A Multivariate Genome-Wide Association Analysis of 10 LDL Subfractions, and Their Response to Statin Treatment, in 1868 Caucasians

Heejung Shim[1], Daniel I. Chasman[2], Joshua D. Smith[3], Samia Mora[2], Paul M. Ridker[2], Deborah A. Nickerson[3], Ronald M. Krauss[4]*, Matthew Stephens[1,5]*

1 Department of Human Genetics, University of Chicago, Chicago, IL, USA, 2 Center for Cardiovascular Disease Prevention, Brigham and Womens Hospital and Harvard Medical School, Boston, MA, USA, 3 Department of Genome Sciences, University of Washington, Seattle, WA, USA, 4 Childrens Hospital Oakland Research Institute, Oakland, CA, USA, 5 Department of Statistics, University of Chicago, Chicago, IL, USA

* RKrauss@chori.org (RMK); stephens@galton.uchicago.edu (MS)

Abstract

We conducted a genome-wide association analysis of 7 subfractions of low density lipoproteins (LDLs) and 3 subfractions of intermediate density lipoproteins (IDLs) measured by gradient gel electrophoresis, and their response to statin treatment, in 1868 individuals of European ancestry from the Pharmacogenomics and Risk of Cardiovascular Disease study. Our analyses identified four previously-implicated loci (SORT1, APOE, LPA, and CETP) as containing variants that are very strongly associated with lipoprotein subfractions ($\log_{10}$Bayes Factor > 15). Subsequent conditional analyses suggest that three of these (APOE, LPA and CETP) likely harbor multiple independently associated SNPs. Further, while different variants typically showed different characteristic patterns of association with combinations of subfractions, the two SNPs in CETP show strikingly similar patterns - both in our original data and in a replication cohort - consistent with a common underlying molecular mechanism. Notably, the CETP variants are very strongly associated with LDL subfractions, despite showing no association with total LDLs in our study, illustrating the potential value of the more detailed phenotypic measurements. In contrast with these strong subfraction associations, genetic association analysis of subfraction response to statins showed much weaker signals (none exceeding $\log_{10}$Bayes Factor of 6). However, two SNPs (in APOE and LPA) previously-reported to be associated with LDL statin response do show some modest evidence for association in our data, and the subfraction response proles at the LPA SNP are consistent with the LPA association, with response likely being due primarily to resistance of Lp(a) particles to statin therapy. An additional important feature of our analysis is that, unlike most previous analyses of multiple related phenotypes, we analyzed the subfractions jointly, rather than one at a time. Comparisons of our multivariate analyses with standard univariate analyses demonstrate that multivariate analyses can

substantially increase power to detect associations. Software implementing our multivariate analysis methods is available at http://stephenslab.uchicago.edu/software.html.

Introduction

Levels of plasma lipids and lipoproteins are related to risk of cardiovascular disease, and because of this, considerable attention has been devoted to genetic association analyses of lipid-related measures. The largest of these studies are genetic association analyses for plasma concentrations of the common clinical lipid phentoypes: total cholesterol (TC), LDL-cholesterol (LDL-C), HDL-cholesterol (HDL-C), and triglycerides (TG). For example, [1] performed a meta-analysis of these traits in > 100,000 individuals of European ancestry, and identified a total of 95 associated loci. Other (smaller, although still substantial) studies considered genetic associations with more detailed lipid-related measurements, specifically plasma concentrations of subfractions of very low density lipoproteins (VLDLs), intermediate density lipoproteins (IDLs), LDLs, and HDLs, measured by NMR [2–4]. And, motivated by the fact that statin drugs are widely used to treat lipid phenotypes, and that response to these drugs has a genetic component, several studies have searched for genetic associations with response of LDL-C, HDL-C, TC and TG to statin treatment [5–7].

In this paper we describe genetic association analyses of 7 subfractions of LDLs and 3 subfractions of IDLs measured by gradient gel electrophoresis, and of their response to statin therapy. Although our sample size is smaller than many recent lipid association studies (1868 individuals), our phenotypic measurements provide higher resolution information for IDLs and LDLs than prior genetic association studies of either statin-treated or untreated samples. Specifically, our 10 subfractions of IDLs and LDLs compare with 3–4 size subfractions in the NMR-based studies, and no previous genome-wide association study of lipoprotein response to statin therapy has considered subfraction data. While our smaller sample size limits what we can say about genetic variants with small effects, for variants with sufficiently strong associations our data provide a more detailed picture of their associations with the entire IDL/LDL subfraction profile than any previous study. We find several examples of SNPs that are only very weakly associated with total LDL-C in our study, but are much more strongly associated with one or more individual subfractions. Of particular note, we highlight two independently-associated variants in the CETP gene that have no overall effect on total LDL-C in our study, but a strong effect on several individual IDL/LDL subfractions (in addition to a well-established strong effect on total HDL-C).

In addition to the detailed nature of the phenotypes, our study also differs from most previous studies in our use of *multivariate* association analysis of related phenotypes, rather than treating each phenotype separately. Our results illustrate that association analysis of multivariate phenotypes can substantially increase the strength of association signals compared with conventional univariate analyses.

Methods

Study populations and genotype data

All samples in our analysis were derived from the Pharmacogenomics and Risk of Cardiovascular Disease (PARC) study. The study population, experimental design, and genotyping procedures have been described in detail previously [6]. Briefly, this study contains individuals from two statin trials: the Cholesterol and Pharmacogenetics (CAP) study [8], and the

Pravastatin Inflammation/CRP Evaluation (PRINCE) study [9]. The PRINCE study consists of two cohorts, one containing individuals with history of CVD (secondary prevention cohort) and the other containing individuals with no history of CVD (primary prevention cohort). Participant characteristics are summarized in Table 1.

Genotyping was conducted in two stages. The first stage individuals were genotyped on the Illumina HumanHap300 bead chip and the second stage individuals were genotyped on the Illumina HumanQuad610 bead chip and a custom-made iSelect chip. The HumanHap300 and the HumanQuad610 chips (henceforth referred to as the 300K chip and the 610K chip) were designed to tag common variation among individuals of European ancestry while 12,959 SNPs in the iSelect chip were selected to increase coverage of candidate SNPs for cardiovascular disease regardless of minor allele frequency (MAF). Our analyses reported here utilized a total of 1,868 Caucasian individuals for whom complete LDL subfraction phenotype data were available (see below).

To maximize genomic coverage and combine the multiple groups genotyped on different SNP chips, we performed genotype imputation [10] [11], using an imputation protocol that has been previously described [12]. Briefly, genotype imputation was performed using IM-PUTE2 [10] with an integrated reference panel that included 120 CEU haplotypes from the 1000 Genomes Pilot Project ("1000G") [13] and 1910 worldwide haplotypes from the HapMap Phase 3 Project ("HM3") [14]. This procedure generated genotypes (either genotyped or imputed) for 7,836,525 SNPs.

Phenotypic measurements and normalisation

LDL cholesterol was estimated by the formula of [15] using measured total cholesterol, HDL cholesterol, and triglyceride. All lipid measurements were consistently in range as determined by the NHLBI-CDC Lipid Standardization Program. Gradient gel electrophoresis estimates of LDL subclass concentrations and LDL peak diameter were determined from whole plasma by 2%–14% non-denaturing polyacrylamide gradient gel electrophoresis as described previously [16] except that Sudan Black was used as the lipid stain [17], and the size range was extended to include IDL [18] since the Friedewald formula provides a measure of cholesterol in the density range that includes IDL as well as LDL particles. The subfractions analyzed and their corresponding particle size intervals were: LDL4b (22.0–23.2 nm), LDL4a (23.3–24.1 nm), LDL3b (24.2–24.6 nm), LDL3a (24.7–25.5 nm), LDL2b (25.6–26.4 nm), LDL2a (26.5–27.1 nm), LDL1 (27.2–28.5 nm), IDL3 (28.6–29.0 nm), IDL2 (29.1–29.6 nm) and IDL1 (29.7–30.3nm). The cholesterol concentrations of these subfractions were determined by multiplying percent of the total stained LDL and IDL for each subfraction by the LDL cholesterol value [17]. In both studies, the LDL subfractions and LDL cholesterol values were estimated from blood samples taken on one visit before statin treatment and one visit during the treatment (after 6 weeks in CAP and 12 weeks in PRINCE).

Following [6] we separately normalized the 12 LDL-related phenotypes (total LDL-C, LDL peak diameter, and the 10 LDL subfractions), by the following normalization procedure, which aims to reduce the influence of outlying observations and avoid false positive association due to systematic differences between studies/cohorts. We partitioned the individuals into six strata, corresponding to the two stages of genotyping in each of three groups (CAP, PRINCE primary prevention and PRINCE secondary prevention). Within each stratum we separately quantile transformed the pre-treatment and post-treatment measurements to a standard normal distribution to yield pre-treatment and post-treatment phenotypes, P and T. Then we calculated both the average $A = (T + P)/2$, and the difference $D = (T − P)$ of the transformed measurements for each individual. (The motivation is that analyses of A are well-powered to

Table 1. Summary of phenotype distributions (mean ± standard deviation) for individuals in our analysis, stratified by study/cohort.

	CAP	PRINCE (secondary prevention)	PRINCE (primary prevention)
N	581	797	490
Gender, N males	310 (53.4%)	631 (79.2%)	366 (74.7%)
Age	54.5±12.6	69.7±10.7	56.8±12.3
BMI	27.7±5.5	28.9±5.3	29.1±5.3
Smoking (# of subjects)	78 (13.4%)	110 (13.8%)	62 (12.7%)
Diabetic (# of subjects)	0 (0%)	225 (28.2%)	37 (7.6%)
type of statins	simvastatin	pravastatin	pravastatin
dose of statins	40mg/day for 6 weeks	40mg/day for 12 weeks	40mg/day for 12 weeks
LDL-C (L)			
Untreated	132±33.2	125±29.7	142.4±24.9
Treated	76.7±23.7	91.8±26.3	107.5±24.3
LDL peak particle diameter (Ld)			
Untreated	266.9±8.9	262.4±9.1	262.9±8.3
Treated	267.6±8.6	262.4±8.7	262.8±8.1
IDL1 (i1)			
Untreated	6.3±2.6	5±2.1	5.8±2.1
Treated	3.5±1.9	3.4±1.6	4.1±1.8
IDL2 (i2)			
Untreated	6.8±2.5	5.5±2.1	6.4±2.1
Treated	3.7±1.6	3.8±1.6	4.6±1.7
IDL3 (i3)			
Untreated	8.4±3.4	6.7±2.7	7.8±2.9
Treated	4.7±2.1	4.7±2	5.5±2.2
LDL1 (l1)			
Untreated	34.1±17.3	27.2±15.5	31.4±16.4
Treated	17.9±9.1	18.1±9.6	21.8±11.2
LDL2a (l2a)			
Untreated	25.3±12.9	23.5±12.3	28.3±13
Treated	13.6±6.6	16.5±8.6	20.5±9.5
LDL2b (l2b)			
Untreated	22.9±12.3	25.8±12.8	30.7±14.5
Treated	13.8±6.7	19.5±9.6	23.4±10.4
LDL3a (l3a)			
Untreated	14.5±12.1	17.6±12.6	18.5±13.6
Treated	9.1±6.2	13.7±8.9	15±10.1
LDL3b (l3b)			
Untreated	3.9±3.2	4.5±3.3	4.4±3.4
Treated	2.8±1.3	3.8±2.4	3.8±2.5
LDL4a (l4a)			
Untreated	4.7±1.9	4.6±2.5	4.5±2.3
Treated	3.6±1.2	4.1±1.8	4.2±1.9
LDL4b (l4b)			
Untreated	5.1±2.1	4.6±2.5	4.7±2.2
Treated	3.9±1.5	4.2±1.9	4.6±2.2

Abbreviations for each phenotype are in parentheses.

doi:10.1371/journal.pone.0120758.t001

identify associations that occur in both T and P, whereas analyses of D are aimed at identifying associations with statin response.) We then separately corrected A and D for covariates (age, log(BMI), sex, and smoking status) within each stratum using a standard multiple linear regression, to yield covariate-corrected phenotypes A' and D' (obtained from the residuals of the multiple regression). Finally, each of A' and D' were again quantile transformed to a standard normal distribution within each strata, to yield final phenotypes $\tilde{A}$ and $\tilde{D}$. Since previous analyses of the PARC study suggest little potential for spurious associations due to cryptic population stratification within strata [6], we did not further control for population substructure within strata.

Replication cohort, genotype data, phenotypic measurements and normalization

Replication of our CETP associations was performed among individuals of European descent from the JUPITER study [7, 19]. Specifically, we examined genotype data at our two replication SNPs, and phenotype data among up to 6745 JUPITER participants on VLDL, IDL and LDL subfractions measurements determined by an ion mobility assay [20, 21]. We focused on the 12 ion-mobility phenotypes that correspond most closely to the 12 phenotypes in our study, namely: total LDL (L), LDL peak diameter (Ld), VLDL small (i1), IDL1 (i2), IDL2 (i3), LDL1 (l1), LDL2a (l2a), LDL medium small (l2b), LDL small (l3a), LDL3b (l3b), LDL4a (l4a), LDL4b (l4b).

Before association analysis each phenotype was quantile transformed to a standard normal distribution, and then adjusted for covariates (age, sex, and region: North America vs Europe) using a standard multiple linear regression. The covariate-corrected phenotypes (i.e. the residuals of the multiple regression) were used in replication analyses. For the top SNP we regressed these covariate-corrected phenotypes against SNP genotype to estimate an effect for each phenotype. We then took the residuals from this regression (to control for the effect of the top SNP) and regressed them against SNP genotype at the secondary SNP to estimate remaining effects at the secondary SNP.

Multivariate association analyses

We performed multivariate association analyses using methods described in detail in [22]. To outline these methods, consider analyzing d phenotypes $Y_1, \ldots, Y_d$, and assessing their association with a single SNP, whose genotypes are denoted by g. Instead of performing a single test of association, the idea is to compare different models for the way that $Y_1, \ldots, Y_d$ could be associated with g. The null model is that none of them are associated with g. Then there are a large number of alternative models, in which different subsets of the variables are associated with g. For example, one alternative is that Y_1 alone is associated with g, and the others are unassociated. Another is that Y_2 alone is associated with g. Another is that all of them are associated with g, etc. In addition, the method allows that some phenotypes might be "indirectly" associated with g, which means that their association with g is mediated entirely through the directly associated phenotypes. (Specifically, these phenotype are associated with g, but are conditionally independent of g given the directly associated phenotypes.)

Formally, consider partitioning the phenotypes into three groups, U, D and I where U denotes the phenotypes that are unassociated with g, D denotes the phenotypes that are "directly" associated with g, and I denotes the phenotypes that are "indirectly" associated with g. The support in the data for each possible partition $\gamma = (U, D, I)$, compared with the null ($\gamma = \gamma_0$, where γ_0 is the partition with all phenotypes in U), is given by the Bayes Factor $\mathrm{BF}_\gamma = \mathrm{P}(Y|g, \gamma)/\mathrm{P}(Y|\gamma_0)$. Here we use expressions for these Bayes Factors from a multivariate normal model for Y [22]. Since in

practice any of the alternative hypotheses may hold, the overall evidence against the null hypothesis is given by a weighted average of these Bayes Factors

$$\text{BF}_{\text{av}} := \sum_{\gamma \neq \gamma_0} w_\gamma \text{BF}_\gamma \tag{1}$$

where the weights w_γ can be chosen to reflect the relative prior plausibility of different values of γ. Here we use the default weights suggested in [22], which come from putting uniform prior probability on the number of associated phenotypes $A = |D \cup I| = 1, \ldots, d$ and, conditional on A, uniform prior probability on $|D| = 1, \ldots, A$.

In cases where there is strong evidence against the null hypothesis, these methods can be used to compare the evidence for different alternative hypotheses. In particular, we summarize the evidence for Y_j being associated with g using the posterior probability of association $P(Y_j \in D \text{ or } I)$. (In figures presented here we give these posterior probabilities conditional on $\gamma \neq \gamma_0$.)

In comparisons below, we compare the multivariate Bayes Factor BF_{av} with the Bayes Factor from the d univariate tests $\text{BF}_{\text{uni}} = (1/d) \sum_{j=1}^d \text{BF}_j$ where BF_j denotes the Bayes Factor from a univariate association test of Y_j. Note that BF_j is highly correlated with the usual p value testing for association between Y_j and g [23], and BF_{uni} is highly correlated with the minimum of these p values across the d phenotypes [22]. For more details on the connections between these Bayes Factors and standard tests see [22].

Calculation of the Bayes Factors requires specification of a single hyperparameters, σ_a, which controls how large the expected genetic effects are for truly associated SNPs. In all the Bayes Factor calculations used here we averaged over $\sigma_a = 0.05, 0.1, 0.2, 0.4$. (See [24] for discussion.)

Initial filtering to reduce computation of genome-wide analysis

Because the total number of possible values for γ in (Eq 1) is large, computing this value genome-wide is computationally intensive (e.g., 30 min/SNP for 12 phenotypes). To reduce computation we performed an initial filtering step in which, for all SNPs (7,836,525), we compute BF_{all}, which is the BF for γ where all variables are in D, and BF_{uni}. (Note that BF_{all} corresponds to a standard multivariate test of association, such as MANOVA; see [22].) We then performed the full analysis only on SNPs (24,601 for $\tilde{A}$ and 25,175 for $\tilde{D}$) that appear promising based on this filter ($\log_{10} \text{BF}_{\text{uni}} > 1.3$ or $\log_{10} \text{BF}_{\text{all}} > 1.3$). S2 Text and S3 Text in Supporting Information show detailed results from the full analysis.

Identifying deviations from multivariate normality

To identify individuals with phenotypes that appear to deviate from multivariate normality we performed multivariate outlier detection based the Mahalanobis distance as follows. Given d-dimensional phenotype vectors $y_1, \ldots, y_n$, the Mahalanobis distance of y_i from the mean $\hat{\mu} := (1/n) \sum_{i=1}^n y_i$ is computed using

$$m_i^2 := (y_i - \hat{\mu})^t \hat{S}^{-1} (y_i - \hat{\mu}), \tag{2}$$

where $\hat{S} = (1/(n-1)) \sum_{i=1}^n (y_i - \hat{\mu})(y_i - \hat{\mu})^t$ is the sample covariance matrix. For each individual we obtain a p value testing whether the individual's phenotype is consistent with the multivariate normal model by using the fact that, under a multivariate normal distribution, $m_i^2 \sim \chi_d^2$.

Results

The effect of statin treatment on LDL subfractions

Fig 1 shows the mean and standard deviation of the absolute and fractional change in LDL subfractions, by study/cohort. The three cohorts show similar trends, with stronger effects in CAP than PRINCE, presumably reflecting both the different statins and doses used, and the different enrollment criteria. In particular, all three studies show a consistent trend of larger fractional change in the larger (lower-density) particles, and lower fractional change in the smaller (higher-density) particles. This is consistent with studies showing that very small LDL have low affinity for the LDL receptor [25].

Genome-wide association analyses, irrespective of statin exposure

To identify genetic variants that are associated with LDL-C subfractions, irrespective of statin exposure, we conducted a multivariate genome-wide scan of an average of the pre-treatment measurements (P) and post-treatment measurements (T) for all twelve LDL-C related phenotypes simultaneously (10 subfractions plus Ld, and total LDL-C; S1 Fig in Supporting Information shows correlations among the phenotypes). The normalization and data processing steps used to derive this phenotype ($\tilde{A}$) are described in the methods section. The rationale for
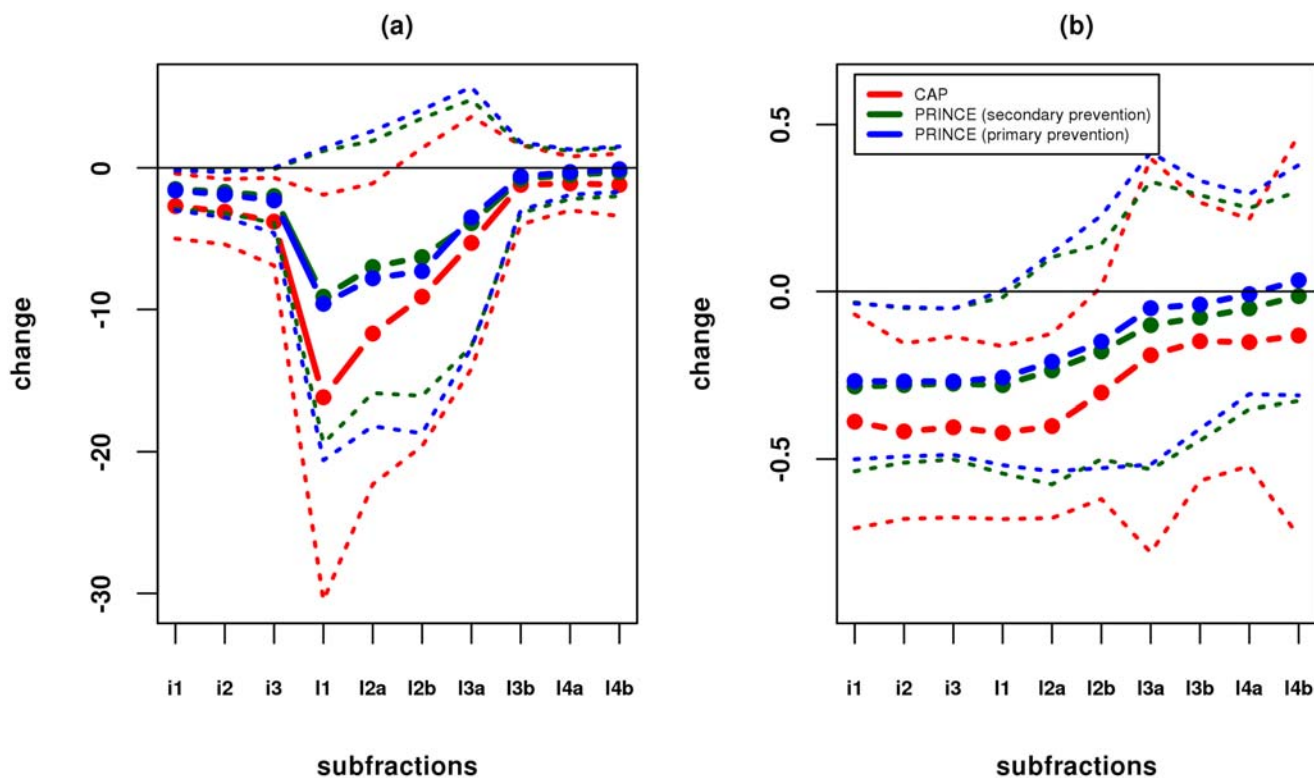

Fig 1. Effects of statin on cholesterol concentrations in LDL subfractions. Dashed lines show (a) mean absolute change, and (b) mean fractional change, in each study/cohort in each subfraction (see Table 1 for abbrevations); dotted lines show ± one standard deviation; subfractions are in order from lower (left) to higher (right) density lipoprotein.

Table 2. Summary of the strongest associations with LDL subfractions in our analysis.

Gene	chr	type	SNP	$\log_{10}$ BF$_{av}$ (LDL subfractions)	$\log_{10}$ BF (LDL-C)	Associated[1] lipids	r^2 with[2] top SNP	Imputation[3] quality	MAF[4]
SORT1	1	top	rs7528419	30.3	3.1	-	-	0.99	0.20
		lead	rs629301	29.9	3.1	LDL-C (TC)	1	0.99	0.20
APOE	19	top	rs7412	22.8	7.2	-	-	0.81	0.07
		lead[5]	rs439401	-0.6	0.6	TG	0.02	+	0.38
LPA	6	top	rs55730499	16.8	-0.3	-	-	0.62	0.07
		lead	rs1564348	-0.6	-0.3	LDL-C(TC)	0.001	+	0.16
		lead	rs1084651	-0.6	-0.4	HDL-C	0.01	0.82	0.13
CETP	16	top	rs247616	16.2	-0.1	-	-	+	0.33
		lead	rs3764261	15.6	-0.1	HDL-C (TC,LDL-C, TG)	0.96	+	0.32

Rows corresponding to the "top" SNPs summarize the results of our analysis for the SNP showing strongest association in each gene (see S1 Table in Supporting Information for other detailed results). Rows corresponding to the "lead" SNP summarize results for the most significant SNP reported in the large meta-analysis of four lipid traits (LDL-C, HDL-C, total cholesterol, TC, and Triglycerides, TG) by [1].

[1] The associated lipid traits from [1]. (the strongest associated trait is outside parentheses)

[2] r^2 between the top and lead SNPs.

[3] Imputation quality reported from the software IMPUTE2 [10]; "+" for the typed SNP.

[4] Minor allele frequency.

[5] [1] reported another lead SNP for APOE associated with LDL-C (the most strongly associated trait), TC, and HDL-C, but that SNP was not included in our reference panel for imputation.

averaging pre-treatment and post-treatment measures is that it reduces the influence of environmental and temporal fluctuations in phenotype, and of measurement errors (see also [6]).

Our analysis summarizes the evidence for each SNP being associated with the multivariate phenotype data (against the null of no association) by a Bayes Factor, BF$_{av}$, which compares the likelihood of the data under a range of alternative models, with the likelihood under the null model (see methods). Although there is no direct correspondence between Bayes Factors and p values, typically $\log_{10}$ BF$_{av} \approx 5 - 6$ corresponds roughly to conventional levels of genome-wide significance.

We found SNPs in or near four genes (SORT1, CETP, APOE, and LPA; S2 Fig in Supporting Information) that showed very strong association signals ($\log_{10}$ BF$_{av} > 15$), and no other SNPs had $\log_{10}$ BF$_{av} > 5$. The association results, including details of the most strongly associated SNPs in each gene, are summarized in Table 2. Notably, all these signals were *much* stronger (15+ orders of magnitude) than in a univariate analysis of LDL-C alone, indicating the potential gains from using both more detailed phenotypic measurements (vs LDL-C) and multivariate analysis methods (vs univariate analysis), an issue we return to below.

The multivariate association analysis methods we use are based on an assumption of multivariate normality (formally, of the residuals, although when effect sizes are small this is similar to multivariate normality of the phenotypes). Although we transformed each phenotype to be univariate normal, this does not guarantee that the phenotypes will be, jointly, multivariate normal. Therefore, to check for sensitivity to deviations from multivariate normality, we repeated the analysis after removing 89 individuals who were identified as potential outliers from multivariate normality on the basis of their Mahalanobis distance from the mean ($p < 0.01$ in a test for multivariate normality; see Methods). All four of the strongest associations remained

strong in this analysis; the biggest drop in signal was at rs247616 in CETP, which dropped from 16.2 to 12.2.

All these genes have been previously identified as containing SNPs associated with LDL-C in a large (univariate) meta-analysis of over 100,000 individuals [1], and also identified as being associated with lipid subfractions measured by an NMR-based method, which distinguishes three subfractions of IDLs and LDLs [2]. However, our more detailed measurements, which distinguish 10 subfractions of IDLs and LDLs, provide greater resolution to examine effects of these genetic variants on subfractions of IDLs and LDLs than previous studies. Fig 2 shows the mean value of each (normalized) subfraction phenotype by genotype class, for the most strongly associated SNP in each gene ("top SNP"). The four SNPs exhibit distinctively different patterns, presumably reflecting different roles these genes play in lipoprotein metabolism. Among them, only rs7528419 in SORT1 has an effect that is consistent in direction among all subfractions, although it primarily affects the higher-density LDL subfractions (l3b, l4a and l4b; see Table 1 for abbreviations) as reported previously [21]. This SNP is in high LD ($r^2 = 1$) with the SNP rs12740374 that has been identified as a causal variant contributing to change in plasma LDL-C [26]. The SNP rs247616 in CETP shows a complex pattern of association, with the C allele showing an increase in both the lowest density (i1 and i2) and highest density subfractions (l3a and l3b), but a decrease in the middle of the range (l1 and l2a). Similarly, SNP rs7412 in APOE also affects a wide range of subfractions, with the minor allele increasing some (i1) and decreasing others (l1, l2a, l2b, l3a and l3b). The effect of SNP rs55730499 in LPA is concentrated entirely on the lowest density subfraction (i1) and this SNP is in high LD ($r^2 = 0.9$) with the SNP rs10455872 that has been reported as associated with plasma lipoprotein(a) [Lp(a)] levels [27, 28]. This result probably reflects the fact that Lp(a) particles are in the same size range as the IDL1 subfraction, and likely contribute to this subfraction measurement; that is, this result may reflect an association of this SNP with Lp(a), rather than with the IDL1 subfraction [29].

Multiple Independent Associations in CETP, APOE and LPA.   Previous analyses (e.g. [1] and [3]) have indicated that many lipid-affecting genes contain multiple variants that independently affect lipid-related phenotypes. In addition, for example, [3] has shown that multiple variants in some genes affect phenotypes similarly (or dissimilarly). We therefore searched for additional independently-associated SNPs near these four genes (±150kb). We did this for each gene by performing a multivariate association analysis on the residuals obtained by regressing out the effects of the most strongly associated SNP in that gene. Although no secondary SNPs showed signals that would be considered "genome-wide" significant, CETP, APOE and LPA all contained secondary SNPs with moderately strong signals ($\log_{10} BF_{av} > 2$; see Table 3 for details of the secondary SNPs). Furthermore, the secondary SNPs at CETP (rs11076175) and LPA (6–161069320) showed effects on the subfraction profiles that were, qualitatively, strikingly similar to those of the top SNP in those genes (Fig 2).

In contrast the secondary SNP at APOE (rs157581) showed a very different pattern to the top SNP, with the minor allele showing a consistent increase in all subfractions. That is, the top and secondary SNPs in CETP (and also LPA) have similar phenotypic consequences, possibly reflecting a similar underlying molecular effect, whereas the two SNPs in APOE have different phenotypic consequences, presumably reflecting different underlying molecular effects. The top SNP (rs7412) for APOE is one of two nonsynonymous SNPs in APOE exon 4 that define the $\epsilon_1/\epsilon_2/\epsilon_3$ haplotype system. The other variant (rs429358) has a modest association in the secondary analysis ($\log_{10} BF_{av} > 1.77$) and is not in LD with the top SNP ($r^2 < 0.01$).

The CETP locus is of considerable interest in light of controversies surrounding the relation of CETP genetic variation to risk of cardiovascular disease, and the failure of pharmacologic inhibitors of CETP to reduce this risk in recent clinical trials [30]. Given the very striking
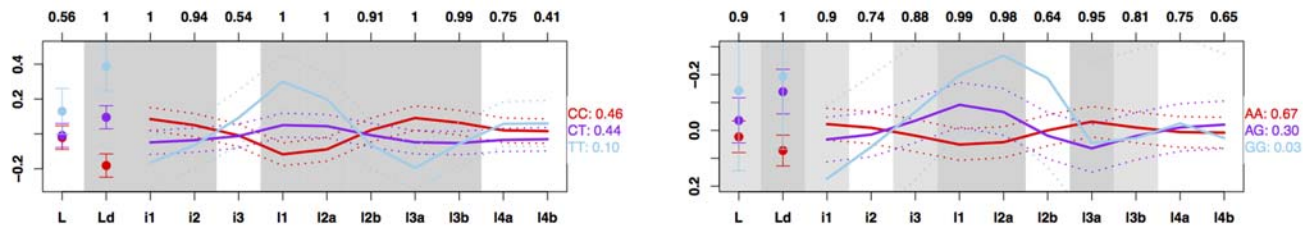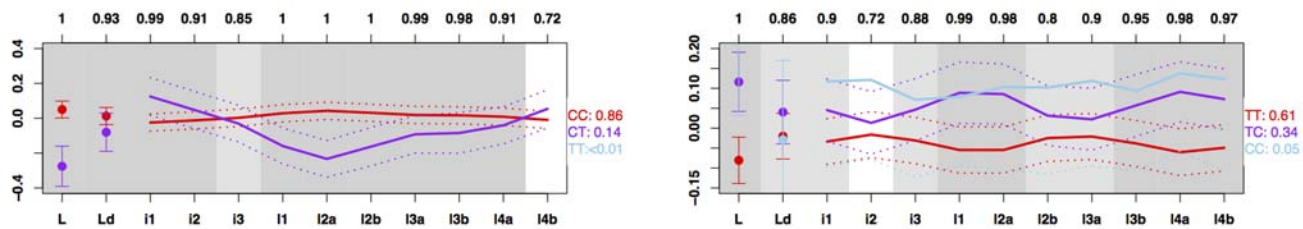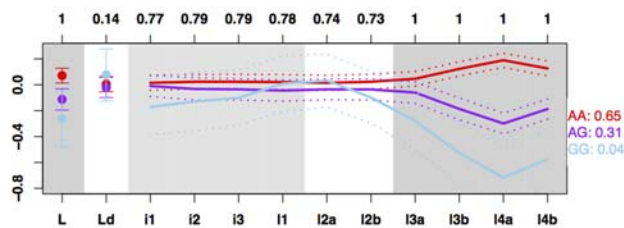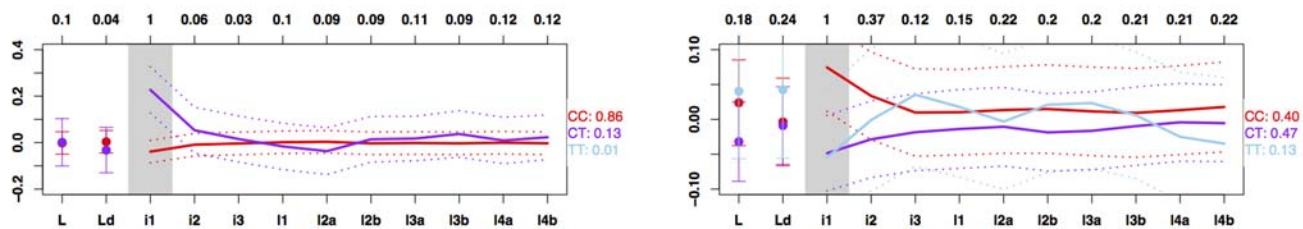
Fig 2. Effects of associated SNPs on LDL subfractions. Solid lines show mean (normalized) phenotype for each subfraction (see Table 1 for abbrevations) by genotype class (reference homozygotes: red, heterozygotes: purple, non-reference homozygotes: sky blue; the proportion of individuals is shown next to the genotype; sky blue lines are omitted if the proportion is ≤ 0.01); dotted lines show ±2 standard errors. Results for secondary SNPs are based on residuals from regressing out top SNPs. Grey shading indicates posterior probability of association (either directly or indirectly) for each phenotype (> 0.9: dark grey, > 0.75: light grey, < 0.75: white; raw numbers given at top of figure). Note that, because the minor alleles have opposite effects on total HDL-C at the two SNPs in CETP, the *y*-axis is reversed for the secondary SNP to emphasize the similar shapes of the curves.

doi:10.1371/journal.pone.0120758.g002

Table 3. Summary of the secondary associations with LDL subfractions in our analysis.

Gene	chr	SNP	log$_{10}$ BF$_{av}$(LDL subfractions)[1]	log$_{10}$ BF(LDL-C)[1]	Imputation quality[2]	MAF[3]
APOE	19	rs157581	2.68	3.62	0.94	0.22
LPA	6	6–161069320	2.65	-0.44	0.61	0.36
CETP	16	rs11076175	2.34	-0.05	+	0.18

See S1 Table in Supporting Information for other detailed results.

[1] Results from a multivariate association analysis on the residuals obtained by regressing out the effects of the most strongly associated SNP in each gene.

[2] Imputation quality reported from the software IMPUTE2 [10]; "+" for the typed SNP.

[3] Minor allele frequency.

doi:10.1371/journal.pone.0120758.t003

patterns of association at CETP, we sought to replicate this association in an independent population. To this end, we obtained LDL subfraction measurements, measured by ion mobility technology, for up to 6745 individuals of European ancestry from the JUPITER study (see Methods). To replicate the results for our top SNP, we estimated its effect on each subfraction using simple linear regression; to replicate results for the secondary SNP we estimated its effect on each subfraction after controlling for the top SNP. In both cases these replication data show effect size estimates that are highly concordant with those in our original study population (Fig 3), demonstrating that the highly significant initial findings also generalize to other populations.

As CETP has been known for strong effect on total HDL-C (the top and secondary SNPs have log$_{10}$ BF > 20 in our study), we investigated whether the association of the top and secondary SNPs with subfractions is partially mediated by association with HDL-C. To do this, for the five phenotypes, 'Ld', 'i1', 'l1', 'l2a', and 'l3a', with strong association signals without controlling for HDL-C (p-value < 0.005 from a standard simple linear regression), we examined how effect on the subfractions changes after controlling for HDL-C. The effect size of each SNP after controlling HDL-C was obtained by including the SNP and HDL-C as a covariate in a standard multiple linear regression. The percentage changes in effect size of the top SNP rs247616 after controlling for HDL-C are -72.7, 4.4, -89.6, -67.4, and -97.5 for 'Ld', 'i1', 'l1', 'l2a', and 'l3a', respectively (similar changes for the secondary SNP rs11076175); see S3 Fig in
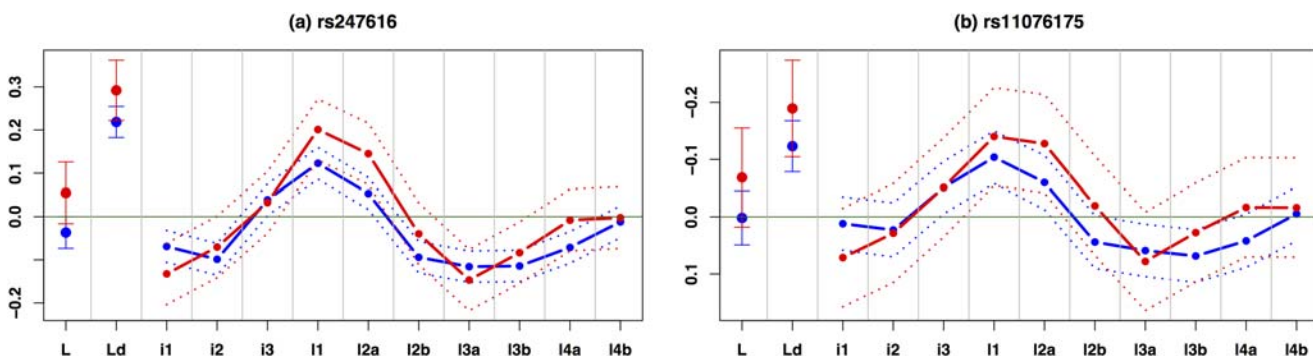

Fig 3. Replication of two independent associations in CETP (top SNP rs247616 and secondary SNP rs11076175). Dashed lines show estimated effect sizes on (normalized) phenotype for each subfraction in our study (red) and JUPITER (blue); dotted lines show ±2 standard errors. Note that the y-axis is reversed for the secondary SNP as in Fig 2.

doi:10.1371/journal.pone.0120758.g003

Supporting Information. Thus effect size for 'i1' is essentially unaffected by controlling for HDL-C, whereas effects on other subfractions are very substantially reduced (after controlling for HDL-C, the signal for the top SNP in analyses of 12 subfractions dropped from 16.2 to 3.43). These results are consistent with the SNP's effect on LDL subfractions being largely mediated through its effect on HDL-C.

Comparisons with previous studies

Chasman et al. [2] performed association studies on lipoprotein subfraction measurements obtained by NMR-based methods. Tukiainen et al. [3] used an NMR-based serum metabolomics platform to quantify 216 metabolic variables, some of which could be related to the subfraction measurements used here. For example, large LDL and small LDL subfractions measured by an NMR-based method correspond most closely to LDL1/LDL2a and LDL3a/LDL3b, respectively, in our study (see Table 2 in [31]).

Compared with these previous studies, the association results we report here for the top SNP in CETP are mostly novel. Our top SNP in CETP (rs247616) is not in high LD ($r^2 < 0.5$) with any SNP reported as associated with NMR-based subfractions in [2]. The SNP is in high LD with a SNP (rs3764261) reported by [3], but they found no significant association between this SNP and any LDL-related measurements (their Table 1, bottom line).

The secondary SNP we report for CETP (rs11076175) is in high LD ($r^2 = 0.95$) with one (rs7499892) of the SNPs reported in [2]. Although [2] makes no mention of the distinctive patterns of association at this SNP, the estimated effects reported in their supplementary information (Figure S4 in [2]) shows patterns concordant with those we highlight here: the effects on total IDL and small (i.e. high density) LDL particles are opposite in direction to the effect on large (low-density) LDL particles. Indeed, further examining their Figure S4 we note that there are other SNPs in CETP, not in high LD ($r^2 < 0.5$) with either the top or secondary SNP in our analysis, that show similar distinctive association patterns, lending further support to the idea that many of the SNPs affecting lipids in CETP may have similar affects on subfractions, perhaps through a shared mechanism.

At APOE, both our top and secondary associations are distinct from those reported in [2] ($r^2 < 0.5$ with any association reported there). Our top SNP (rs7412) was reported in the serum metabolomics study [3], with their most strongly associated phenotype being 'enzymatically measured LDL-C' (Table 2 and Table S10 in [3]).

The top SNP in SORT1 from our analysis is in high LD ($r^2 = 1$) with SNPs reported as associated in the previous NMR-based studies (rs646776 in [2] and rs629301 in [3]). However, neither of these previous studies has the resolution to highlight the distinctive patterns of effects highlighted in the gradient gel electrophoresis results, with by far the strongest associations being with the smallest and densest LDL subfractions (3b/4a/4b).

The two independent associations for LPA from our analysis were not reported in either of [2, 3]. Since these association may be due to associations with Lp(a), rather than with an LDL subfraction, this difference may reflect different sensitivities of the technologies used to Lp(a). Interestingly, our secondary SNP (6–161069320) is not in LD ($r^2 < 0.25$) with any of five SNPs that have been reported in the NHGRI GWAS catalog [32] as associated with Lp(a) levels, and so may represent a novel association with Lp(a).

Genome-wide association analysis of statin response

To identify SNPs associated with response to statin we performed a multivariate genome-wide association analysis using the difference between post-treatment and pre-treatment measures of each phenotype as a multivariate outcome. These differences, $\tilde{D}$, were obtained using the

standardized (quantile-normalized) phenotypes (see Methods), so this association analysis targets genetic variants whose *standardized* effect on phenotype changes pre- and post- treatment. For example, a SNP that explains 1% of phenotypic variance before treatment and 2% after treatment would be associated with response to statin in our analysis.

Only one SNP (rs6430626) had $\log_{10} BF_{av} > 5$ (actual $\log_{10} BF_{av} = 5.3$). The phenotypes contributing to this strong multivariate signal included a modest (univariate) association with total LDL-C response ($\log_{10} BF = 1.7$). However, imputation quality at this SNP was low (0.25), and a substantial part of the association signal appeared to be driven by deviations from multivariate normality: on removing 96 individuals detected as having outlying (multivariate) phenotypes at the 0.01 significance threshold (see Methods), $\log_{10} BF_{av}$ fell to 2.3. Furthermore, neither this SNP, nor other SNPs in LD with it, showed even nominal associations with (total) LDL response to statins in the JUPITER study (*p* values near 0.9).

Although no other SNPs showed strong evidence for association with statin response in our analysis, our data also provide the potential opportunity to assess which subfractions of LDL are most likely responsible for previously-observed associations with LDL-C response to statin in the much larger JUPITER study [7]. To this end we examined the three SNPs (near the genes ABCG2, APOE and LPA) highlighted in [7] as being associated with fractional change in LDL-C (at *p* value $< 5 \times 10^{-8}$).

The SNP rs1481012 near ABCG2 showed no signal for association with response to statin in either the subfractions or total LDL-C ($\log_{10} BF_{av} = -0.65$). Possibly this lack of signal could reflect our small sample size, although it could also be due to the different statin drugs used (CAP used simvastatin; PRINCE used pravastatin; JUPITER used rosuvastatin).

The SNPs rs7412 in APOE and rs10455872 in LPA had modest association signals with statin response in our data (rs7412: $\log_{10} BF_{av} = 1.6$ and rs10455872: $\log_{10} BF_{av} = 0.8$). Both these SNPs are associated with LDL phenotypes both before and after statin treatment, so these associations with statin response must reflect a difference in the *strengths* of these associations in the untreated vs treated condition. That is, they show a statistical interaction with treatment. To gain additional insights into these interactions we examined whether the effect sizes are stronger in the treated or the untreated data. The APOE SNP showed consistently stronger estimated effects in the treated condition compared with the untreated condition, across the smaller particles (l2a-l4a; Fig 4a). The LPA SNP also showed stronger effects in treated vs untreated conditions, but with the effects concentrated in the IDL1, and, to a lesser extent, IDL2 subfraction (Fig 4b). As noted above, the association of this SNP with the IDL1 subfraction likely reflects, at least in part, the effect of this SNP on Lp(a) particles. The stronger association with IDL1 post-treatment may therefore be explained by the fact that Lp(a) levels are resistant to statin treatment [27], and so will contribute more strongly to the post–statin measures than the pre-statin measures.

Comparison of multivariate and univariate association analyses

In our genome-wide association analysis irrespective of statin exposure, all four of the strongest associations (in or near CETP, APOE, SORT1, and LPA) showed very much stronger association signals in the multivariate analysis than in a univariate analysis of LDL-C. Two factors could contribute to this increased signal: first, our use of more-detailed phenotypes (the subfractions), any of which could show individually stronger association signals than total LDL-C; and second, our analysis of these phenotypes simultaneously, in a multivariate way, rather than one at a time. We now investigate the contributions of these two factors to increased association signals in more detail. To do this we compare three different Bayes Factors that test for association: the BF based on univariate analysis of LDL-C (BF_{ldl}), the BF based on univariate

(a) rs7412 (APOE)
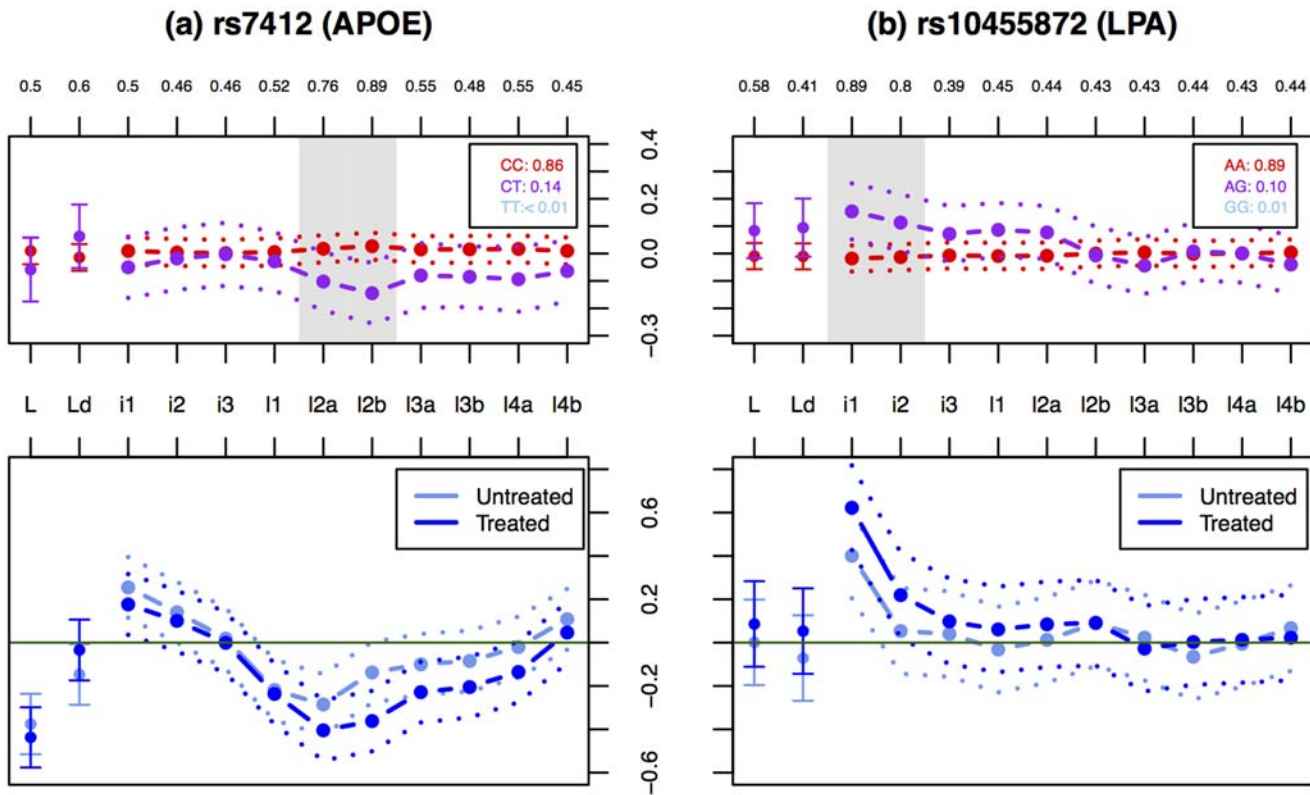
(b) rs10455872 (LPA)

Fig 4. Effects of associated SNPs (rs7412 in APOE and rs10455872 in LPA) on treated and untreated measures of LDL subfractions, and on their difference. Upper panels show the effects of the SNPs on the difference between treated and untreated measures of LDL subfractions. Labels and colors are as in Fig 2. Lower panels show estimated effect size (dashed lines) and ±2 standard errors (dotted lines) in the treated and untreated (normalized) phenotype for each subfraction.

analysis of all 12 phenotypes (BF_{uni}) and the BF based on multivariate analysis of all 12 phenotypes (BF_{av}). We observed above that for the strongest associations BF_{av} was many orders of magnitude larger than BF_{ldl}. We now attempt to decompose this gain in signal into two parts, using the identity

$$\log_{10}(BF_{av}/BF_{ldl}) = \log_{10}(BF_{uni}/BF_{ldl}) + \log_{10}(BF_{av}/BF_{uni}). \qquad (3)$$

Intuitively, we can think of the first term as capturing the gain (or loss) from the more detailed subfraction measurements, and the second term as capturing the gain (or loss) in going from a univariate to a multivariate analysis.

To increase the number of SNPs contributing to this assessment, we compare not only the very strongest associations, but also more modest association signals. To attempt to ensure that these more modest associations nonetheless reflect likely genuine associations we focus on SNPs in or near the 95 genes (±150kb) identified as being associated with lipid traits in [1] ("Global Lipids genes"). We included all the genes from this study, and not only those associated with LDL-C, because we wanted to allow for the fact that, as illustrated in our main analysis, SNPs may be strongly associated with LDL subfractions without being strongly associated with LDL-C. We include in our comparisons any SNP for which any of the BFs (BF_{av}, BF_{ldl}, BF_{uni}) exceed 10^3. Given the modest nature of some of these associations we cannot be completely

confident that all of them are genuine, although it seems reasonable to believe that most of them are.

[Fig 5](#) breaks down the gain (or loss) from using BF_{av} vs BF_{ldl} (panel a) into a component due to the more detailed measurements (panel b) and a component due to the multivariate analysis (panel c). For the majority of associations, both the use of more detailed phenotypic measurements and the use of multivariate association analysis contribute to an increase in the association signal. For some associations (e.g. in SORT1, APOA1) the main gain comes from the more detailed phenotype measurements; for others (e.g. some associations in APOE, LPA) the increase in association signal is greater in moving from the univariate to multivariate analysis.

It is natural to ask what features of the different associations explain these different behaviors, and, more generally, under what circumstances more detailed measurements or multivariate analyses may prove most helpful. For multivariate analyses, the general answer is that the gains in multivariate analyses are largest when the effects are discordant with the correlation among phenotypes. For example, for two positively correlated phenotypes, the gains of multivariate analysis will be greatest when the effects on the two phenotypes have opposite signs; there will also be substantial gains when one phenotype has an effect and the other has zero effect; the gains will be smallest when both phenotypes have an effect of the same sign. See [22] for more discussion. This observation helps explain the gain of the multivariate analyses for APOE and LPA associations, both of which exhibit some discordance between effect and correlation structure: for APOE, some negatively-correlated subfractions (e.g. l1 and l3a; see [S1 Fig](#) in Supporting Information) show effects in the same direction ([Fig 2](#)); for LPA, subfraction i1 shows a substantive effect, and many subfractions correlated with this subfraction (both positively and negatively) show no effect. It is harder to generalize as to when more detailed measurements will increase power to detect associations. In our study, the largest gains in power from the subfraction measurements were at SORT1 and at CETP, and these gains appear to be for different reasons. At CETP, the gain is due to the genetic variants affecting subfractions having effects in different directions, which approximately "cancel out" so that the effect on total LDL-C is small compared with the effects on individual subfractions. At SORT1, the gain is because the genetic variants affect primarily the smaller (l3b, l4a, l4b) subfractions, which together represent only a small proportion of total LDL-C, and contribute correspondingly little to overall variation in total LDL-C (See [Table 1](#)). That is, the genetic effects on these subfractions are diluted by variation in the other more abundant subfractions (l1-l3a) when testing for association with LDL-C.

We note that, of course, power is only one potential benefit of the subfraction measurements: even when the more detailed measurements do not increase power to detect associations, being able to estimate the effects on different subfractions may still be useful (e.g. because the different subfractions have different biologic and clinical implications).

Discussion

We have presented an association analysis of genetic variants with detailed measurements of IDL/LDL subfractions, and highlighted several SNPs that, while only weakly associated with total LDL-C in our study, are strongly associated with the subfractions. Perhaps the most interesting of the associations are the two independently-associated SNPs in CETP. From previous studies these variants are known to be very strongly associated with plasma HDL-C ($\log_{10}$ BF > 20 in our study), and less strongly associated with total LDL-C. For example, the lead SNP in [1] explains 1.5% of the variance in HDL-C, but only 0.06% of the variance in LDL-C (see [S1 Text](#) in Supporting Information). Our data demonstrate that, nonetheless, these
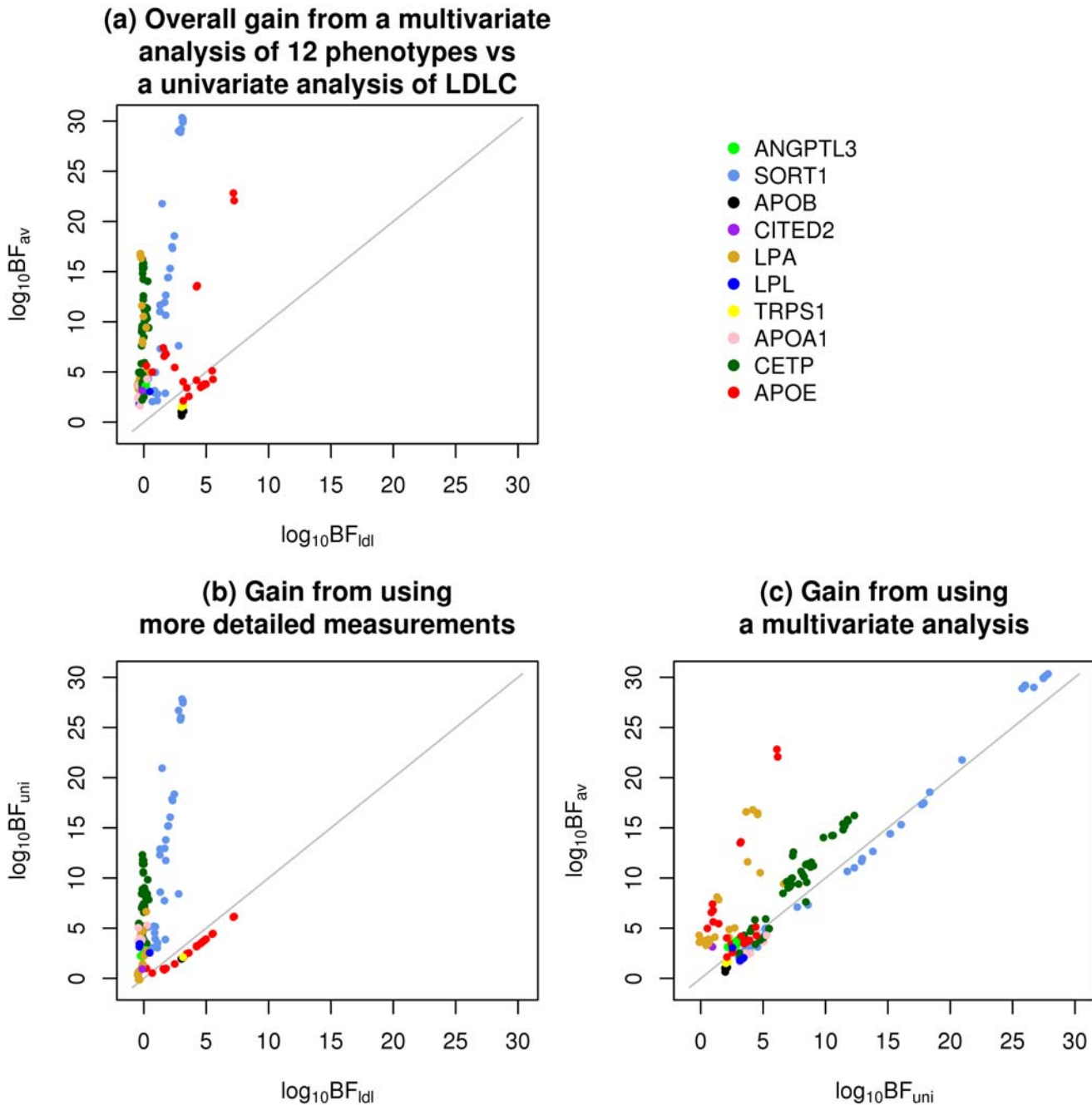
Fig 5. Decomposition of the gain (or loss) from a multivariate analysis of 12 phenotypes vs a univariate analysis of LDL-C into two components: one from using more detailed measurements, and one from using a multivariate analysis. Plotted are (a) $\log_{10} BF_{av}$ (the BF based on multivariate analysis of all 12 phenotypes) vs $\log_{10} BF_{ldl}$ (the BF based on univariate analysis of LDL-C), (b) $\log_{10} BF_{uni}$ (the BF based on univariate analysis of all 12 phenotypes) vs $\log_{10} BF_{ldl}$, and (c) $\log_{10} BF_{av}$ vs $\log_{10} BF_{uni}$. SNPs are colored according to the nearest gene.

doi:10.1371/journal.pone.0120758.g005

variants are very strongly associated with multiple subfractions of LDL, with effects in different directions on different subfractions approximately cancelling each other out to yield a small overall effect on total LDL-C. Interestingly, these independent variants have similar effects on the different subfractions, consistent with a shared underlying functional mechanism that may involve selective CETP-mediated cholesterol-triglyceride exchange between HDL and specific subclasses of apoB-containing particles [33]. The findings complement previous analyses of LDL subfractions in two hyperalphalipoproteinemic patients with genetic deficiency of CETP [34], and recent data on the effects of the CETP inhibitor anacetrapib on lipoprotein subfraction concentrations [33], both of which indicate that CETP influences LDL subclasses in a manner consistent with the CETP genotype associations shown here. Specifically, reduced CETP activity in both instances results in an increase in small LDL particles together with a reduction in mid-sized LDL. While the mechanism for this effect is not known, it has been suggested that it is related to the existence of separate pathways giving rise to larger vs. mid-sized and smaller LDL [18, 34], with CETP deficiency selectively resulting in triglyceride enrichment of particles in the second of these pathways, and a resultant shift from mid-sized to smaller LDL through the action of lipase activity [33, 34].

Our study also provides a practical illustration of the potential benefits of joint association analysis of multiple related phenotypes in genetic association studies, rather than considering each phenotype one at a time. While in principle the potential benefits of joint analysis are well established, and have been highlighted in several recent papers (e.g. [35–42]), in practice the vast majority of association studies of multiple phenotypes continue to rely on univariate analyses. We hope that our comparisons provide further motivation to investigators to consider joint multivariate analyses as a tool to help improve power to detect genetic associations.

Supporting Information

S1 Text. Computing proportion of variance in phenotype explained by a given SNP (PVE). (PDF)

S2 Text. Detailed results from the full analysis, irrespective of statin exposure. Results from the full analysis for 24,689 SNPs (24,601 SNPs which pass through our initial filtering step and 88 SNPs reported in the large meta-analysis of four lipid traits by [1]): rs number, chromosome, position, two alleles, minor allele frequency, imputation quality, $\log_{10} BF_{av}$, $\log_{10} BF_{all}$, $\log_{10} BF$ from a univariate analysis of each phenotype, and marginal posterior probability of direct, indirect, no association for each phenotype. (TXT)

S3 Text. Detailed results from the full analysis of statin response. Results from the full analysis for 25,175 SNPs which pass through our initial filtering step: rs number, chromosome, position, two alleles, minor allele frequency, imputation quality, $\log_{10} BF_{av}$, $\log_{10} BF_{all}$, $\log_{10} BF$ from a univariate analysis of each phenotype, and marginal posterior probability of direct, indirect, no association for each phenotype. (TXT)

S1 Fig. Spearman correlations among 16 phenotypes. In addition to 12 phenotypes considered in our main analysis, correlations with triglycerides (Tg), HDL-cholesterol (H), total cholesterol (T), ApoB levels (A) are also included. See Table 1 in the main text for abbreviations for 12 phenotypes. Correlations are computed by using (a) (normalized) pre-treatment measurements (*P* in the methods section); (b) (normalized) post-treatment measurements (*T* in the methods); (c) (normalized) averages of pre-treatment measurements and post-treatment

measurements ($\tilde{A}$ in the methods); (d) (normalized) differences between post-treatment and pre-treatment measures ($\tilde{D}$ in the methods).
(TIFF)

S2 Fig. Regional plots of the SORT1, APOE, LPA and CETP. Regional plots of the four loci are created by using the tool LocusZoom [43] (with $\log_{10} BF_{av}$ in y-axis). Purple diamond is the top SNP with the strongest association in each locus. Each circle corresponds to a SNP whose color indicates the linkage disequilibrium with the top SNP. LDL subfraction associations around (a) SORT1, (b) APOE, (c) LPA, and (d) CETP.
(TIFF)

S3 Fig. Effect size with ±2 standard error before (blue) and after (red) controlling for HDL-C for each subfraction. (a) The top SNP rs247616 and (b) the secondary SNP rs11076175 in CETP. Note that, because the minor alleles have opposite effects on total HDL-C at the two SNPs, the y-axis is reversed for the secondary SNP to emphasize the similar pattern of effect size.
(TIFF)

S1 Table. Summary of the top and secondary associations with LDL subfractions in our analysis.
(PDF)

Acknowledgments

Author Contributions

Conceived and designed the experiments: RMK. Performed the experiments: HS DIC JDS SM PMR DAN RMK MS. Analyzed the data: HS RMK MS. Contributed reagents/materials/analysis tools: DIC JDS SM PMR DAN RMK. Wrote the paper: HS RMK MS. Conceived and designed the analyses: HS MS.

References

1. Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M, et al. Biological, clinical and population relevance of 95 loci for blood lipids. Nature. 2010; 466:707–713. doi: 10.1038/nature09270 PMID: 20686565

2. Chasman DI, ParNi G, Mora S, Hopewell JC, Peloso G, Clarke R, et al. Forty-three loci associated with plasma lipoprotein size, concentration, and cholesterol content in genome-wide analysis. PLoS Genet. 2009; 5(11):e1000730. doi: 10.1371/journal.pgen.1000730 PMID: 19936222

3. Tukiainen T, Kettunen J, Kangas AJ, Lyytiäinen LP, Soininen P, Sarin AP, et al. Detailed metabolic and genetic characterization reveals new associations for 30 known lipid loci. Human Molecular Genetics. 2012 Mar; 21(6):1444–55. doi: 10.1093/hmg/ddr581 PMID: 22156771

4. Kettunen J, Tukiainen T, Sarin AP, Ortega-Alonso A, Tikkanen E, Lyytikinen LP, et al. Genomewide association study identifies multiple loci influencing human serum metabolite levels. Nature Genetics. 2012; 44(3):269–76. doi: 10.1038/ng.1073 PMID: 22286219

5. Thompson JF, Hyde CL, Wood LS, Paciga SA, Hinds DA, Cox DR, et al. Comprehensive wholegenome and candidate gene analysis for response to statin therapy in the Treating to New Targets (TNT) cohort. Circulation Cardiovascular genetics. 2009 Apr; 2(2):173–81. doi: 10.1161/CIRCGENETICS.108.818062 PMID: 20031582

6. Barber MJ, Mangravite LM, Hyde CL, Chasman DI, Smith JD, McCarty Ca, et al. Genome-wide association of lipid-lowering response to statins in combined study populations. PloS one. 2010 Jan; 5(3): e9763. doi: 10.1371/journal.pone.0009763 PMID: 20339536

7. Chasman DI, Giulianini F, Macfadyen J, Barratt BJ, Nyberg F, Ridker PM. Genetic Determinants of Statin Induced LDL-C Reduction: The JUPITER Trial. Circulation Cardiovascular Genetics. 2012 Feb; 5:257–264. doi: 10.1161/CIRCGENETICS.111.961144 PMID: 22331829

8. Simon Ja, Lin F, Hulley SB, Blanche PJ, Waters D, Shiboski S, et al. Phenotypic predictors of response to simvastatin therapy among African-Americans and Caucasians: the Cholesterol and Pharmacogenetics (CAP) Study. The American journal of cardiology. 2006 Mar; 97(6):843–50. doi: 10.1016/j.amjcard.2005.09.134 PMID: 16516587

9. Albert MA, Danielson E, Rifai N, Ridker PM. Effect of Statin Therapy on C-Reactive Protein Levels: The Pravastatin Inammation/CRP Evaluation (PRINCE): A Randomized Trial and Cohort Study. The Journal of the American Medical Association. 2001; 286:64–70. doi: 10.1001/jama.286.1.64

10. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS genetics. 2009 Jun; 5(6):e1000529. doi: 10.1371/journal.pgen.1000529 PMID: 19543373

11. Servin B, Stephens M. Imputation-based analysis of association studies: candidate regions and quantitative traits. PLoS genetics. 2007 Jul; 3(7):e114. doi: 10.1371/journal.pgen.0030114 PMID: 17676998

12. Mangravite LM, Engelhardt BE, Medina MW, Smith JD, Brown CD, Chasman DI, et al. A statindependent QTL for GATM expression is associated with statin-induced myopathy. Nature. 2013 Oct; 502 (7471):377–80. doi: 10.1038/nature12508 PMID: 23995691

13. Durbin RM, Altshuler DL, Abecasis GR, Bentley DR, et al AC. A map of human genome variation from population-scale sequencing. Nature. 2010 Oct; 467(7319):1061–73. doi: 10.1038/nature09534 PMID: 20981092

14. The International HapMap Consortium. A haplotype map of the human genome. Nature. 2005 Oct; 437 (7063):1299–320. doi: 10.1038/nature04226 PMID: 16255080

15. Friedewald WT, Levy RI, Fredrickson DS. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. Clin Chem. 1972; 18:499–502. PMID: 4337382

16. Williams PT, Vranizan KM, Krauss RM. Correlations of plasma lipoproteins with LDL subfractions by particle size in men and women. J Lipid Res. 1992; 33:765–774. PMID: 1619368

17. Rainwater DL, Mitchell BD, Comuzzie AG, Haffner SM. Relationship of low-density lipoprotein particle size and measures of adiposity. Int J Obes Relat Metab Disord. 1999; 23:180189. doi: 10.1038/sj.ijo.0800813

18. Berneis K, Krauss RM. Metabolic origins and clinical significance of LDL heterogeneity. J Lipid Res. 2002; 43:1363–1379. doi: 10.1194/jlr.R200004-JLR200 PMID: 12235168

19. Ridker PM, Danielson E, Fonseca FAH, Genest J, Gotto AM, Kastelein JJP, et al. Rosuvastatin to prevent vascular events in men and women with elevated C-reactive protein. The New England journal of medicine. 2008 Nov; 359(21):2195–207. doi: 10.1056/NEJMoa0807646 PMID: 18997196

20. Caulfield MP, Li S, Lee G, Blanche PJ, Salameh WA, Benner WH, et al. Direct determination of lipoprotein particle sizes and concentrations by ion mobility analysis. Clinical chemistry. 2008 Aug; 54 (8):1307–16. doi: 10.1373/clinchem.2007.100586 PMID: 18515257

21. Musunuru K, Orho-Melander M, Caulfield MP, Li S, Salameh Wa, Reitz RE, et al. Ion mobility analysis of lipoprotein subfractions identifies three independent axes of cardiovascular risk. Arteriosclerosis, thrombosis, and vascular biology. 2009 Nov; 29(11):1975–80. doi: 10.1161/ATVBAHA.109.190405 PMID: 19729614

22. Stephens M. A unified framework for association analysis with multiple related phenotypes. PloS one. 2013 Jan; 8(7):e65245. doi: 10.1371/journal.pone.0065245 PMID: 23861737

23. Wakefield J. Bayes factors for genome-wide association studies: comparison with P-values. Genetic epidemiology. 2009 Jan; 33(1):79–86. doi: 10.1002/gepi.20359 PMID: 18642345

24. Stephens M, Balding DJ. Bayesian statistical methods for genetic association studies. Nature reviews Genetics. 2009 Oct; 10(10):681–90. doi: 10.1038/nrg2615 PMID: 19763151

25. Campos H, Arnold KS, Balestra ME, Innerarity TL, Krauss RM. Differences in receptor binding of LDL subfractions. Arteriosclerosis, Thrombosis, and Vascular Biology. 1996 Jun; 16(6):794–801. doi: 10.1161/01.ATV.16.6.794 PMID: 8640407

26. Musunuru K, Strong A, Frank-Kamenetsky M, Lee NE, Ahfeldt T, Sachs KV, et al. From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. Nature. 2010 Aug; 466(7307):714–9. doi: 10.1038/nature09266 PMID: 20686566

27. Deshmukh HA, Colhoun HM, Johnson T, Mckeigue PM, John D, Durrington PN, et al. Genomewide Association Study of Genetic Determinants of LDL-c Response to Atorvastatin Therapy: Importance of Lp (a). Journal of lipid research. 2012; 53(5):1000–11. doi: 10.1194/jlr.P021113 PMID: 22368281

28. Qi Q, Workalemahu T, Zhang C, Hu FB, Qi L. Genetic variants, plasma lipoprotein(a) levels, and risk of cardiovascular morbidity and mortality among two prospective cohorts of type 2 diabetes. European heart journal. 2012 Feb; 33(3):325–34. doi: 10.1093/eurheartj/ehr350 PMID: 21900290

29. Dreon DM, Fernstrom HA, Williams PT, Krauss RM. LDL subclass patterns and lipoprotein response to a low-fat, high-carbohydrate diet in women. Arteriosclerosis, Thrombosis, and Vascular Biology. 1997; 17(4):707–714. doi: 10.1161/01.ATV.17.4.707 PMID: 9108784

30. Rader DJ, DeGoma EM. Future of cholesteryl ester transfer protein inhibitors. Annual review of medicine. 2014 Jan; 65:385–403. doi: 10.1146/annurev-med-050311-163305 PMID: 24422575

31. Williams PT, Zhao XQ, Marcovina SM, Otvos JD, Brown BG, Krauss RM. Comparison of four methods of analysis of lipoprotein particle subfractions for their association with angiographic progression of coronary artery disease. Atherosclerosis. 2014 Apr; 233(2):713–20. doi: 10.1016/j.atherosclerosis.2014.01.034 PMID: 24603218

32. Hindorff L, MacArthur J, Morales J, Junkins H, Hall P, Klemm A, et al. A Catalog of Published Genome-Wide Association Studies.;. Available at: www.genome.gov/gwastudies. Accessed [date of access].

33. Krauss RM, Wojnooski K, Orr J, Geaney JC, Pinto CA, Liu Y, et al. Changes in lipoprotein subfraction concentration and composition in healthy individuals treated with the CETP inhibitor anacetrapib. Journal of lipid research. 2012 Mar; 53(3):540–7. doi: 10.1194/jlr.M018010 PMID: 22180633

34. Sakai N, Matsuzawa Y, Hirano K, Yamashita S, Nozaki S, Ueyama Y, et al. Detection of Two Species of Low Density Lipoprotein Particles in Cholesteryl Ester Transfer Protein Deficiency. Arteriosclerosis, thrombosis, and vascular biology. 1991; 11:71–79. doi: 10.1161/01.ATV.11.1.71

35. Verzilli CJ, Stallard N, Whittaker JC. Bayesian modelling of multivariate quantitative traits using seemingly unrelated regressions. Genet Epidemiol. 2005 May; 28(4):313–25. doi: 10.1002/gepi.20072 PMID: 15789447

36. Banerjee S, Yandell BS, Yi N. Bayesian quantitative trait loci mapping for multiple traits. Genetics. 2008 Aug; 179(4):2275–89. doi: 10.1534/genetics.108.088427 PMID: 18689903

37. Kim S, Sohn KA, Xing EP. A multivariate regression approach to association analysis of a quantitative trait network. Bioinformatics. 2009 Jun; 25(12):i204–12. doi: 10.1093/bioinformatics/btp218 PMID: 19477989

38. Kim S, Xing EP. Statistical estimation of correlated genome associations to a quantitative trait network. PLoS Genet. 2009 Aug; 5(8):e1000587. doi: 10.1371/journal.pgen.1000587 PMID: 19680538

39. Baker AR, Goodloe RJ, Larkin EK, Baechle DJ, Song YE, Phillips LS, et al. Multivariate association analysis of the components of metabolic syndrome from the Framingham Heart Study. BMC Proc. 2009; 3 Suppl 7:S42. doi: 10.1186/1753-6561-3-s7-s42 PMID: 20018034

40. Zhang L, Pei YF, Li J, Papasian CJ, Deng HW. Univariate/multivariate genome-wide association scans using data from families and unrelated samples. PLoS One. 2009; 4(8):e6502. doi: 10.1371/journal.pone.0006502 PMID: 19652719

41. Ferreira MAR, Purcell SM. A multivariate test of association. Bioinformatics. 2009 Jan; 25(1):132–3. doi: 10.1093/bioinformatics/btn563 PMID: 19019849

42. O'Reilly PF, Hoggart CJ, Pomyen Y, Calboli FCF, Elliott P, Jarvelin MR, et al. MultiPhen: Joint Model of Multiple Phenotypes Can Increase Discovery in GWAS. PLoS One. 2012; 7(5):e34861. doi: 10.1371/journal.pone.0034861 PMID: 22567092

43. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics (Oxford, England). 2010 Sep; 26 (18):2336–7. doi: 10.1093/bioinformatics/btq419

Author/s:
Shim, H; Chasman, DI; Smith, JD; Mora, S; Ridker, PM; Nickerson, DA; Krauss, RM; Stephens, M

Title:
A Multivariate Genome-Wide Association Analysis of 10 LDL Subfractions, and Their Response to Statin Treatment, in 1868 Caucasians

Date:
2015-04-21

Citation:
Shim, H., Chasman, D. I., Smith, J. D., Mora, S., Ridker, P. M., Nickerson, D. A., Krauss, R. M. & Stephens, M. (2015). A Multivariate Genome-Wide Association Analysis of 10 LDL Subfractions, and Their Response to Statin Treatment, in 1868 Caucasians. PLOS ONE, 10 (4), https://doi.org/10.1371/journal.pone.0120758.

Persistent Link:
http://hdl.handle.net/11343/246301

File Description:
published version
License:
CC BY