td>0.009 1.38E-03</td>
<td>rs769449, rs5167</td>
</tr>
<tr>
<td>LDL</td>
<td>6:161018174</td>
<td>rs7770628-3,8</td>
<td>C/T</td>
<td>45.47</td>
<td>0.030</td>
<td>0.009 1.38E-03</td>
<td>rs3798220</td>
</tr>
<tr>
<td></td>
<td>12:121660770</td>
<td>rs145814749-1,4</td>
<td>G/A</td>
<td>99.91</td>
<td>0.030</td>
<td>0.009 1.38E-03</td>
<td>rs34606562</td>
</tr>
<tr>
<td></td>
<td>19:45448036</td>
<td>rs1128299-4,5</td>
<td>T/C</td>
<td>47.49</td>
<td>0.030</td>
<td>0.009 1.38E-03</td>
<td>rs7412, rs769449, rs445925, rs3208856</td>
</tr>
<tr>
<td>TC</td>
<td>2:21229160</td>
<td>rs5742904-1,4</td>
<td>T/C</td>
<td>0.05</td>
<td>1.676</td>
<td>0.260 2.53E-11</td>
<td>rs541041, rs1367117, rs533617</td>
</tr>
<tr>
<td>TG</td>
<td>2:27550967</td>
<td>rs1049817-11</td>
<td>A/G</td>
<td>59.83</td>
<td>0.053</td>
<td>0.009 5.42E-09</td>
<td>rs1260326</td>
</tr>
<tr>
<td></td>
<td>11:116701355</td>
<td>rs76353203-3,10</td>
<td>C/T</td>
<td>99.96</td>
<td>1.258</td>
<td>0.320 8.60E-05</td>
<td>rs2266788, rs35120633</td>
</tr>
<tr>
<td></td>
<td>15:42436237</td>
<td>rs19789907-1,4</td>
<td>G/A</td>
<td>0.04</td>
<td>1.665</td>
<td>0.421 7.67E-05</td>
<td>rs2412710</td>
</tr>
<tr>
<td></td>
<td>16:57015091</td>
<td>rs5880-4,4</td>
<td>C/G</td>
<td>5.09</td>
<td>0.085</td>
<td>0.021 3.85E-05</td>
<td>rs3764261</td>
</tr>
</tbody>
</table>

1 nonsynonymous variant, 2 common variant, 3 intron variant, 4 missense variant, 5 nonsense-mediated mRNA decay transcript variant, 6 noncoding exon variant, 7 non-coding transcript variant, 8 NHGRI catalogue, 9 stop gained variant, 10 splice region variant, 11 synonymous

rs7770628 has been previously published for LPA eQTL

*change in the direction of the conditional effect
Table 3. Association results from the genetic score analyses estimating the combined effect of all statistically significant SNPs within a region by regressing a genetic score against each lipid trait

<table>
<thead>
<tr>
<th>Trait</th>
<th>Locus</th>
<th>rsIDs</th>
<th>N</th>
<th>β</th>
<th>SE</th>
<th>p-value</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>HDL</td>
<td>LPL</td>
<td>rs328, rs268, rs13702, rs1801177</td>
<td>9025</td>
<td>0.098</td>
<td>0.010</td>
<td>1.07E-21</td>
<td>0.010</td>
</tr>
<tr>
<td></td>
<td>ABCA1</td>
<td>rs3905000, rs1883025</td>
<td>9025</td>
<td>0.050</td>
<td>0.011</td>
<td>1.53E-05</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>APOA1</td>
<td>rs2266788, rs35120633, rs186808413, rs79554110, rs138407155</td>
<td>9025</td>
<td>0.060</td>
<td>0.013</td>
<td>8.90E-14</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>LIPC</td>
<td>rs1800588, rs10468017, rs140029729, rs200684324</td>
<td>7634</td>
<td>0.116</td>
<td>0.013</td>
<td>6.19E-18</td>
<td>0.010</td>
</tr>
<tr>
<td></td>
<td>CETP</td>
<td>rs3764261, rs5880, rs9939224, rs5882</td>
<td>9025</td>
<td>0.124</td>
<td>0.007</td>
<td>1.2E-75</td>
<td>0.031</td>
</tr>
<tr>
<td></td>
<td>LCAT</td>
<td>rs2271293, rs4986970</td>
<td>9025</td>
<td>0.064</td>
<td>0.020</td>
<td>1.15E-03</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LDL</td>
<td>ABCG5/8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>rs6756629, rs4245791</td>
<td>8697</td>
<td>0.057</td>
<td>0.014</td>
<td>3.87E-05</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>LPA</td>
<td>rs7770628, rs3798220</td>
<td>8697</td>
<td>0.037</td>
<td>0.016</td>
<td>4.31E-05</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>PCSK9</td>
<td>rs505151, rs11591147</td>
<td>8697</td>
<td>0.192</td>
<td>0.034</td>
<td>3.73E-14</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TC</td>
<td>APOE</td>
<td>rs1132899, rs3208856, rs445925, rs769449, rs7412</td>
<td>8697</td>
<td>0.116</td>
<td>0.013</td>
<td>5.13E-110</td>
<td>0.011</td>
</tr>
<tr>
<td></td>
<td>APOB</td>
<td>rs41288783, rs5742904, rs533617, rs541041, rs1367117</td>
<td>8697</td>
<td>0.099</td>
<td>0.011</td>
<td>1.57E-19</td>
<td>0.010</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ABCG5/8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>rs4245791, rs6756629</td>
<td>9029</td>
<td>0.057</td>
<td>0.013</td>
<td>2.09E-05</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>PCSK9</td>
<td>rs11591147, rs505151</td>
<td>9029</td>
<td>0.166</td>
<td>0.033</td>
<td>2.37E-11</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>APOA1</td>
<td>rs2075290, rs35120633</td>
<td>9029</td>
<td>0.091</td>
<td>0.022</td>
<td>5.66E-05</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>LIPC</td>
<td>rs1532085, rs1800588, rs920915</td>
<td>9029</td>
<td>0.017</td>
<td>0.014</td>
<td>1.73E-02</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>LIPG</td>
<td>rs4939883, rs77960347</td>
<td>9029</td>
<td>0.063</td>
<td>0.019</td>
<td>4.72E-04</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>APOE</td>
<td>rs7412, rs439401, rs769449, rs445925</td>
<td>9029</td>
<td>0.046</td>
<td>0.011</td>
<td>4.05E-57</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>APOB</td>
<td>rs541041, rs1367117, rs533617, rs5742904</td>
<td>9029</td>
<td>0.086</td>
<td>0.011</td>
<td>2.30E-18</td>
<td>0.008</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>LPL</td>
<td>rs15285, rs268, rs1801177, rs328</td>
<td>8729</td>
<td>0.111</td>
<td>0.011</td>
<td>3.26E-26</td>
<td>0.013</td>
</tr>
<tr>
<td></td>
<td>TRIB1</td>
<td>rs2954033, rs2954029</td>
<td>8729</td>
<td>0.049</td>
<td>0.009</td>
<td>7.72E-08</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TG</td>
<td>APOA1</td>
<td>rs76353203, rs35120633, rs7350481, rs2266788</td>
<td>8729</td>
<td>0.172</td>
<td>0.015</td>
<td>2.70E-37</td>
<td>0.017</td>
</tr>
<tr>
<td></td>
<td>LIPC</td>
<td>rs1800588, rs1532085</td>
<td>8729</td>
<td>0.030</td>
<td>0.012</td>
<td>9.63E-03</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>CETP</td>
<td>rs5880, rs3764261</td>
<td>8729</td>
<td>0.035</td>
<td>0.015</td>
<td>2.87E-02</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>GCKR</td>
<td>rs1049817, rs1260326</td>
<td>8729</td>
<td>0.077</td>
<td>0.018</td>
<td>3.32E-17</td>
<td>0.002</td>
</tr>
</tbody>
</table>

*Results taken from random-effects model (due to having p<0.01 for the Q statistic). Betas, standard errors (SEs) and R² estimates taken from the non-weighted models; betas are expressed as per 1-allele increment in the risk score, p-values estimated from the weighted models;
only regions with a p-value < 0.05 are presented. Note that R^2 estimates were synthesised by taking a weighted average over contributing studies (with weights based on sample size).
ABBREVIATIONS

MAF - Minor Allele Frequency
SNP - Single Nucleotide Polymorphism
GWAS - Genome-wide association studies
TC - Total Cholesterol
TG - Triglycerides
LDL-C - low-density lipoprotein cholesterol
HDL-C - high-density lipoprotein cholesterol
CAD - Coronary Artery Disease
CVD - Cardiovascular Disease
QC - Quality Control
LD - Linkage Disequilibrium