Pharmacogenetic meta-analysis of genome-wide association studies of LDL cholesterol response to statins


It is advisable to refer to the publisher’s version if you intend to cite from the work. http://dx.doi.org/10.1038/ncomms6068

For more information about UCLan’s research in this area go to http://www.uclan.ac.uk/researchgroups/ and search for <name of research Group>.

For information about Research generally at UCLan please go to http://www.uclan.ac.uk/research/

All outputs in CLoK are protected by Intellectual Property Rights law, including Copyright law. Copyright, IPR and Moral Rights for the works on this site are retained by the individual authors and/or other copyright owners. Terms and conditions for use of this material are defined in the http://clok.uclan.ac.uk/policies/
Pharmacogenetic meta-analysis of genome-wide association studies of LDL cholesterol response to statins


Statins effectively lower LDL cholesterol levels in large studies and the observed interindividual response variability may be partially explained by genetic variation. Here we perform a pharmacogenetic meta-analysis of genome-wide association studies (GWAS) in studies addressing the LDL cholesterol response to statins, including up to 18,596 statin-treated subjects. After the top SNPs were stratified, the meta-analysis included 13 studies (GWAS) in studies addressing the LDL cholesterol response to statins, including up to 18,596 statin-treated subjects.
The 3-hydroxymethyl-3-methylglutaryl coenzyme A (HMG-CoA) reductase inhibitors, also known as statins, are widely prescribed and are highly effective in the management and prevention of cardiovascular disease. Statin therapy results in a lowering of low-density lipoprotein cholesterol (LDL-C) levels by up to 55%1 and a 20–30% reduction of cardiovascular events2,3. Despite the clinical efficacy of statins in a wide range of patients2, interindividual variability exists with regard to LDL-C-lowering response as well as efficacy in reducing major cardiovascular events3. The suggestion that some of this variability may be due, in part, to common pharmacogenetic variation is supported by previous studies that have identified genetic variants associated with differential LDL-C response to statin therapy4,5.

A small number of genome-wide association studies (GWAS) have previously identified loci associated with statin response on a genome-wide level. A GWAS in the JUPITER trial identified three genetic loci, ABCG2 (rs2199936), LPA (rs10455872) and APOE (rs7412), that were associated with percentage LDL-C lowering effect compared with non-carriers. The second strongest association was with a SNP at chromosome 19, at APOE (rs445925, minor allele frequency (MAF) = 0.069, \( \beta = 0.041 \), s.e. = 0.006, \( P = 1.95 \times 10^{-11} \), Fig. 2b), indicating a 5.9% smaller LDL-C lowering per minor allele for carriers of the SNP compared with non-carriers. Associations at both loci have previously been described7,8. A third genome-wide significant association was found with a SNP at RICTOR on chromosome 5 (rs13166647, MAF = 0.230, \( \beta = -0.253 \), s.e. = 0.046, \( P = 4.50 \times 10^{-8} \)), although genotypes for this SNP were only available in two studies within the first stage (n = 2,144).

**Second-stage meta-analysis.** We selected 246 SNPs with \( P < 5 \times 10^{-8} \) from 158 loci for further investigation in three additional studies comprising up to 22,318 statin-treated subjects (see Methods; Supplementary Tables 1 and 5; Supplementary Note 3). This second stage confirmed the genome-wide significant associations between variations within the APOE and LPA loci and LDL-C response, as observed in the first stage (Table 1; Supplementary Fig. 2; Supplementary Table 5). In addition, SNPs at two new loci with \( P \) values between 6.70 \times 10^{-7} and 2.26 \times 10^{-6} in the first phase were shown to be significantly associated with statin-induced LDL-C lowering after statin treatment in the total combined meta-analysis at a genome-wide level: SORT1/CERS2/PSRC1 (rs646776, \( \beta = -0.013, \) s.e. = 0.002, \( P = 1.05 \times 10^{-9} \)) and LPA (rs10455872) and APOE (rs445925) and rs4420638) were associated with LDL-C response to atorvastatin treatment6.

A combined GWAS in three statin trials identified a SNP within CLMN (rs8014194) that is associated with the magnitude of statin-induced reduction in plasma cholesterol9. However, two other GWAS identified no genetic determinants of LDL-C response to statin therapy at a genome-wide significant level6,10.

On the basis of these studies, as well as previous candidate gene studies4,6, only the genetic variants that have been consistently identified to be associated with variation in LDL-C response to statin therapy, irrespective of statin formulation, are located at or nearby APOE and LPA. To determine whether additional loci may influence LDL-C response to statins, we formed the Genomic Investigation of Statin Therapy (GIST) consortium and conducted a pharmacogenetic meta-analysis using GWAS data sets from randomized controlled trials (RCTs) and observational studies. We identify two loci not previously identified in GWAS, SORT1/CERS2/PSRC1 and SLCO1B1. In addition, we confirm the associations within the APOE and LPA genes. These findings will extend the knowledge of the pharmacogenetic architecture of statin response.

**Results**

**First-stage meta-analysis.** The GIST consortium includes 6 RCTs (n = 8,421 statin recipients) and 10 observational studies (n = 10,175 statin recipients) that participated in the first stage (see Methods; Supplementary Tables 1 and 2; Supplementary Notes 1 and 2). To search for genetic variants associated with differential LDL-C response to statin therapy, each study independently performed a GWAS among statin users, using the difference between the natural log-transformed LDL-C levels on- and off-treatment as the response variable (see Methods).

The first-stage meta-analysis identified three loci, including 13 SNPs, that attained genome-wide significance (\( P < 5 \times 10^{-8} \)) for association with LDL-C response to statin treatment (Fig. 1; Table 1). The most significant association was for a SNP on chromosome 19, at APOE (rs445925, minor allele frequency (MAF) = 0.098, \( \beta = -0.043 \), s.e. = 0.005, \( P = 1.58 \times 10^{-18} \), Fig. 2a), indicating that carriers of the rs445925 SNP respond to statins with an additional 4.3% increase per allele in LDL-C lowering effect compared with non-carriers. The second strongest association was with a SNP at LPA on chromosome 6 (rs10455872, MAF = 0.069, \( \beta = 0.041 \), s.e. = 0.006, \( P = 1.95 \times 10^{-11} \), Fig. 2b), indicating a 5.9% smaller LDL-C lowering per minor allele for carriers of the SNP compared with non-carriers. Associations at both loci have previously been described7,8. A third genome-wide significant association was found with a SNP at RICTOR on chromosome 5 (rs13166647, MAF = 0.230, \( \beta = -0.253 \), s.e. = 0.046, \( P = 4.50 \times 10^{-8} \)), although genotypes for this SNP were only available in two studies within the first stage (n = 2,144).
LDL-C response to statins (rs445925) showed a small and non-
significant association with change in very small LDL (Table 2).
For the minor allele of rs2900478 (SLCO1B1), the borderline
significant association with smaller magnitude of LDL-C reduc-
tion showed a trend for preferential association with larger versus
smaller LDL subfractions. The lack of association of rs10455872
(LPA) with changes in LDL subfractions is consistent with evi-
dence discussed below that this locus affects levels of lipopro-
tein(a) (Lp(a)) and not LDL particles. Using generalized
estimating equations, we tested the association of log change in
each of the LDL subfractions with interactions of the four SNPs.
For very small LDL, the association with the rs646776 minor
allele was significantly different from that of the other minor
alleles ($P = 0.03$ after adjustment for multiple testing).

**Effects of off-treatment LDL-C.** To demonstrate that our find-
ings for LDL-C response to statin treatment are unlikely to be
explained through associations with baseline LDL-C levels, we
performed a number of additional analyses (see Methods). First,
Supplementary Table 7 shows regression coefficients for baseline-
adjusted and measurement noise-corrected estimates of the direct
effect of genotype on on-treatment LDL-C at the strongest SNPs
in the GIST meta-analysis ($P < 1 \times 10^{-8}$), which were available
in the CARDS data set. Correcting our effect size estimate further
and modelling measurement noise at baseline reduced the
apparent effect only slightly for all the markers, suggesting that
there is little effect of measurement noise. Next, within the
JUPITER trial, additional analyses were performed to determine
whether there was an interaction between LDL-C change and
statin or placebo allocation. Supplementary Table 8 shows sig-
nificant $P$ values for interaction ($< 5 \times 10^{-2}$) for SNPs at the
two genome-wide significant loci in the GIST meta-analysis, also
suggesting that genetic effects on baseline LDL-C as manifested in
the placebo group contribute at most only in part to genetic
effects on LDL-C response in the statin group.

**Genome-Wide Conditional Analysis.** To investigate whether
there were multiple SNPs within any gene and multiple loci
associated with differential LDL-C lowering to statin therapy, we
performed a conditional analysis across the genome using the
summary statistics of the combined meta-analysis. The results of
the Genome-Wide Conditional Analysis (GWCA; see Methods;
Supplementary Table 9) showed 14 SNPs independently asso-
ciated with statin response and these explained ~5% of the
variation in LDL-C response to statin treatment. Of the 14
independent SNPs, 6 were genome-wide significant in the com-
bined GWAS meta-analysis (Supplementary Table 5).

**Previous findings.** In Supplementary Table 10, we performed a
look-up in our GWAS meta-analysis for SNPs previously
described in the literature (NHGRI Catalogue$^{11}$ of Published
GWAS and Candidate gene studies) to be associated with statin
response, besides the loci associated at a genome-wide level in the
current study. None of these SNPs was associated with statin
response in our GWAS after correcting for multiple testing.

**Functional analyses.** Functional characterization of the 246 SNPs
selected for the second stage was performed using a range of
bioinformatics tools (see Methods). A total of 420 expression
quantitative trait loci (eQTL) associations were identified across a
wide range of tissues (Supplementary Data 1), which comprised 67
independent gene eQTL associations. Eleven genes, including
APOE, SORT1, CELSR2 and PSRC1, showed eQTLs in liver, which
considering its primary role in mediating statin-induced LDL
reduction may be particularly relevant to statin response. Putative
gene eQTLs were combined with genes annotated to variants in
linkage disequilibrium (LD) with LDL-C response-associated
variants, resulting in a list of 185 candidate gene loci, defined by
2,681 SNPs (Supplementary Data 2 and 3). To identify statin
response genes among the candidate loci, gene expression data
measured in response to statin treatment in a range of cell lines
was retrieved from the Connectivity Map resource$^{12}$ (see
Methods). Five genes (APOE, BRCAI, GRPELI, ADRB2 and
ETV1) showed convincing evidence of statin responsiveness on
the basis of greater than twofold differential expression in
response to statin treatment. Eight genes showed suggestive
evidence (1.5- to 2-fold change; TOMM40, SREBP1, PSRC1, BCL3,
BCAM, ANK3, SIVA1 and RANBP9; Supplementary Data 3).
Finally, involvement in statin response was investigated at a
pathway level using GeneGo Metacore (Thomson Reuters$^{13}$).
Brieﬂy, 87 literature-reported genes linked to statin response were
combined with the 185 candidate gene loci reported here

---

**Table 1 | Genome-wide significant associations in stage 1, stage 2 and combined meta-analysis.**

<table>
<thead>
<tr>
<th>Chr</th>
<th>Position</th>
<th>Lead SNP</th>
<th>Gene</th>
<th>Coding allele</th>
<th>Noncoding allele</th>
<th>Phase</th>
<th>N</th>
<th>Frequency-coding allele</th>
<th>Beta*</th>
<th>s.e.</th>
<th>% Extra reduction</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2109620053</td>
<td>rs646776</td>
<td>SOR/T/ CEILSR2/ PSRC1</td>
<td>C</td>
<td>T</td>
<td>Stage 1</td>
<td>16,697</td>
<td>0.230</td>
<td>−0.015</td>
<td>0.003</td>
<td>1.5</td>
<td>$6.70 \times 10^{-7}$</td>
</tr>
<tr>
<td>2</td>
<td>2109620053</td>
<td>rs646776</td>
<td>SOR/T/ CEILSR2/ PSRC1</td>
<td>C</td>
<td>T</td>
<td>Combined</td>
<td>16,697</td>
<td>0.230</td>
<td>−0.015</td>
<td>0.003</td>
<td>1.5</td>
<td>$6.70 \times 10^{-7}$</td>
</tr>
</tbody>
</table>

Chr, chromosome; SNP, single nucleotide polymorphism.
*Beta for difference between the natural log-transformed on- and off-treatment low-density lipoprotein cholesterol (LDL-C) levels adjusted for natural log-transformed off-treatment LDL-C, age-, sex- and study-specific covariates. The beta reflects the fraction of differential LDL-C lowering in carriers versus non-carriers of the SNP; a negative beta indicates a better statin response (stronger LDL-C reduction), a positive beta a worse statin response. Betas and P values were generated using linear regression analysis.

This percentage reflects the % extra LDL-C lowering in carriers versus non-carriers of the SNP.
A conservative network of direct interactions was constructed between query genes (Supplementary Data 3). The network included 24 genes located in the LDL-C-associated loci (Supplementary Fig. 4). Collectively, our functional and pathway analysis confirms a strong biological and functional role in statin response for several strongly associated gene loci, including APOE/TOMM40/PVRL2 and SORT1/CERS2/PSRC2.

Discussion

We have performed a meta-analysis of GWAS including more than 40,000 subjects, investigating genetic variants associated with variation in LDL-C lowering on statin treatment independent from associations with baseline LDL-C. We identified four loci at genome-wide significance, including the previously identified APOE and LPA, and the novel GWAS loci SORT1/CERS2/PSRC1 and SLC01B1.

Nine SNPs in the APOE gene region reached genome-wide significance for LDL-C response. The minor allele of the lead SNP rs445925, which is a proxy for the apoE e2 protein variant defining SNP rs7412 (ref. 14), was associated with a larger LDL-C-lowering response to statins compared with carriers of the major allele. The magnitude and direction of the effect size was similar to previously reported findings for the rs445925 variant in...
the GWAS study performed in CARDS and ASCOT\textsuperscript{8} and of the SNP rs7412 in JUPITER\textsuperscript{2}. Since the apoE c2 protein results in increased hepatic cholesterol synthesis, it may also predispose to stronger inhibition of cholesterol synthesis by statin treatment\textsuperscript{8,10}.

Three independent SNPs at \textit{LPA} were significantly associated with LDL-C response to statins. The minor G allele of the lead SNP rs10455872 was associated with smaller LDL-C reduction than the major allele. This result was similar to the previous GWAS findings for this SNP in the JUPITER trial and the combined ASCOT and CARDS study\textsuperscript{2,8}. The rs10455872 SNP was strongly associated with the KIV-2 copy number variant in Lp(a), which encodes variability in apo(a) size and is responsible for \~30% of variance in Lp(a) levels\textsuperscript{8,15}. Furthermore, rs10455872 was shown to be strongly associated with plasma Lp(a) levels\textsuperscript{16}. Standard assays of LDL-C, as well as the Friedewald formula, include cholesterol that resides in Lp(a)\textsuperscript{8,17}. Carriers of this LPA variant are characterized by higher Lp(a) levels and a larger proportion of their measured LDL-C resides in Lp(a) particles\textsuperscript{8,17}. Since statin therapy does not reduce the number of Lp(a) particles\textsuperscript{5}, their presence attenuates the measured LDL-C response to statins.

Two SNPs at \textit{SORT1}/\textit{CELSR2}/\textit{PSRC1} (rs646776 and rs12740374) on chromosome 1p were associated with an enhanced statin LDL-C response. A similar association was previously observed in a large candidate gene study in HPS\textsuperscript{5}; however, we demonstrate this finding now first at a genome-wide significance level. The minor allele of rs12740374 has been shown to generate a binding site for the transcription factor CEBPB\textsuperscript{18}. Transcription results in upregulation of hepatic expression of three genes at this locus, \textit{SORT1}, \textit{CELSR2} and \textit{PSRC1} (ref. 18), which we also showed in our eQTL analysis (Supplementary Data 1). Of these, \textit{SORT1} is most notable, in that it encodes the multifunctional intracellular trafficking protein sortilin, which has been shown to bind tightly to apoB\textsuperscript{19}. Sortilin-induced lowering of plasma LDL-C results from two mechanisms: reduced secretion of apoB-containing precursors, and, perhaps of greater importance, increased hepatic LDL uptake via binding to sortilin at the cell surface, with subsequent internalization and lysosomal degradation\textsuperscript{19}. Notably, the minor allele of rs646776 is preferentially associated with lower levels of small and very small LDL (Table 2), suggesting that sortilin is of particular importance for regulating levels of these particles\textsuperscript{18}. Smaller LDL subfractions have been shown to be relatively enriched in particles with reduced LDL receptor binding affinity and cellular uptake\textsuperscript{20}, a property that may contribute to their associations with increased risk for cardiovascular disease\textsuperscript{21}. This property may also underlie the diminished efficacy of statins for reduction of these particles (Supplementary Fig. 3)\textsuperscript{22}, since statins act to reduce LDL-C levels to a large extent by increasing LDL receptor expression as a result of upregulation of the transcription factor SREBP2, whereas \textit{SORT1} is not regulated by this mechanism. Hence, the greater statin-mediated reduction of LDL-C among carriers of the rs646776 minor allele could be attributed to relative depletion of LDL particles dependent on sortilin for clearance and hence a residually greater proportion of those LDL particles whose uptake is more dependent on the LDL receptor than on sortilin.

Notably, the strong association of rs646776 with statin-induced reductions in small and very small LDL particles contrasts to the weaker associations of changes in these particles with rs445925, likely the result of differing mechanisms underlying the effects of these SNPs on statin response. As noted above, rs445925 is a proxy for the SNP defining the apoE c2 protein variant that is thought to predispose to heightened statin response as a result of greater statin inhibition of cholesterol synthesis and hence upregulation of SREBP and LDL receptor activity.

The \textit{SLCO1B1} rs2900478 minor allele was associated with a smaller LDL-C reduction in response to statin treatment. \textit{SLCO1B1} encodes the organic anion-transporting polypeptide OATP1B1 and facilitates the hepatic uptake of statins\textsuperscript{23}. SNP rs2900478 is in strong LD ($r^2 = 0.89$) with rs4149056, which represents the Val174Ala substitution resulting in complete loss of function. In the HPS trial, which used simvastatin, this candidate gene SNP was associated with a 1% lower LDL-C reduction per allele\textsuperscript{23}. Single-dose studies have shown that the observed area under the curve of plasma level of active simvastatin after a dose of 40 mg was 221% higher in rs4149056 CC homozygotes compared with rs4149056 TT homozygotes, as compared with atorvastatin 20 mg (144% higher for CC versus TT) and rosuvastatin 40 mg (117% higher for CC versus TT)\textsuperscript{24}. This finding results from the slower hepatic uptake of statins caused by the genetic variant, which would also be expected to result in a reduction in the cholesterol-lowering effect\textsuperscript{25}. In a GWAS of the genetic risk factors for simvastatin-induced myopathy, \textit{SLCO1B1} showed the strongest association\textsuperscript{25}. Homozygous carriers of the \textit{SLCO1B1} variant had a 16.9 times higher risk for myopathy compared with non-carriers. This might have led to a decrease in study medication adherence, and consequently a decreased effect on LDL-C in carriers of this SNP.

In addition, previous analysis in the GoDARTS study showed that the effect of the \textit{SLCO1B1} gene on statin efficacy was abolished after removal of individuals who showed signs of intolerance\textsuperscript{26}. GWCA identified three independent loci in the \textit{APOE} gene region and two loci in the \textit{LPA} gene region (Supplementary Table 9). GWCA also showed several other loci with $P<5 \times 10^{-8}$ that were not GWAS significant on single-SNP analysis (HGD, RNF175, ISCA1-HTRA1, GLI3-SLC1A1, LOC100128657, NKX2-3-SLC25A28 and PEL2). These findings will require replication in independent, larger data sets. The significant SNPs in the GWCA analysis explained \~5% of the variation in LDL-C response to statin treatment. Whether this 5% is clinically relevant should be investigated by other studies. For example, it would be of interest to investigate whether this differential LDL-C lowering is also associated with differential event reduction by statin treatment.

In the current study, we combined the results of 6 randomized clinical trials and 10 observational studies in the first stage. This approach resulted also in combining several types of statins, since different statins were studied in the trials and within the observational studies (Supplementary Table 2). This, and the variation in statin dosage during follow-up for an individual, is a limitation of the current study, since, for example, the impact of the \textit{SLCO1B1} variant on statin pharmacogenetics is known to be highly dependent on statin type and dose\textsuperscript{24,27}. To overcome this limitation, the individual study analyses were adjusted for statin dose. Dividing the actual statin dose given by the statin-specific dose equivalent (Supplementary Table 3) gives the statin-adjusted range of doses used. To correct for between-study variance, we used a fixed effect meta-analysis with inverse variance weighting. Since we observed that the \textit{SLCO1B1} gene was genome-wide significantly associated with LDL lowering, this highlights the thoroughness of our analytical approach, in which the analyses were correctly adjusted for the type and dose of statins used (Supplementary Table 3). Moreover, a comparison of the estimates of the SNPs between the RCTs (where there are no intra-individual differences in dosages) with the estimates of the SNPs in the observational studies showed large homogeneity between the estimates in the various study designs (Supplementary Fig. 2), indicating that our adjustment for dosage seems to be sufficient within this study.
Another possible limitation of the current study is the influence of the identified genetic variants on baseline LDL-C levels. In pharmacogenetic studies investigating the LDL-C-lowering response to statins, it is important to eliminate the effect of association between the genetic variant and baseline LDL-C levels, since those findings may confound the response to treatment associations. Previous large GWAS studies have shown strong associations between baseline LDL-C levels and genetic variants in \textit{SORT1/CERS2/PSPC1}, \textit{APOE}, and \textit{LPA} \cite{3}. To eliminate those possible confounding effects, our response to treatment analyses were adjusted for baseline LDL-C levels. In addition, additional analysis in CARDS and JUPITER suggests no or little influence of genetic associations with baseline LDL-C on the genetic effects on LDL-C-lowering response.

In conclusion, this study is the largest meta-analysis of GWAS for LDL-C response to statin therapy conducted to date. Our results demonstrate that apart from the previously identified \textit{APOE} and \textit{LPA} loci, two new loci, \textit{SORT1/CERS2/PSPC1} and \textit{SLCO1B1}, also have a modest but genome-wide significant effect on LDL-C response. The minor alleles of the \textit{APOE} rs445925 and \textit{SORT1/CERS2/PSPC1} rs646775 SNPs were associated with a larger statin response, whereas the minor alleles of the \textit{LPA} rs10455872 and \textit{SLCO1B1} rs2900478 SNPs were associated with a smaller statin response. Our findings advance the understanding of the pharmacogenetic architecture of statin response.

Methods

\textbf{Study populations.} The meta-analysis was conducted in the GIST consortium, which includes data from 8 randomized controlled statin trials (RCTs) and 11 prospective, population-based studies. The initial analysis (first stage) was performed in 8,421 statin-treated subjects from 6 RCTs (ASCOT, CARDS, CAP, PRINCE, PROSPER and TONT) and 10,175 statin-treated subjects from 10 observational studies (AGES, ARIC, BioVU, CHS, FHS, GoDARTS I, GoDARTS II, Health ABC, IVH and MESA). Further investigation (second stage) was performed in 21,975 statin-treated subjects from two randomized trials (HPS and JUPITER) and one observational study (Rotterdam Study). Six SNPs were additionally genotyped in the Scandinavian participants of the ASCOT study. The details of the first- and second-stage studies can be found in the Supplementary Tables 1 and 2 and Supplementary Notes 1 and 2.

\textbf{Subjects.} Response to statin treatment was studied in statin-treated subjects only and not in those treated with placebo. Subjects included in the observational studies’ analysis should be treated with statins and have LDL-C measurements before and after start of statin treatment. Subjects of reported or suspected non-European ancestry were excluded. All participants gave written informed consent and the study was approved by all institutional ethics committees.

\textbf{Outcome measurements.} The response to statin treatment was defined as the difference between the natural log-transformed on- and off-treatment LDL-C levels. The beta of the corresponding regression thus reflects the fraction of differential LDL lowering in carriers versus non-carriers of the SNP. For observational studies, the on-treatment LDL-C levels were taken into account for all kinds of prescribed statins, at any dosage, for any indication and for at least 4 weeks before measurement. Characteristics of on- and off-treatment LDL-C levels and statins used in each study are shown in Supplementary Table 2. For each individual, at least one off-treatment LDL-C measurement and at least one on-treatment LDL-C measurement were required. When multiple on- or off-treatment measurements were available, the mean of the cholesterol measurements was used. Subjects with missing on- or off-treatment measurements were excluded, with the exception of the GoDARTS cohorts for which missing off-treatment LDL-C levels were estimated using imputation methods (Supplementary Note 2). In the HPS, proportional LDL-C response was defined by the changes in natural log lipid levels from the screening visit before starting statin therapy to the randomization visit \cite{2}.

\textbf{Genotyping and imputation.} Genotyping, quality control, data cleaning and imputation were performed independently in each study using different genetic platforms and software as outlined in Supplementary Table 4. In all studies, genotyping was performed using Illumina, Affymetrix or Perlegen genotyping arrays, and MACH, Impute or BIMBAM software was used for imputation.

\textbf{GWAS analysis.} Each study independently performed the GWAS on the difference between natural log-transformed on- and off-treatment LDL-C levels. To control for possible associations with off-treatment LDL-C levels, analyses were adjusted for the natural log-transformed off-treatment LDL-C level. An additive genetic model was assumed for all studies tested using a linear regression model. For observational imputed SNPs, regression analysis was performed on expected allele dosage. Analyses were additionally adjusted for age, sex and study-specific covariates (for example, ancestry principal components or country). Analyses in the observational studies were, if available, additionally adjusted for the statin dose by the natural logarithm of the dose equivalent as defined in Supplementary Table 3. This table shows the dose equivalent per statin type; dividing the statin dosage of an individual by the dose equivalent shown in Supplementary Table 3 will give the adjusted statin dosage.

\textbf{Quality control and meta-analysis.} Centrally, within each study, SNPs with MAF < 1% or imputation quality < 0.3 were excluded from the analysis. QQ-plots were assessed for each study to identify study-difference (Fig. 1). The software package METAL was used for performing the meta-analysis (http://www.sph.umich.edu/csg/abecasis/Metal/index.html). A fixed effects, inverse variance weighted approach was used. Using an inverse variance weighted meta-analysis will give smaller weights to studies with large s.e., To correct for possible population stratification, genomic control was performed by adjusting the within-study findings and the meta-analysis results for the genomic inflation factor.

\textbf{Second stage.} SNPs with \( P \) values < \( 5 \times 10^{-8} \) in the first-stage meta-analysis were selected for further investigation in a second stage. A maximum of two SNPs per locus were selected, based on statistical significance, except for the APOE locus, for which all genetic variants included were selected for validation. A total of 6 SNPs, within 1 independent loci, were included in the second stage, which was performed in the JUPITER trial, HPS study and the Rotterdam Study, which all had GWAS data and response to statin treatment available. For 2 of the 246 SNPs, a proxy was used in the JUPITER trial, and 31 SNPs were not available, nor was a proxy SNP. HPS provided data on 151 directly genotyped SNPs from the second stage, including 48 of the 6 SNPs and 103 proxy SNPs (rs\( r^2 > 0.8 \)). Analysis in HPS was not adjusted for baseline LDL-C levels. In addition, the number of subjects with data varied from SNP-to-SNP and ranged from \( \sim 4,000 \) for variants with GWAS data to \( \sim 18,000 \) for some candidate genes. Results of the first and second stage were combined using fixed effects, with allele weighted meta-analysis and analysed by the meta stage. Six SNPs with \( P \) values \( 5 \times 10^{-8} < P < 5 \times 10^{-7} \) in the combined meta-analysis were selected for additional genotyping in the Scandinavian participants of the ASCOT study. Kaspar assays were designed for four of the SNPs using the KBioscience Primmerpick software, and oligos were provided by Integragen DNA technologies (http://eu.idtdna.com/site). Full Kaspar methodology is available from LGC SNP genotyping (http://www.lgcgenomics.com/genealogy/kaspar-genotyping-reagents/). Two SNPs (rs981844 and rs13166647) were genotyped using Taqman assays supplied by Life Technologies (http://www.lifetechnologies.com/en/home.html) using the standard Taqman protocol. Results of the additional genotyping were combined with results from the first and second stages using a fixed effects, inverse variance weighted meta-analysis and analysed by METAL.

\textbf{Determination of changes in LDL subfractions.} LDL subclasses were analysed as described previously \cite{20} using non-denaturing gradient gel electrophoresis of fasting plasma samples taken at baseline and after 6 weeks of simvastatin 40 mg per day (CAP study, \( n = 579 \)) or 12 weeks of pravastatin 40 mg per day (PRINCE study, \( n = 1,284 \)). Aliquots of 3.0 ml of whole plasma were mixed 1:1 with a sampling buffer of 20% sucrose and 0.2% bromophenol blue. Electrophoresis of samples and size calibration standards was performed using 2–14% polyacrylamide gels at 150 V for 3 h following a 15 min pre-run at 75 V. Gels were stained with \( 0.07 \% \) Sudan black for 1 h and stored in 0.81% acetic acid, 4% methanol solution until they were scanned by computer-assisted densitometry for determination of areas of LDL IVa (22.0–23.2 nm), LDL IVb (24.2–24.6 nm), LDL IIA (24.7–25.3 nm) and LDL IIB (25.6–26.4 nm). LDL IIA (26.5–27.1 nm) and LDL I (27.2–28.5 nm). The cholesterol concentrations of the subfractions (mg dl \(^{-1} \) plasma) were determined by multiplying percent of the total stained LDL area for each subfraction by the LDL I for that sample. For genetic association analyses, subfractions were grouped into large LDL (LDL I + IIA), medium LDL (LDL IIB), small LDL (LDL IIB) and very small LDL (LDL IIB + IVA + IVb) as described previously \cite{21}. A generalized estimating equation method was used to test the association of log change with the interaction of the four SNPs by LDL subfraction.

\textbf{Effect of off-treatment LDL-C.} Effects of genetic variation on treatment response as measured by on-treatment LDL-C could be mediated through effects on the off-treatment LDL-C. To evaluate whether genetic on-treatment LDL-C likely reflects residual effect on off-treatment LDL-C, it is necessary to adjust for the off-treatment LDL-C levels and to correct the maximum likelihood estimate of the adjusted effect of genotype on on-treatment value for the noise in off-treatment values (the noise is both random measurement error and intra-individual variation in usual LDL-C). This analysis was only performed in CARDS (Supplementary baseline measurements were available. From the rules of path analysis, we calculated the direct effect of genotype on an on-treatment trait value as \( \beta = \alpha(1 - \rho) \), where...
$\beta$ is the coefficient of regression for on-treatment trait value on genotype adjusted for measured off-treatment value, $s$ is the coefficient of regression of baseline LDL on on-treatment trait value and $\gamma$ is a within-person correlation including possible replicate measurements of on-treatment values and $\delta$ is the coefficient of regression for on-treatment trait value on observed off-treatment value. For these calculations, we used $p = 0.8$ as a plausible value for the intraclass correlation based on the within-person correlation in LDL-C values taken over two off-treatment visits in CARDS. The interaction of candidate locus with statin versus placebo allocation was assessed in the JUPITER trial, since this study was not involved in the first-stage meta-analysis. Regression models were applied to the combined population of statin- and placebo-treated subjects by including extra terms encoding placebo allocation and the product of placebo allocation with SNP minor allele dose.

GWCA using Genome-Complex Trait Analysis. There may be multiple causal variants in a gene and the total variation that could be explained at a locus may be underestimated if only the most significant SNP in the region is selected. To identify independent SNPs, we ideally can perform a conditional analysis, starting with the top associated SNP, across the whole genome followed by a stepwise procedure of selecting additional SNPs, one by one, according to their conditional $P$ values. This would allow the discovery of more than two associated SNPs at a locus. To identify independent SNPs across the genome-wide data, we used an approximate conditional and joint analysis approach implemented in Genome-Complex Trait Analysis (GCTA) software (http://www.complextraitgenomics.com/software/gcta/). We used summary-level statistics from the first- and second-stage GWAS meta-analysis and LD corrections between SNPs estimated from CARDS GWAS data. SNPs on different chromosomes or more than 10 Mb distant are assumed to be in linkage equilibrium. The model selection process in GCTA starts with the most significant SNP in the single-SNP meta-analysis across the whole genome with $P < 5 \times 10^{-8}$. In the next step, it calculates the $P$ values of the remaining SNPs conditional on the top SNP that have already been selected in the model. To avoid problems due to colinearity, if the squared multiple correlations between a SNP to be tested and the selected SNP(s) is larger than a cut-off value, such as 0.9, the conditional $P$ value for that SNP will be set to 1. Select the SNPs with minimum conditional $P$ value that is lower than the cut-off $P$ value. Fit all the selected SNPs jointly in a model and drop the SNPs with the $P$ value that is greater than the cut-off $P$ value. This process is repeated until no SNPs can be added or removed from the model.

Pathway analysis and construction of a statin response network. Genes showing evidence of association (based on direct association or LD (HapMap CEU $r^2 > 0.8$)) were reviewed for evidence of involvement in statin response at a pathway level using GeneGo MetaCore (Thomson Reuters portal.genego.com). A statin response pathway network was used in two stages. First, all genes with a literature-reported involvement in statin response (based on Medical Subject Headings (MeSH)) were identified using GeneGo MetaCore (Supplementary Data 3). Second, these genes were combined with all genes in associated loci (including genes in LD and a marker was constructed based on direct interactions only, including direct interactions only, we created a conservative network of direct gene interactions that have been consistently linked to statin response in the literature.

eQTL analysis. LDL-C-associated index SNPs (246 SNPs) were used to identify 1,443 LD proxy SNPs displaying complete LD ($r^2 = 1$) across four HapMap builds in European ancestry samples (CEU) using the SNP tool (http://www.broadinstitute.org/mpg/snap/). The primary index SNPs and LD probes were searched against a collected database of expression SNP (eSNP) results, including the following tissues: fresh lymphocytes, fresh leukocytes, samples from individuals with Celiac disease, whole-blood samples, lymphoblastoid cell lines (LCL) derived from asthmatic children, HapMap LCL from three populations, a separate study on HapMap CEU LCL, additional LCL population samples, primary phytohaemagglutinin-stimulated T cells, CD4+ T cells, primary blood monocytes, CD11+ dendritic cells before and after Mycobacterium tuberculosis infection, oral, subcutaneous adipose, abdominal fat, endometrial carcinomas, ER+ and ER− breast cancer tumour cells, endometrial cancer cells, breast cancer cells, prefrontal cortex, frontal cortex, temporal cortex, pons, cerebellum, three additional large studies of brain regions including prefrontal cortex, visual cortex and cerebellum, respectively, liver, ileum, lung, skin, and primary fibroblasts. Micro-RNA QTLs were also queried for LCL and gluteal and abdominal adipose. The collected eSNP results met the criteria and were associated with expression levels as defined in the original papers. In each case where a LDL-C-associated SNP or proxy was associated with a transcript, we further examined the strongest eSNP for that transcript within that data set (best eSNP), and the LD between the best eSNP and GIST-selected eSNPs to estimate the concordance of the LDL-C and expression signals.

Statin response connectivity map analysis. The Connectivity Map (Cmap) data set is available at the Broad Institute (http://www.broadinstitute.org/cmap) and contains more than 7,000 expression profiles representing 1,309 compounds used on five different cultured human cancer cell lines (MCF7, SHMCF7, HL60, PC3 and SKMEL5). We selected (prostate tumour-derived) PC3 cells as they showed the most responsiveness to statins at a genome-wide level. Four statins were included in the original study with pravastatin, simvastatin and rosuvastatin. The IncRNA expression data set (best eSNP), and the LD between the best eSNP and GIST-selected eSNPs to estimate the concordance of the LDL-C and expression signals.

Exploration of functional impact among directly and indirectly associated variables. Genes and variants at associated loci were investigated for evidence of functional perturbation using a range of bioinformatics tools and databases. Variants showing LD (CEU $r^2 > 0.8$) with associated variants were explored for impact on coding gene function using Annovar and regulatory function using a combination of HaploReg and Regulomedb, which both draw on comprehensive data from the Encyclopedia of DNA Elements (ENCODE) and NIH Roadmap Epigenomics, respectively. Building on the functional annotation, we also identified variants that were shown to mediate eQTLs. Genes in associated loci were also used to query the NIH connectivity map for evidence of differential expression in PC3 cell lines treated with pravastatin, simvastatin and rosuvastatin. By combining a wide range of functional data and pathway support, we were able to build up a view of genes with the highest level of support in statin response.

References


Acknowledgements

Data from the Heart Protection Study have been contributed by Heart Protection Study Collaborative Group. The Prospective Study of Pravastatin in the Elderly at Risk (PROSPER) trial was supported by an investigator-initiated grant from Bristol-Myers Squibb, USA. The study was conducted, analysed and reported independently of the company. The GWAS project PHASE has received funding from the European Union’s Seventh Framework Programme (FP7/2007–2013) under grant agreement HEALTH-F2-2009-228004. A part of the genotyping was funded by The Netherlands Consortium for Healthy Ageing (NGI: 0568010). Professor Dr J.W.C. is an established clinical investigator of The Netherlands Heart Foundation (2001 D 032).

ASCOT: The Anglo-Scandinavian Cardiac Outcomes Trial (ASCOT) trial was funded by an investigator-initiated grant from Pfizer, USA. The study was investigator led and was conducted, analysed and reported independently of the company. The GWAS project PHASE has received funding from the European Union’s Seventh Framework Programme (FP7/2007–2013) under grant agreement HEALTH-F2-2009-228004. A part of the genotyping was funded by The Netherlands Consortium for Healthy Ageing (NGI: 0568010). Professor Dr J.W.C. is an established clinical investigator of The Netherlands Heart Foundation (2001 D 032).

CARDS: We thank the other investigators, the staff and the participants of the CARDS study. A full list of CARDS investigators can be found in original CARDS paper’.

CARDs was funded by grants to the Universities of London and Manchester by Pfizer, Diabetes UK and the Department of Health.

PARC: This research was supported by the National Institutes of Health: grant U19 HL069757 from the National Heart, Lung, and Blood Institute, and grant UL1TR000124 to the NIH Roadmap Epigenomics Mapping Consortium. The NIH Roadmap Epigenomics Mapping Consortium is supported by the National Human Genome Research Institute, the National Institute of General Medical Sciences, the National Institute of Mental Health, the National Institute of Drug Abuse, and the Office of the NIH Director’s Strategic Initiatives. The NNPTN: The TNT study was funded by Pfizer, who also provided support for genotyping.
Author contributions


Additional information
Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: B.M.P. serves on the Data and Safety Monitoring Board of a clinical trial funded by the device manufacturer (Zoll LifeCor). N.P.F. and A.S. received funding from Pfizer for the extended follow-up of the ASCOT UK participants. D.C. and P.M.R. received research support for independent genetic analysis in JUPITER from AstraZeneca. F.N. and B.J.B. have employment, stock and stock options in AstraZeneca, a for-profit company engaged in the discovery, development, manufacture and marketing of proprietary therapeutics such as rosuvastatin, but do not consider that this creates any conflict of interest with the subject–matter of this publication. R.M.K. serves on the Merck Global Atherosclerosis Advisory Board. The remaining authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Postmus, I. et al. Pharmacogenetic meta-analysis of genome-wide association studies of LDL cholesterol response to statins. Nat. Commun. 5:5068 doi: 10.1038/ncomms6068 (2014).

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/
Membership of Wellcome Trust Case Control Consortium


Data and Analysis Group
Chris C.A. Spencer77, Gavin Band77, Céline Bellenguez77, Colin Freeman77, Garrett Hellenthal77, Eleni Giannoulatou77, Matti Pirinen77, Richard Pearson77, Amy Strange77, Zhan Su77, Damjan Vukcevic77, Peter Donnelly77,78.

DNA, Genotyping, Data QC and Informatics Group

Publications Committee
Christopher G. Mathew92, Jenefer M. Blackwell80,81, Matthew A. Brown83, Aiden Corvin86, Mark I. McCarthy98, Chris C.A. Spencer77.

77Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford, UK; 78Department of Statistics, University of Oxford, Oxford, UK; 79Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK; 80Telethon Institute for Child Health Research, Centre for Child Health Research, University of Western Australia, 100 Roberts Road, Subiaco, Western Australia, Australia; 81Cambridge Institute for Medical Research, University of Cambridge School of Clinical Medicine, Cambridge, UK; 82Department of Psychosis Studies, NIHR Biomedical Research Centre for Mental Health at the Institute of Psychiatry, King’s College London and The South London and Maudsley NHS Foundation Trust, Denmark Hill, London, UK; 83University of Queensland Diamantina Institute, Princess Alexandra Hospital, University of Queensland, Brisbane, Queensland, Australia; 84Department of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, UK; 85Department of Epidemiology and Public Health, University College London, London, UK; 86Neuropsychiatric Genetics Research Group, Institute of Molecular Medicine, Trinity College Dublin, Dublin, Ireland; 87Molecular and Physiological Sciences, The Wellcome Trust, London, UK; 88Centre for Digestive Diseases, Queen Mary University of London, London, UK; 89Digestive Diseases Centre, Leicester Royal Infirmary, Leicester, UK; 90Department of Clinical Pharmacology, Old Road Campus, University of Oxford, Oxford, UK; 91Clinical Neurosciences, St George’s University of London, London, UK; 92King’s College London, Department of Medical and Molecular Genetics, School of Medicine, Guy’s Hospital, London, UK; 93Medical Research Institute, University of Dundee, Ninewells Hospital and Medical School, Dundee, UK; 94King’s College London Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Denmark Hill, London, UK; 95University of Cambridge, Department of Clinical Neurosciences, Addenbrooke’s Hospital, Cambridge, UK; 96NIHR Biomedical Research Centre for Ophthalmology, Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology, London, UK; 97Department of Molecular Neuroscience, Institute of Neurology, Queen Square, London, UK; 98Oxford Centre for Diabetes, Endocrinology and Metabolism (ICDEM), Churchill Hospital, Oxford, UK.