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


Our findings advance the understanding of the pharmacogenetic architecture of statin treatment response.

1 Department of Gerontology and Geriatrics, Leiden University Medical Center, Leiden 2300 RC, The Netherlands. 2 The Netherlands Consortium for Healthy Ageing, Leiden 2300 RC, The Netherlands. 3 Department of Cardiology, Leiden University Medical Center, Leiden 2300 RC, The Netherlands. 4 Medical Research Institute, Ninewells Hospital and Medical School, University of Dundee, Dundee DD1 9DT, UK. 5 Genome Centre, William Harvey Research Institute, Barts and The London School of Medicine, Queen Mary University of London, London E1 2AB, UK. 6 NIHR Barts Cardiovascular Biomedical Research Unit, Queen Mary University of London, London E1 2AB, UK. 7 Institute for Translational Genomics and Population Sciences, Los Angeles Biomedical Research Institute at Harbor-UCLA Medical Center, Torrance, California 90730, USA. 8 Department of Cardiology, William Harvey Research Institute, Barts and The London School of Medicine, Queen Mary University of London, London E1 2AB, UK. 9 Department of Cardiology, William Harvey Research Institute, Barts and The London School of Medicine, Queen Mary University of London, London E1 2AB, UK. 10 Department of Pharmacology and Therapeutics, University College Cork, Cork 30, Ireland. 11 Department of Gerontology and Geriatrics, Leiden University Medical Center, Leiden 2300 RC, The Netherlands. 12 The Netherlands Consortium for Healthy Ageing, Leiden 2300 RC, The Netherlands. 13 Department of Pharmacology and Therapeutics, University College Cork, Cork 30, Ireland. 14 Faculty of Medicine, Institute of Cardiovascular and Medical Sciences, University of Glasgow, Glasgow G12 8QQ, UK. 15 Department of Pharmacology and Therapeutics, University College Cork, Cork 30, Ireland. 16 Department of Cardiology, University College Dublin, Dublin 4, Ireland. 17 School of Medicine and Medical Sciences, University College Dublin, Dublin 4, Ireland. 18 Department of Genome Sciences, University of Washington, Seattle, Washington 98101, USA. 19 Department of Cardiology, Academic Medical Center, Amsterdam 1006 DD Amsterdam, The Netherlands. 20 Department of Vascular Medicine, Academic Medical Center, 1006 DD Amsterdam, The Netherlands. 21 University of Edinburgh, Edinburgh EH9 3JR, UK. 22 Massachusetts General Hospital, Boston, Massachusetts 02114, USA. 23 Department of Medicine, Children’s Hospital Oakland Research Institute, Oakland, California 94609, USA. 24 Interuniversity Cardiology Institute of the Netherlands, 3511 GC Utrecht, The Netherlands. 25 Members of Welcome Trust Case Control Consortium are listed at the end of the paper. 1 These authors jointly supervised the work. Correspondence and requests for materials should be addressed to I.J.L.W. (email: ijlw@jeanmillier.com) or M.C.T. (email: m.j.cauldfield@ncl.ac.uk).

ARTICLE

Received 27 Feb 2014 | Accepted 22 Aug 2014 | Published 28 Oct 2014

DOI: 10.1038/ncomms6068 | OPEN

NATURE COMMUNICATIONS | 5:5068 | DOI: 10.1038/ncomms6068 | www.nature.com/naturecommunications

© 2014 Macmillan Publishers Limited. All rights reserved.
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% and a 20–30% reduction of cardiovascular events. Despite the clinical efficacy of statins in a wide range of patients, interindividual variability exists with regard to LDL-C-lowering response as well as efficacy in reducing major cardiovascular events. 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 therapy. 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 reduction following rosuvastatin therapy. In the CARDs and ASCOT studies, single nucleotide polymorphisms (SNPs) at LPA (rs10455872) and APOE (rs445925 and rs4420638) were associated with LDL-C response to atorvastatin treatment. A combined GWAS in three statin trials identified a SNP within CLMN that is associated with the magnitude of statin-induced reduction in plasma cholesterol. However, two other GWAS identified no genetic determinants of LDL-C response to statin therapy at a genome-wide significant level.

On the basis of these studies, as well as previous candidate gene studies, the only 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/CERSLR2/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 × 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, β = −0.043, s.e. = 0.005, P = 1.58 × 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, β = 0.041, s.e. = 0.006, P = 1.95 × 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 described. A third genome-wide significant association was found with a SNP at RICTOR on chromosome 5 (rs13166647, MAF = 0.230, β = −0.253, s.e. = 0.046, P = 4.50 × 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 × 10^−3 from 158 loci for further investigation in three additional studies comprising up to 22,318 statin-treated subjects (see Methods; Supplementary Tables 1 and 2; 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 × 10^−7 and 2.26 × 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/CERSLR2/PSRC1 (rs646776, β = −0.013, s.e. = 0.002, P = 1.05 × 10^−9 and rs12740374, β = −0.013, s.e. = 0.002, P = 1.05 × 10^−9; Fig 2c) and SLCO1B1 (rs2900478, β = 0.016, s.e. = 0.003, P = 1.22 × 10^−9; Fig 2d), indicating an additional 1.5% increase per allele in LDL-C lowering effect for carriers of the SORT1/CERSLR2/PSRC1 SNP and a 1.6% smaller LDL-C lowering per minor allele for carriers of the SLCO1B1 SNP.

The six next-ranked SNPs with P values just below 5 × 10^−8 in the combined meta-analysis, including the two SNPs at RICTOR (rs13166647 and rs13172966), were selected for additional genotyping in the Scandinavian ASCOT participants (see Methods). None of these six SNPs reached genome-wide significance after this additional genotyping (Supplementary Table 6). Therefore, our overall genome-wide significant findings were the SNPs at APOE, LPA, SORT1/CERSLR2/PSRC1 and SLCO1B1.

Subfraction analyses. To extend our results for the novel GWAS finding SORT1/CERSLR2/PSRC1, we performed additional association analyses, using measurements of cholesterol levels in four LDL subfractions (large, medium, small and very small) from two of the trials in GIST, CAP and PRINCE (Table 2; see Methods). The minor allele of SORT1 rs646776 was associated with greater statin-induced reductions in levels of all LDL subfractions, and there was a nonsignificant trend for larger effect sizes and greater statistical significance for lowering of small and very small LDL (Table 2). In contrast, the APOE SNP associated with greater...
whether there was an interaction between LDL-C change and the JUPITER trial, additional analyses were performed to determine if there is little effect of measurement noise. Next, within the apparent effect only slightly for all the markers, suggesting that correcting our effect size estimate further in the CARDS data set. Correcting our effect size estimate further 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) 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.

**Effects of off-treatment LDL-C.** To demonstrate that our findings for LDL-C response to statins 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 significant $P$ values for interaction (all $<5 \times 10^{-2}$) for SNPs at the four 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 associated 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 combined GWAS meta-analysis (Supplementary Table 5).

### 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>109620053</td>
<td>rs646776</td>
<td>SOR/T CELSR2/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 × 10^{-7}</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>160930108</td>
<td>rs10455872</td>
<td>LPA</td>
<td>G</td>
<td>A</td>
<td>Stage 1</td>
<td>12,981</td>
<td>0.069</td>
<td>0.041</td>
<td>0.006</td>
<td>4.1</td>
<td>1.95 × 10^{-11}</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>21260064</td>
<td>rs2900478</td>
<td>SLCO1B1</td>
<td>A</td>
<td>T</td>
<td>Stage 1</td>
<td>16,749</td>
<td>0.165</td>
<td>0.016</td>
<td>0.003</td>
<td>−1.6</td>
<td>2.26 × 10^{-6}</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>50107480</td>
<td>rs445925</td>
<td>APOE</td>
<td>A</td>
<td>G</td>
<td>Stage 1</td>
<td>13,909</td>
<td>0.098</td>
<td>0.043</td>
<td>0.005</td>
<td>4.3</td>
<td>1.58 × 10^{-18}</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></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 ½ extra LDL-C lowering in carriers versus non-carriers of the SNP.

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 reduction 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 evidence discussed below that this locus affects levels of lipoprotein(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).

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 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 responsive 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 (see Methods). Five genes (APOE, BRCA1, GRPEL1, 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). Briefly, 87 literature-reported genes linked to statin response were combined with the 185 candidate gene loci reported here.
with the lead SNP (shown in purple). The RefSeq genes in the region are shown in the lower panel. The plots show the genome-wide significant associated loci in the combined meta-analysis (4). A conservative network of direct interactions was constructed between query genes (Supplementary Fig. 4). The colour of the SNPs is based on the LD with the lead SNP (shown in purple). The RefSeq genes in the region are shown in the lower panel. P values were generated using linear regression analysis.

Table 2 | Associations of the minor alleles of rs646776, rs445925, rs2900478 and rs10455872 with changes in LDL-C and LDL subfractions in response to statin in the combined CAP and PRINCE studies.

<table>
<thead>
<tr>
<th>Change*</th>
<th>SORT1/CELSR2/PSRC1</th>
<th>APOE</th>
<th>SLC01B1</th>
<th>LPA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>rs646776 (MAF 0.2)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LDL-C total</td>
<td>Beta</td>
<td>s.e.</td>
<td>P value</td>
<td>Beta</td>
</tr>
<tr>
<td>Large LDL-C</td>
<td>-0.023</td>
<td>0.008</td>
<td>0.003</td>
<td>-0.046</td>
</tr>
<tr>
<td>Medium LDL-C</td>
<td>-0.028</td>
<td>0.014</td>
<td>0.042</td>
<td>-0.075</td>
</tr>
<tr>
<td>Small LDL-C</td>
<td>-0.027</td>
<td>0.015</td>
<td>0.075</td>
<td>-0.079</td>
</tr>
<tr>
<td>Very small LDL-C</td>
<td>-0.047</td>
<td>0.018</td>
<td>0.009</td>
<td>-0.071</td>
</tr>
<tr>
<td>LDL-C, low-density lipoprotein cholesterol; MAF, minor allele frequency.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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/CELSR2/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 ε2 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...
Notably, the minor allele of rs646776 is at the cell surface, with subsequent internalization and lysosomal secretion of apoB-containing precursors, and, perhaps of greater lowering of plasma LDL-C results from two mechanisms: reduced Friedewald formula, include cholesterol that resides in Lp(a)6,8. A stronger inhibition of cholesterol synthesis by statin treatment8,10. Increased hepatic cholesterol synthesis, it may also predispose to heightened statin response as a result of 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 e2 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 SLCO1B1 rs2900478 minor allele was associated with a smaller LDL-C reduction in response to statin treatment. SLCO1B1 encodes the organic anion-transporting polypeptide OATP1B1 and facilitates the hepatic uptake of statins25. SNP rs2900478 is in strong LD (r2 = 0.89) with rs4149056, which represents the Val174 Ala 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 allele25. 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)23. 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 effect23. In a GWAS of the genetic risk factors for simvastatin-induced myopathy, SLCO1B1 showed the strongest association25. Haploinsufficient carriers of the 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 SLCO1B1 gene on statin efficacy was abolished after removal of individuals who showed signs of intolerance26.

GWCA identified three independent loci in the APOE gene region and two loci in the LPA gene region (Supplementary Table 9). GWCA also showed several other loci with P < 5 × 10−8 that were not GWAS significant on single-SNP analysis (HGD, RNF175, ISCA1-HTRIA, GLIS3-SLC1A1, LOC100128657, NKKX2-3-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 SLCO1B1 variant on statin pharmacogenetics is known to be highly dependent on statin type and dose24,27. 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 additive analysis with inverse variance weighting. Since we observed that the SLCO1B1 variant on statin pharmacogenetics is known to be highly dependent on statin type and dose24,27. To overcome this limitation, 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 equivalent based on the daily dosages required to achieve a mean 30% LDL-C reduction. Using this table, we made the different statin dosages and types comparable within the studies. To correct for between-study variance, we used a fixed effect meta-analysis with inverse variance weighting. Since we observed that the 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.

The GWAS study performed in CARDS and ASCOT8 and of the SNP rs7412 in JUPITER2. Since the apoE e2 protein results in increased hepatic cholesterol synthesis, it may also predispose to stronger inhibition of cholesterol synthesis by statin treatment8,10.
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 SORT1/CERS2/PSRC1, APOE and LPA\(^2\). 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 small 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 APOE and LPA loci, two new loci, SORT1/CERS2/PSRC1 and SLC01B1, also have a modest but genome-wide significant effect on LDL-C response. The minor alleles of the APOE rs445925 and SORT1/CERS2/PSRC1 rs646776 SNPs were associated with a larger statin response, whereas the minor alleles of the LPA rs10455872 and SLC01B1 rs2990478 SNPs were associated with a smaller statin response. Our findings advance the understanding of the pharmacogenetic architecture of statin response.

**Methods**

**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 TNT) and 10,175 statin-treated subjects from 10 observational studies (AGES, ARIC, BioVU, CHS, FHS, GoDARTS I, GoDARTS II, Health ABC, HVH and MESA). Further analysis (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.

**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.

**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 on-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 screening visit before starting statin therapy to the randomization visit\(^2\).

**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.

**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 when testing for association. To correct for multiple 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.

**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 by each study to identify between-study differences (Supplementary 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.

**Determination of changes in LDL subfractions.** LDL subclasses were classified as described previously\(^2\) 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 0.1 ml of whole plasma were mixed 1:1 with a sampling buffer of 20% sucrose and 0.25% bromophenol blue. Electrophoresis of samples and size calibration standards was performed using 2–14% polyacrylamide gradients 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 a 0.81% acetic acid, 4% methanol solution until they were scanned by computer-assisted densitometry for determination of areas of LDL IVb (22.0–23.2 nm), LDL IIIa (23.3–24.1 nm), LDL IIIb (24.2–24.6 nm), LDL IIb (24.7–25.5 nm) and LDL IIa (25.6–26.4 nm). The cholesterol concentrations of the subfractions (mg d/l \(^{-1}\) plasma) were determined by multiplying percent of the total stained LDL area for each subfraction by the LDL-C for that sample. For genetic association analyses, subfractions were grouped into large LDL (LDL I + IIa), intermediate LDL (LDL IIb), small LDL (LDL IIIa and IIIb) and very small LDL (LDL IVa + IVb) as described previously\(^2\). A generalized estimating equation method was used to test the association of log change with the interaction of the four SNPs by LDL subclassation.

**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 CARDOS, where 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
β is the coefficient of regression for on-treatment trait value on genotype adjusted for covariates of interest. Df is the degrees of freedom of the regression model.

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. Such a strategy will 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-combined 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 value < 5 × 10⁻⁸. In the next step, it calculates the P values of all 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 SNPs 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 r² > 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 network was constructed 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 network constructed based on direct drug-target 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 (r² = 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 proxies were selected against a collected database of expression information eQTLs, including the following tissues: fresh lymphocytes, fresh leukocytes, samples in individuals with Celiac disease, whole-blood samples, lymphoblastoid cell lines (LCL) derived from asthmatic children, LCL HapMap LCL from three populations, a separate study on HapMap CEU LCL, additional LCL population samples, primary phytohaemagglutinin-stimulated T cells, CD4 + T cells, peripheral blood monocytes, CD11 + dendritic cells before and after Mycobacterium tuberculosis infection, men, and subcutaneous adipose, smooth muscle, endomysial and subendomysial, skeletal muscle, endomyocardial, endocardial, myocardioma, ER + and ER − breast cancer tumour cells, brain cortex, ER + breast cancer tumour cells, brain cortex, fetal frontal cortex, peripheral frontal cortex, temporal frontal cortex, prefrontal cortex, pons, cerebellum, three additional large studies of brain regions including prefrontal cortex, visual cortex and cerebellum, respectively, liver, ileum, osteoblast, ileum, lung, skin, primary fibroblasts. Micro-RNA QTLs were also queried for LCL. We and glued and abdominal adipose. The collected eQTL results with gene expression levels as defined in the original papers. In each case where a LCL-associated SNP or proxy was associated with a transcript, we further examined the strongest eSNP for that transcript within the data set (best eSNP), and the LD between the best eSNP and GIST-selected eSNPs to estimate the concordance of the LCL-C and expression signals.

Statin response connectivity map analysis. The Connectivity Map (Cmap) data set is available at the Broad Institute (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, MCF7/CE, 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 our analysis, relating pravastatin, atorvastatin, simvastatin and rosvastatin. PC3 Instance reference files for each statin treatment were extracted (as defined by Lamb et al.), that is, a treatment associated to its control pair. Transcripts were considered to show evidence of differential expression with a fold change > 2. A fold change > 1.5 was considered to be suggestive of differential expression only.

Exploration of functional impact among directly and indirectly associated variants. Genes and variants across all LDL-C-associated loci were investigated for evidence of functional perturbation using a range of bioinformatics tools and databases. Variants showing LD (CEU r² > 0.8) with associated variants were explored for impact on coding gene function using Annovar and regulatory function using a combination of HaploReg® and RegulomedDB, which both draw on comprehensive data from the Encyclopedia of DNA Elements (ENCODE) and the NIH Roadmap Epigenomics consortium. 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 rosvastatin. 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


34. Fehrmann, R. S. et al. Trans-eQTLs reveal that independent genetic variants associated with a complex phenotype converge on intermediate genes, with a major role for the HLA. PLoS Genet. 7, e1001977 (2011).


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.S. and A.S. received funding from Pfizer for the extended follow-up of the ASCOT UK participants. D.I.C. and P.M.R. received research support for independent genetic analysis in JUPITER from AstraZeneca. F.N. and B.J. 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/

Membership of Wellcome Trust Case Control Consortium

Management Committee: Peter Donnelly77,78, Ines Barroso79, Jenefer M. Blackwell80,81, Elvira Bramon82, Matthew A. Brown83, Juan P. Casas84,85, Aiden Corvin86, Panos Deloukas79, Audrey Duncanson87, Janusz Jankowsk78,89,90, Hugh S. Markus91, Christopher G. Mathew92, Colin N.A. Palmer93, Robert Plomin94, Anna Rautanen77, Stephen J. Sawcer95, Richard C. Trembath92, Ananth C. Viswanathan96, Nicholas W. Wood97.

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.