nature genetics

Exome-wide association study of plasma lipids in >300,000 individuals

We screened variants on an exome-focused genotyping array in >300,000 participants (replication in >280,000 participants) and identified 444 independent variants in 250 loci significantly associated with total cholesterol (TC), high-density-lipoprotein cholesterol (HDL-C), low-density-lipoprotein cholesterol (LDL-C), and/or triglycerides (TG). At two loci (*JAK2* and *A1CF*), experimental analysis in mice showed lipid changes consistent with the human data. We also found that: (i) beta-thalassemia trait carriers displayed lower TC and were protected from coronary artery disease (CAD); (ii) excluding the *CETP* locus, there was not a predictable relationship between plasma HDL-C and risk for age-related macular degeneration; (iii) only some mechanisms of lowering LDL-C appeared to increase risk for type 2 diabetes (T2D); and (iv) TG-lowering alleles involved in hepatic production of TG-rich lipoproteins (*TM6SF2* and *PNPLA3*) tracked with higher liver fat, higher risk for T2D, and lower risk for CAD, whereas TG-lowering alleles involved in peripheral lipolysis (*LPL* and *ANGPTL4*) had no effect on liver fat but decreased risks for both T2D and CAD.

Plasma lipid levels are modifiable risk factors for atherosclerotic cardiovascular disease. Genome-wide association studies (GWAS) testing common DNA-sequence variation have uncovered 175 genetic loci affecting lipid levels in the population[1–7]. These findings have provided information on lipoprotein biology and elucidated the causal roles of lipid levels in cardiovascular disease[8–11]. Here, we built on these previous efforts to: (i) perform an exome-wide association screen for plasma lipids in >300,000 individuals; (ii) experimentally evaluate discovered alleles; and (iii) test the inter-relationship of mapped lipid variants with CAD, age-related macular degeneration (AMD), fatty liver, and T2D.

We tested the association of genotypes from the HumanExome BeadChip (exome array) with lipid levels in 73 studies including >300,000 participants (**Supplementary Note** and **Supplementary Tables 1–3**) across several ancestries, with maximal sample sizes of 237,050 for participants of European ancestry, 16,935 for participants of African ancestry, 37,613 for participants of South Asian ancestry, and 5,082 for participants of Hispanic or other ancestry. A companion manuscript describes results for 47,532 East Asian participants[12]. A total of 242,289 variants were analyzed after quality control, approximately one-third of which were nonsynonymous, with minor allele frequency (MAF) <0.1% (**Supplementary Table 4**).

Single-variant association statistics and linkage-disequilibrium information summarized across 1-Mb sliding windows were generated from each cohort in RAREMETALWORKER or RVTESTS[13,14] software. Meta-analyses of single-variant and gene-level association tests were performed with rareMETALS (version 6.0). Genomic control values for meta-analysis results were between 1.09 and 1.14 for all four traits (**Supplementary Fig. 1**), suggesting that the population structure in our analysis was well controlled[3,15].

We identified 1,445 single variants associated at $P < 2.1 \times 10^{-7}$ (two-tailed score test with Bonferroni correction of 242,289 variants analyzed) (**Supplementary Figs. 2–5**). Full association results are available ('Data availability' section). Of those variants, 75 were 'novel' variants (located at least 1 Mb from previously reported GWAS signals): 35 were protein-altering variants, and 40 were noncoding variants (**Table 1** and **Supplementary Tables 5–7**). The MAF of the lead variant was >5% at 61 of these 75 loci. Participants of European ancestry provided the most significant associations for the 75 novel loci, with the exception of two LDL-associated variants (rs201148465 and rs147032017), which were driven by the South Asian participants (**Supplementary Table 8**). Gene-level association analyses identified an additional five genes whose signal was driven by multiple rare variants ($P < 4.2 \times 10^{-7}$, two-tailed score test with Bonferroni-correction threshold for performing five tests on ~20,000 genes, **Supplementary Table 9**).

We sought replication in up to 286,268 independent participants from three studies: the Nord-Trøndelag Health Study[16] (HUNT; maximum n = 62,168), the Michigan Genomics Initiative (MGI; maximum n = 6,411; URLs) and the Million Veteran Program (MVP; maximum n = 218,117). Of the novel primary trait associations, 73/73 associations were directionally consistent (**Supplementary Table 10**); two SNPs were not available for replication (rs201148465 and rs75862065). Furthermore, we were able to replicate the associations of 66/73 (90%) at $\alpha = 0.05$.

At any given genetic locus, multiple variants may independently contribute to plasma lipid levels. We quantified this phenomenon by iteratively performing association analyses conditional on the top variants at each locus. We identified 444 variants independently associated with one or more of the four lipid traits in 75 novel and 175 previously implicated loci (**Supplementary Fig. 6** and **Supplementary Tables 11** and **12**).

The identification of lipid-associated coding variants may help to refine association signals at previously identified GWAS loci. We were able to evaluate this possibility in 131 of the 175 previously reported GWAS loci where the index or proxy variant was available on the

A full list of authors and affiliations appears at the end of the paper.

Received 25 January; accepted 26 September; published online 30 October 2017; doi:10.1038/ng.3977

exome array and associated with lipid levels with $P < 2.1 \times 10^{-7}$ (two-tailed score test; **Supplementary Tables 13** and **14**). For example, an intronic SNP (rs11136341, close to the *PLEC* gene) associated with LDL-C was the original lead SNP in its GWAS locus ($P = 2 \times 10^{-13}$). In the current study, a protein-altering variant in *PARP10* was the top variant in the same locus (rs11136343; p.Leu395Pro; $P = 7 \times 10^{-26}$). After conditioning on PARP10 p.Leu395Pro, the evidence for rs11136341 diminished ($P = 0.02$); in contrast, PARP10 p.Leu395Pro remained significant ($P = 9 \times 10^{-13}$) after conditioning on rs11136341. PARP10 has been shown to affect the hepatic secretion of apolipoprotein B (apoB) in human hepatocytes[17]; these results prioritize *PARP10* as a causal gene at this locus.

Experimental analysis of discovered mutations in model systems is a powerful approach to validate the results of human genetic analyses. We prioritized two coding alterations for experimental analysis: Janus kinase 2 (JAK2) p.Val617Phe and APOBEC1 complementation factor (A1CF) p.Gly398Ser.

JAK2 p.Val617Phe is a recurrent somatic alteration that arises in hematopoietic stem cells and can lead to myeloproliferative disorders or clonal hematopoiesis of indeterminate potential[18–22]. We

Table 1 Protein-altering variants at novel loci associated with lipid levels

Chromosome: position (hg19)	rs ID	Alleles (reference/ alternative)	Gene	Protein change	*n*	Alternative-allele frequency	Trait	*P* value	Beta	s.e.
Total cholesterol										
2:101627925	rs1062062	C/T	*TBC1D8*	p.Gly954Arg	292,898	0.12	TC	1×10^{-7}	–0.021	0.0040
4:69343287	rs976002	A/G	*TMPRSS11E*	p.Tyr303Cys	293,961	0.23	TC	5×10^{-20}	0.029	0.0031
							LDL-C	3×10^{-12}	0.023	0.0033
4:155489608	rs6054	C/T	*FGB*	p.Pro206Leu	307,997	0.0038	TC	5×10^{-12}	0.14	0.021
							TG	3×10^{-11}	0.14	0.021
9:5073770	rs77375493	G/T	*JAK2*	p.Val617Phe	188,412	0.0011	TC	1×10^{-11}	–0.32	0.047
							LDL-C	2×10^{-9}	–0.30	0.049
9:117166246	rs2274159	A/G	*DFNB31*	p.Val400Ala	319,677	0.48	TC	2×10^{-7}	0.013	0.0026
17:8216468	rs871841	T/C	*ARHGEF15*	p.Leu277Pro	298,725	0.52	TC	2×10^{-8}	0.015	0.0026
19:18304700	rs874628	A/G	*MPV17L2*	p.Met72Val	319,677	0.26	TC	2×10^{-7}	0.015	0.0029
LDL cholesterol										
1:155106227	rs4745	A/T	*EFNA1*	p.Asp137Val	291,361	0.49	LDL-C	5×10^{-8}	–0.015	0.0027
4:187120211	rs13146272	C/A	*CYP4V2*	p.Gln259Lys	295,826	0.62	LDL-C	1×10^{-7}	–0.015	0.0027
5:176520243	rs351855	G/A	*FGFR4*	p.Gly388Arg	233,058	0.29	LDL-C	4×10^{-8}	–0.018	0.0033
9:139368953	rs3812594	G/A	*SEC16A*	p.Arg1039Cys	293,723	0.24	LDL-C	2×10^{-8}	–0.018	0.0031
10:118397971	rs10885997	A/G	*PNLIPRP2*	p.Gln387Arg	258,146	0.41	LDL-C	9×10^{-8}	0.015	0.0029
10:124610027	rs1891110	G/A	*FAM24B*	p.Pro2Leu	295,826	0.55	LDL-C	8×10^{-15}	0.021	0.0026
							TC	2×10^{-13}	0.019	0.0025
12:72179446	rs61754230	C/T	*RAB21*	p.Ser224Phe	292,762	0.015	LDL-C	1×10^{-7}	0.057	0.011
14:94844947	rs28929474	C/T	*SERPINA1*	p.Glu366Lys	290,263	0.015	LDL-C	4×10^{-14}	0.081	0.011
							TC	6×10^{-14}	0.078	0.010
17:26694861	rs704	G/A	*VTN*	p.Thr400Met	295,826	0.49	LDL-C	6×10^{-16}	0.021	0.0026
							TC	1×10^{-8}	0.015	0.0025
19:42584958	rs201596848	C/T	*ZNF574*	p.Arg734Cys	273,744	0.0014	LDL-C	5×10^{-12}	–0.255	0.037
Triglycerides										
2:202122995	rs3769823	A/G	*CASP8*	p.Lys14Arg	295,956	0.69	TG	1×10^{-9}	0.017	0.0028
5:131008194	rs26008	T/C	*FNIP1*	p.Gln620Arg	305,699	0.92	TG	5×10^{-9}	–0.028	0.0048
10:52573772	rs41274050	C/T	*A1CF*	p.Gly398Ser	299,984	0.0072	TG	4×10^{-11}	0.10	0.015
							TC	1×10^{-7}	0.08	0.015
13:45970147	rs138358301	A/G	*SLC25A30*	p.Phe280Leu	301,087	0.0035	TG	3×10^{-11}	0.15	0.022
15:40751555	rs3803357	C/A	*BAHD1*	p.Gln298Lys	305,699	0.55	TG	1×10^{-10}	–0.017	0.0026
17:17409560	rs7946	C/T	*PEMT*	p.Val212Met	304,420	0.67	TG	1×10^{-8}	–0.016	0.0029
20:56140439	rs41302559	G/A	*PCK1*	p.Arg483Gln	299,984	0.0021	TG	9×10^{-8}	–0.154	0.029
22:17625915	rs35665085	G/A	*CECR5*	p.Thr149Met	302,582	0.050	TG	5×10^{-8}	0.032	0.0059
HDL cholesterol										
2:272203	rs11553746	C/T	*ACP1*	p.Thr95Ile	313,148	0.33	HDL-C	5×10^{-8}	0.015	0.0027
2:54482553	rs17189743	G/A	*TSPYL6*	p.Arg246Cys	314,415	0.029	HDL-C	2×10^{-7}	0.040	0.0076
2:179309165	rs75862065	G/A	*PRKRA*	p.Pro116Leu	105,490	0.29	HDL-C	2×10^{-7}	0.026	0.0050
3:48229366	rs146179438	C/A	*CDC25A*	p.Gln25His	288,306	0.020	HDL-C	3×10^{-11}	–0.063	0.0095
5:176637576	rs28932178	T/C	*NSD1*	p.Ser457Pro	310,567	0.17	HDL-C	8×10^{-9}	0.020	0.0035
11:64031241	rs35169799	C/T	*PLCB3*	p.Ser778Leu	314,415	0.060	HDL-C	4×10^{-13}	–0.039	0.0054
							TG	3×10^{-12}	0.038	0.0055
11:68703959	rs622082	A/G	*IGHMBP2*	p.Thr671Ala	316,391	0.31	HDL-C	6×10^{-10}	–0.017	0.0028
16:4755108	rs78074706	G/A	*ANKS3*	p.Arg286Trp	315,298	0.022	HDL-C	1×10^{-9}	–0.053	0.0087
16:69385641	rs76116020	A/G	*TMED6*	p.Phe6Leu	310,822	0.033	HDL-C	7×10^{-9}	–0.041	0.0071
17:40257163	rs2074158	T/C	*DHX58*	p.Gln425Arg	244,331	0.19	HDL-C	1×10^{-7}	–0.020	0.0038

have recently shown that carriage of p.Val617Phe increases with age and confers higher risk for CAD[23]. Surprisingly, the allele encoding p.617Phe, which increases risk for CAD, was associated with lower LDL-C. Mice with Jak2 p.Val617Phe knock-in were created, as reported previously[24]. Hypercholesterolemia-prone mice that were engrafted with bone marrow obtained from Jak2 p.Val617Phe transgenic mice displayed lower TC than did mice that received control bone marrow (**Supplementary Fig. 7**). This result was consistent with our human genetic observations. The mechanism by which the JAK2 p.Val617Phe alteration leads to lower plasma TC and LDL-C but higher risk for CAD requires further study.

Another new association that emerged from the genetic analyses was between A1CF p.Gly398Ser and TG (MAF 0.7%, 0.10-s.d. increase in TG per copy of alternative allele, $P = 4 \times 10^{-11}$, two-tailed score test); this variant was also associated with higher circulating TC ($P = 4 \times 10^{-7}$, two-tailed score test) and was nominally associated with a higher risk of CAD (odds ratio (OR), 1.12; $P = 0.02$, two-tailed logistic regression). *A1CF* encodes APOBEC1 complementation factor, an RNA-binding protein that facilitates the RNA-editing action of APOBEC1 on the *APOB* transcript[25,26]. We performed CRISPR–Cas9 deletion, rescue, and knock-in experiments to assess whether A1CF p.Gly398Ser is a causal alteration that alters TG metabolism.

CRISPR–Cas9-induced deletion of *A1CF* led to 72% and 65% less secreted APOB100 than that in control Huh7 and HepG2 human hepatoma cells, respectively (**Fig. 1a–c** and **Supplementary Fig. 8**). These findings were consistent with those from previous studies in rat primary hepatocytes, in which significantly decreased apoB secretion has been observed after RNA-interference-based depletion of *A1CF*[27]. Additionally, cellular APOB100 levels were significantly lower in A1CF-deficient cells (**Supplementary Fig. 8b**,**c**). A subsequent 'rescue' experiment involving overexpression of wild-type or A1CF p.Gly398Ser in Huh7 cells with or without endogenous A1CF expression confirmed higher APOB100 secretion in cell lines expressing A1CF p.Gly398Ser (**Fig. 1d**).

We sought to further validate the *A1CF* gene and the p.Gly398Ser variant by using CRISPR–Cas9 to generate knock-in mice. Using a guide RNA targeting *A1cf* exon 9, the site of the codon encoding p.Gly398, and a 162-nt single-stranded DNA oligonucleotide-repair template containing the p.Gly398Ser variant as well as additional synonymous changes to prevent recleavage by CRISPR–Cas9, we generated mice on the C57BL/6J inbred background with an allele encoding A1cf p.Gly398Ser (hereafter referred to as KI) (**Supplementary Fig. 9a**,**b**). We bred mice to achieve KI-allele homozygosity and found that KI/KI mice were viable and healthy. We compared wild-type and KI/KI colony mates ($n = 9$ and 8, respectively) with respect to TG levels (**Supplementary Fig. 9c**,**d**). We found that KI/KI mice had 46% higher TG than did the wild-type mice ($P = 0.05$). In sum, these results indicated that *A1CF* is a causal gene for TG in humans and that the p.Gly398Ser variant is a causal alteration with possible relevance to CAD.

Next, we used the 444 identified DNA sequence variants to address four clinical questions. First, a rare null mutation in *HBB*, encoding beta-globin c.92+1G>A, rs33971440, associated with lower TC (**Supplementary Table 15**) and exhibited the strongest TC-lowering effect after null mutations in *PCSK9*; this result prompted us to question the relationship between beta-thalassemia and risk for CAD. Approximately 80 to 90 million individuals worldwide have been estimated to carry a heterozygous loss-of-function *HBB* mutation, termed the beta-thalassemia trait[28]. Observational epidemiologic studies have shown that the beta-thalassemia trait is associated with lower blood cholesterol levels[29,30]. We found that *HBB* c.92+1G>A was associated with a 17 mg/dl decrease in LDL-C (95% confidence interval (CI), [−23, −11]; $P = 2.7 \times 10^{-8}$ two-tailed score test) and a 21 mg/dl decrease in TC (95% CI, [−27, −14]; $P = 8.9 \times 10^{-11}$) (**Supplementary Fig. 10**). In an analysis of 31,156 CAD cases and 65,787 controls, carriers of loss-of-function variants in *HBB* were protected against CAD (OR for CAD, 0.70; 95% CI, [0.54, 0.90]; $P = 0.005$, two-tailed logistic regression) (**Supplementary Fig. 11**). We display results for all null mutations with association $P < 0.001$, two-sided test for correlation for any of the four lipid traits (**Supplementary Table 15**).

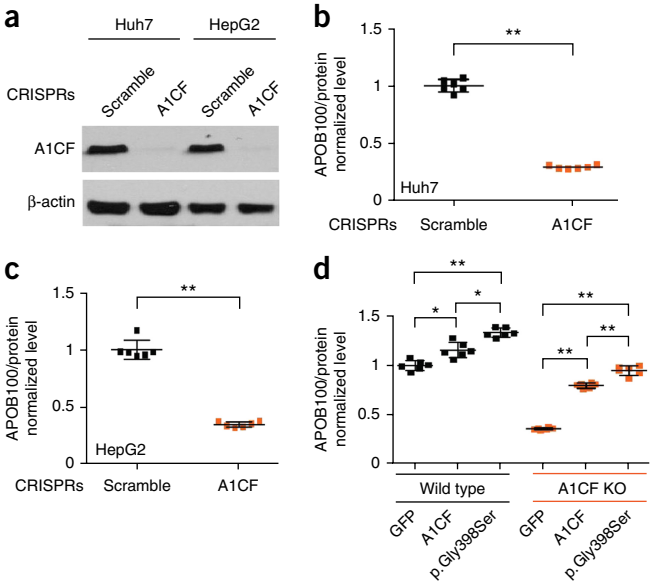


Figure 1 A1CF p.Gly398Ser mutant leads to increased APOB100 secretion. (**a**) Western blot showing the depletion of endogenous A1CF levels via a CRISPR–Cas9 system in both Huh7 and HepG2 cells. Scramble, scrambled control sequence; β-actin, loading control. (**b**,**c**) Lack of A1CF leads to low APOB100 secretion in Huh7 (**b**) and HepG2 (**c**) human hepatoma cells. (**d**) Recombinantly overexpressed A1CF p.Gly398Ser variant, compared with A1CF or GFP control, results in significantly higher APOB100 secretion in both Huh7 wild type and A1CF-knockout (A1CF KO) cells, respectively. Data are shown as mean ± s.d. in **b**–**d** from experiments with $n = 6$ replicates per genotype per experiment, representative of 3 independent experiments (*$P < 0.05$; **$P < 0.01$ by two-tailed Student's t test).

Second, DNA sequence variants in the *CETP* gene that are associated with higher HDL-C also correlate with higher risk for AMD, a leading cause of blindness[31–35]; here, we asked whether any way of increasing plasma HDL-C might predictably lead to increased AMD risk. Across 168 independent HDL-C variants with MAF >1%, we tested the association of each HDL-C variant with AMD risk. The effect size of a variant on HDL-C was positively correlated with its effect on AMD risk (correlation in effect sizes, $r = 0.41$, $P = 4.4 \times 10^{-8}$; **Supplementary Table 16** and **Supplementary Fig. 12**). However, this effect was driven by the ten independent HDL-C-associated variants in *CETP* (heterogeneity across the different HDL-C-raising mechanisms (between-study variance (τ^2) = 0.91, P value for heterogeneity (P_{het}) = 1.8×10^{-15}, two-tailed test of heterogeneity statistic Q)) (**Supplementary Table 17**). When these ten *CETP* variants were removed, there was no longer a relationship between genetically altered HDL-C and AMD risk ($P = 0.17$ two-sided correlation of effect sizes). These results suggested that outside of the *CETP* locus, there is not a predictable relationship between plasma HDL-C and risk for AMD.

Third, we asked whether lowering LDL-C with lipid-modifying medicines would always increase risk for T2D. This question was motivated by the observations that, in randomized controlled trials, statin therapy has been found to increase risk for T2D and that, in

Table 2 Effects of genes involved in hepatic production of triglyceride-rich lipoproteins (*PNPLA3* and *TM6SF2*) versus lipolytic-pathway genes (*LPL* and *ANGPTL4*) on related metabolic traits: blood lipids, fatty liver, type 2 diabetes, and coronary artery disease

Gene	*LPL*	*ANGPTL4*	*PNPLA3*	*TM6SF2*
Variant	p.Ser474Ter	p.Glu40Lys	p.Ile148Met	p.Glu167Lys
Effect allele Frequency	Ter 10%	Lys 2%	Met 23%	Lys 7%
		Blood triglycerides		
Effect direction	↓	↓	↓	↓
Beta	–0.18	–0.27	–0.018	–0.12
CI	[–0.19, –0.17]	[–0.29, –0.25]	[–0.024, –0.012]	[–0.13, –0.11]
P	$P < 1 \times 10^{-323}$	$P = 4 \times 10^{-175}$	$P = 4 \times 10^{-9}$	$P = 4 \times 10^{-125}$
		Blood LDL cholesterol		
Effect direction	–	↓	↓	↓
Beta	0.013	–0.004	–0.018	–0.103
CI	[0.0052, 0.021]	[–0.024, 0.016]	[–0.024, –0.012]	[–0.11, –0.093]
P	$P = 0.005$	$P = 0.70$	$P = 1 \times 10^{-8}$	$P = 7 \times 10^{-93}$
		Fatty liver		
Effect direction	–	–	↑	↑
Beta[a]	0.026	0.112	–0.25	–0.25
CI	[–0.035, 0.087]	[–0.021, 0.25]	[–0.29, –0.2]	[–0.32, –0.18]
P	$P = 0.41$	$P = 0.10$	$P = 2 \times 10^{-30}$	$P = 5 \times 10^{-12}$
		Type 2 diabetes		
Effect direction	↓	↓	↑	↑
OR	0.95	0.91	1.04	1.07
CI	[0.93, 0.97]	[0.83, 0.99]	[1.03, 1.05]	[1.05, 1.09]
P	$P = 7 \times 10^{-9}$	$P = 1 \times 10^{-4}$	$P = 2 \times 10^{-10}$	$P = 5 \times 10^{-12}$
		Coronary artery disease		
Effect direction	↓	↓	↓	↓
OR	0.93	0.85	0.96	0.95
CI	[0.9, 0.96]	[0.8, 0.9]	[0.94, 0.97]	[0.93, 0.98]
P	$P = 4 \times 10^{-7}$	$P = 2 \times 10^{-10}$	$P = 4 \times 10^{-8}$	$P = 3 \times 10^{-4}$

[a]A negative beta reflects liver attenuation in computed tomography, a result indicative of higher liver fat. Association results for lipids are derived from the present study. Association results for T2D are from ref. 39. Association results for coronary artery disease are from ref. 40. *P* values were determined by two-sided correlation of effect sizes.

recent reports, *PCSK9* variants have been associated with higher risk for T2D[36–38]. We confirmed the association of PCSK9 p.Arg46Leu with T2D risk among 222,877 participants (**Supplementary Table 18**). We found that the p.Arg46Leu allele associated with lower LDL-C conferred a 13% higher risk for T2D (OR, 1.13; 95% CI, [1.06–1.20]; 7.0×10^{-5}, Cochran–Mantel–Haenszel statistics for stratified 2 × 2 tables) (**Supplementary Fig. 13**). In addition, across 113 independent LDL-C variants at 90 distinct loci, we compared each variant's effect on LDL-C with its subsequent effect on risk for T2D. Across the 113 variants, there was a weak inverse correlation between a variant's effects on LDL-C and T2D ($r = -0.21$, $P = 0.025$); however, there was evidence for heterogeneity in this relationship ($\tau_2 = 0.50$, $P_{het} = 2.5 \times 10^{-9}$, two-tailed test of heterogeneity statistic *Q*). Five LDL-C-lowering genetic mechanisms had the most compelling evidence for association with higher risk for T2D (TM6SF2 p.Glu167Lys, *APOE* rs769449, HNF4A p.Thr136Ile, PNPLA3 p.Ile148Met, and GCKR p.Leu446Pro) ($P < 4.0 \times 10^{-4}$, two-sided score test for each, Bonferroni-correction threshold for performing tests at 113 variants, **Supplementary Table 19** and **Supplementary Fig. 14**). These results suggested that only some ways of lowering LDL-C are likely to increase risk for T2D.

Finally, two key processes—hepatic production and peripheral lipolysis—contribute to blood TG levels. We asked how genes involved in hepatic production of TG-rich lipoproteins (*PNPLA3* and *TM6SF2*) differed from lipolysis-pathway genes (*LPL* and *ANGPTL4*) in their effects on the related metabolic traits of blood lipids, fatty liver, T2D, and CAD (**Table 2**). The alternative alleles at PNPLA3 p.Ile148Met, TM6SF2 p.Glu167Lys, LPL p.Ser474Ter, and ANGTPL4 p.Glu40Lys were all associated with lower blood TG and less risk for CAD. However, the blood-TG-lowering alleles at *PNPLA3* and *TM6SF2* led to more fatty liver and higher risk for T2D. In contrast, the blood-TG-lowering alleles at *LPL* and *ANGPTL4* were neutral with respect to fatty liver and led to lower risk for T2D. We confirmed the *LPL* observation by using a phenome-wide association study in the UK Biobank (**Supplementary Table 20**). In UK Biobank, a 1-s.d. decrease in TG mediated by *LPL* variants decreased risks for both T2D and CAD (**Fig. 2**).

In summary, by combining large-scale human genetic analysis with experimental evidence, we demonstrated the association of 444 independent coding and noncoding variants at 250 loci with plasma lipids and the use of mouse models and genome editing to pinpoint causal genes and protein-altering variants. We further provided evidence that *LPL* activation can be expected to lower TG and reduce risks for both CAD and T2D without increasing liver fat and thus may be advantageous for patients with metabolic risk factors.

URLs. Michigan Genomics Initiative, http://www.michigangenomics.org/.

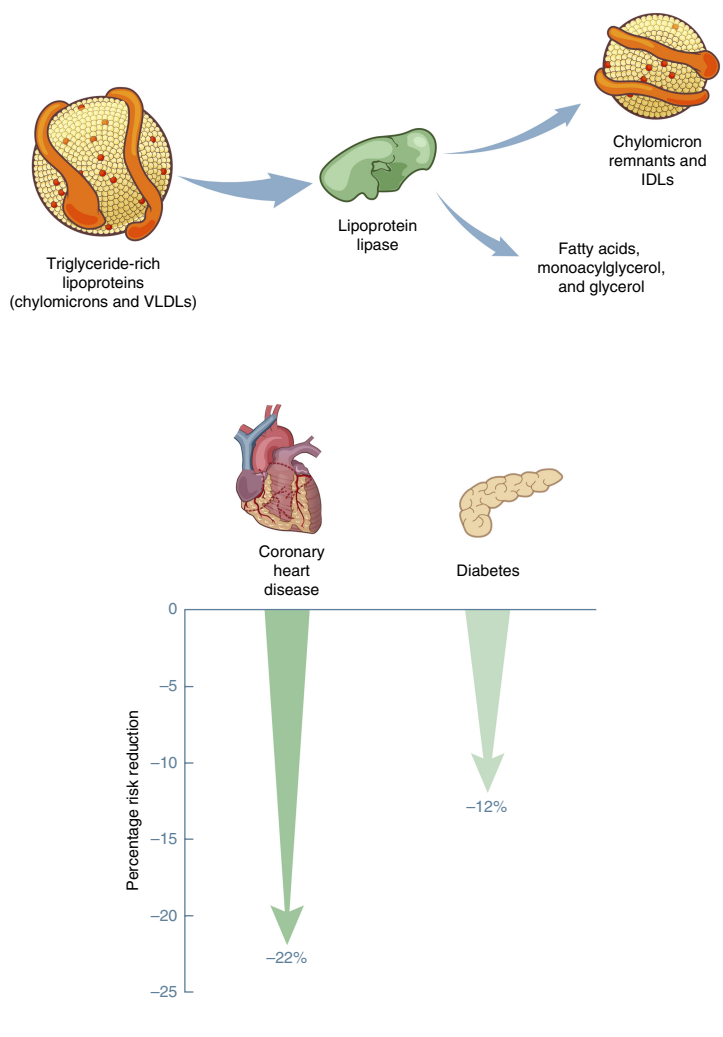
METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

D.J.L. is partially supported by R01HG008983 from the National Human Genome Research Institute of the National Institute of Health, and R21DA040177 and R01DA037904 from the National Institute of Drug Abuse of the National Institute



Outcome	Cases	Controls	OR per s.d. lower triglycerides		*P* value
Cardiovascular					
Coronary heart disease	4,461	107,877	0.78	(0.69, 0.88)	0.0001
Stroke	2,066	110,272	1.06	(0.90, 1.25)	0.5072
Heart failure	599	111,739	0.83	(0.60, 1.13)	0.2397
Atrial fibrillation	2,208	110,130	0.93	(0.79, 1.09)	0.3521
Aortic stenosis	193	112,145	0.63	(0.36, 1.13)	0.1203
Peripheral vascular disease	692	111,646	0.85	(0.64, 1.14)	0.2818
Venous thromboembolism	3,306	109,032	1.00	(0.87, 1.14)	0.9711
Gastrointestinal disease					
Inflammatory bowel disease	1,021	111,317	1.17	(0.93, 1.48)	0.1702
Gastric reflux	4,881	107,457	1.03	(0.92, 1.14)	0.6414
Gallstones	1,831	110,507	1.13	(0.95, 1.34)	0.1682
Irritable bowel syndrome	2,679	109,659	0.93	(0.80, 1.08)	0.3554
Endocrine					
Diabetes	5,741	112,338	0.88	(0.79, 0.97)	0.0139
Gestational diabetes	56	112,282	0.73	(0.26, 2.08)	0.5594
Hyperthyroidism	868	111,470	1.13	(0.88, 1.45)	0.3331
Hypothyroidism	5,433	106,905	1.05	(0.95, 1.17)	0.348
Gout	1,612	110,726	1.00	(0.83, 1.21)	0.9826
Renal/urological					
Enlarged prostate	1,573	110,765	1.08	(0.89, 1.30)	0.4548
Uterine fibroids	1,634	110,704	1.00	(0.83, 1.20)	0.9901
Neurological/psychiatric					
Migraine	3,161	109,177	1.02	(0.89, 1.17)	0.7436
Depression	6,667	105,671	1.06	(0.97, 1.17)	0.1985
Anxiety	1,545	110,793	1.02	(0.84, 1.23)	0.8409
Muscoskeletal					
Back pain	584	111,754	1.03	(0.76, 1.40)	0.8501
Joint pain	377	111,961	0.99	(0.68, 1.46)	0.9737
Osteoporosis	1,740	110,598	1.02	(0.85, 1.22)	0.8528
Osteoarthritis	9,693	102,645	0.96	(0.89, 1.05)	0.3779
Sciatica	1,035	111,303	1.00	(0.79, 1.27)	0.9776
Prolapsed disc	1,856	110,482	1.04	(0.87, 1.24)	0.6542
Respiratory					
Asthma	13,941	98,397	0.99	(0.93, 1.06)	0.8411
COPD/emphysema	2,363	109,975	0.93	(0.80, 1.09)	0.398
Pneumonia	1,581	110,757	1.05	(0.87, 1.26)	0.6297
Hayfever	6,263	106,075	1.06	(0.96, 1.17)	0.2411
Cancer					
Lung cancer	115	112,223	0.92	(0.45, 1.86)	0.8124
Breast cancer	2,383	109,955	0.90	(0.77, 1.06)	0.203
Colorectal cancer	616	111,722	1.16	(0.86, 1.56)	0.3222
Skin cancer	2,482	109,856	0.96	(0.82, 1.12)	0.5883
Prostate cancer	840	111,498	0.99	(0.76, 1.28)	0.9238
Cervical cancer	872	111,466	1.09	(0.85, 1.40)	0.491
Other cancer	2,409	109,929	0.88	(0.75, 1.03)	0.1007
Any cancer	9,530	102,808	0.94	(0.87, 1.02)	0.1692

Figure 2 Association of genetically lowered triglycerides by *LPL* variants with a range of phenotypes. Estimates were derived in UK Biobank by using logistic regression, adjusting for age, sex, ten principal components of ancestry, and an indicator variable for array type. Effect estimates are for 1-s.d.-lower plasma triglycerides. Definitions for all outcomes are provided in **Supplementary Table 20**. VLDLs, very low-density lipoproteins; IDLs, intermediate-density lipoproteins.

of Health. G.M.P. is supported by the National Heart, Lung, and Blood Institute of the National Institutes of Health award K01HL125751. A.P.P. is supported by a research fellowship from the Stanley J. Sarnoff Cardiovascular Research Foundation. H. Tada is supported by a grant from the Japanese Circulation Society to study in the United States. The research was supported by the National Institute for Health Research (NIHR) Exeter Clinical Research Facility and ERC grant 323195; SZ-245 50371-GLUCOSEGENES-FP7-IDEAS-ERC to T.M.F. E.K.S. is supported by NIH grants R01 DK106621 and R01 DK107904, the University of Michigan Biological Sciences Scholars Program, and the University of Michigan Department of Internal Medicine. T.D.S. is supported by an ERC Advanced Principal Investigator award. A.P.M. is supported as a Wellcome Trust Senior Fellow in Basic Biomedical Science (grant no. WT098017). Y.E.C. is supported by HL117491 and HL129778 from the NIH. S.K.G. is supported by HL122684 from the NIH. P.L.A. is supported by NHLBI R21 HL121422-02 from the NIH. C.L., N.J.W., and R.A.S. acknowledge funding from the Medical Research Council, UK (MC_UU_12015/1). J.D. is supported as a British Heart Foundation Professor, European Research Council Senior Investigator, and National Institute for Health Research (NIHR) Senior Investigator. C.J.W. is supported by HL094535 and HL109946 from the NIH. S. Kathiresan is supported by a research scholar award from the Massachusetts General Hospital, the Donovan Family Foundation, and R01 HL127564 and R33 HL120781 from the NIH.

The views expressed in this manuscript are those of the authors and do not necessarily represent the views of the National Heart, Lung, and Blood Institute; the National Institutes of Health or the US Department of Health and Human Services.

This research has been conducted using the UK Biobank resource, application 7089. Funding support for participating studies in the meta-analysis can be found in the **Supplementary Note**.

AUTHOR CONTRIBUTIONS

All authors contributed to and approved the results and comments on the manuscript. Writing: C.J.W., D.J.L., G.M.P., G.A., P.D., X.L., and S. Kathiresan. Study supervision: S. Kathiresan. Primary analysis: D.J.L. and G.M.P. Secondary analysis: A.V.K., A. Mahajan, Charge Diabetes Working Group, C.M.M., C.E., D.J.R., D.F.R., D.P., E.K.S., E.M.S., GOLD Consortium, J.B.M., J. Wessel, L.G.F., M.O.G., M.I.M., M. Boehnke, N. Stitziel, R.S.S., S. Somayajula, VA Million Veteran Program, and X.L. Functional characterization: A.R.T., C.A.C., H. Yu, K.M., N.W., and X.W. Contribution to study-specific analysis: A.S.B., A.C.A., A.C.M., A.D., A.-E.F., A.K.M., A. Langsted, A. Linneberg, A. Malarstig, A. Manichaikul, A. Maschio, A. Metspalu, A. Mulas, A.P., A.P.M., A.P.P., A.P.R., A.R., A.T.-H., A.U.J., A.V., A.V.S., A.Y.C., B.G.N., B.H.S., B.M.P., C.C., C.G., C.H., C.J.O'D., C.J.W., C.L., C.K., C.M.B., C.M.S., C.N.A.P., C.P., D. Alam, D. Arveiler, D.C.M.L., D.I.C., D.J.L., D.K., D.M.R., D.S., E.B., E.d.A., E.M., E.P.B., EPIC-CVD Consortium, The EPIC-InterAct Consortium, E.Z., F.B., F.C., F.G., F. Karpe, F. Kee, F.R., G.B.J., G. Davies, G. Dedoussis, G.E., G.M.P., G.P., H.A.K., H.G., H.M.S., H.R.W., H. Tada, H. Tang, H. Yaghootkar, H.Z., I.B., I.F., I.J.D., I.R., J.C.B., J.B.-J., J.C.C., J.C.D., J.D., J.D.R., J.F., J.G.W., J.H., J.I.R., J.J., J.K., J.M.C., J.M.M.H., J.M.J., J.M.O., J.M.S., J.B.N., J.N.H., J.S.K., J.-C.T., J.T., J.V., J. Weinstock, J.W.J., K.D.T., K.E.S., K.H., K.K., K.S., K.S.S., L.A.C., L.A.L., L.E.B., L.G., L.J.L., L.S., M. Benn, M. Brown, M.J.C., M.-P.D., M.E.G., M.E.J., M. Ferrario, M.F.F.,

M. Fornage, M.-R.J., M.J.N., M.L., M.L.G., M.M.-N., M.O.-M., M.P., M.W., M.X., M.Z., N.G., N.G.D.M., N.J.S., N.J.W., N.P., N.R.R., N.R.v.Z., N. Sattar, N.S.Z., O.L.H., O.M., O. Pedersen, O. Polasek, P.A., P.B.M., P.D., P.E.W., P.F., P.L.A., P. Mäntyselkä, P.M.R., P. Muntendam, P.R.K., P. Sever, P.S.T., P. Surendran, P.W.F., P.W.F.W., R.A.S., R.C., R.F.-S., R.J.F.L., R. Magi, R. Mehran, R.R., R.Y., S.P., S.F.N., S.J., S. Kanoni, S. Kathiresan, S.K.G., S.M.D., S. Sanna, S. Sivapalaratnam, S.S.R., S.T., T.B.H., T.D.S., T. Ebeling, T.E.-I.C., T. Esko, T.H., T.L.A., T. Lakka, T. Lauritzen, T.M.F., T.V.V., U.B., V.F., V.G., V.S., W.G., W. Zhang, W. Zhou, X.S., Y.E.C., Y.H., Y.-D.I.C., Y.L., Y. Zhang, and Y. Zhou.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details are available in the online version of the paper.

1. Teslovich, T.M. *et al.* Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**, 707–713 (2010).
2. Asselbergs, F.W. *et al.* Large-scale gene-centric meta-analysis across 32 studies identifies multiple lipid loci. *Am. J. Hum. Genet.* **91**, 823–838 (2012).
3. Willer, C.J. *et al.* Discovery and refinement of loci associated with lipid levels. *Nat. Genet.* **45**, 1274–1283 (2013).
4. Albrechtsen, A. *et al.* Exome sequencing-driven discovery of coding polymorphisms associated with common metabolic phenotypes. *Diabetologia* **56**, 298–310 (2013).
5. Peloso, G.M. *et al.* Association of low-frequency and rare coding-sequence variants with blood lipids and coronary heart disease in 56,000 whites and blacks. *Am. J. Hum. Genet.* **94**, 223–232 (2014).
6. Surakka, I. *et al.* The impact of low-frequency and rare variants on lipid levels. *Nat. Genet.* **47**, 589–597 (2015).
7. Tang, C.S. *et al.* Exome-wide association analysis reveals novel coding sequence variants associated with lipid traits in Chinese. *Nat. Commun.* **6**, 10206 (2015).
8. Musunuru, K. *et al.* From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. *Nature* **466**, 714–719 (2010).
9. Burkhardt, R. *et al.* Trib1 is a lipid- and myocardial infarction-associated gene that regulates hepatic lipogenesis and VLDL production in mice. *J. Clin. Invest.* **120**, 4410–4414 (2010).
10. Voight, B.F. *et al.* Plasma HDL cholesterol and risk of myocardial infarction: a mendelian randomisation study. *Lancet* **380**, 572–580 (2012).
11. Do, R. *et al.* Common variants associated with plasma triglycerides and risk for coronary artery disease. *Nat. Genet.* **45**, 1345–1352 (2013).
12. Lu, X. *et al.* Exome chip meta-analysis identifies novel loci and East Asian–specific coding variants contributing to lipid levels and coronary artery disease. *Nat. Genet.* http://dx.doi.org/10.1038/ng.3978 (2017).
13. Feng, S., Liu, D., Zhan, X., Wing, M.K. & Abecasis, G.R. RAREMETAL: fast and powerful meta-analysis for rare variants. *Bioinformatics* **30**, 2828–2829 (2014).
14. Liu, D.J. *et al.* Meta-analysis of gene-level tests for rare variant association. *Nat. Genet.* **46**, 200–204 (2014).
15. Locke, A.E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197–206 (2015).
16. Holmen, O.L. *et al.* Systematic evaluation of coding variation identifies a candidate causal variant in TM6SF2 influencing total cholesterol and myocardial infarction risk. *Nat. Genet.* **46**, 345–351 (2014).
17. Shen, X. *et al.* Identification of genes affecting apolipoprotein B secretion following siRNA-mediated gene knockdown in primary human hepatocytes. *Atherosclerosis* **222**, 154–157 (2012).
18. Baxter, E.J. *et al.* Acquired mutation of the tyrosine kinase JAK2 in human myeloproliferative disorders. *Lancet* **365**, 1054–1061 (2005).
19. James, C. *et al.* A unique clonal JAK2 mutation leading to constitutive signalling causes polycythaemia vera. *Nature* **434**, 1144–1148 (2005).
20. Kralovics, R. *et al.* A gain-of-function mutation of JAK2 in myeloproliferative disorders. *N. Engl. J. Med.* **352**, 1779–1790 (2005).
21. Levine, R.L. *et al.* Activating mutation in the tyrosine kinase JAK2 in polycythemia vera, essential thrombocythemia, and myeloid metaplasia with myelofibrosis. *Cancer Cell* **7**, 387–397 (2005).
22. Jaiswal, S. *et al.* Age-related clonal hematopoiesis associated with adverse outcomes. *N. Engl. J. Med.* **371**, 2488–2498 (2014).
23. Jaiswal, S. *et al.* Clonal hematopoiesis and risk of atherosclerotic cardiovascular disease. *N. Engl. J. Med.* **377**, 111–121 (2017).
24. Mullally, A. *et al.* Physiological Jak2V617F expression causes a lethal myeloproliferative neoplasm with differential effects on hematopoietic stem and progenitor cells. *Cancer Cell* **17**, 584–596 (2010).
25. Lellek, H. *et al.* Purification and molecular cloning of a novel essential component of the apolipoprotein B mRNA editing enzyme-complex. *J. Biol. Chem.* **275**, 19848–19856 (2000).
26. Mehta, A., Kinter, M.T., Sherman, N.E. & Driscoll, D.M. Molecular cloning of apobec-1 complementation factor, a novel RNA-binding protein involved in the editing of apolipoprotein B mRNA. *Mol. Cell. Biol.* **20**, 1846–1854 (2000).
27. Galloway, C.A., Ashton, J., Sparks, J.D., Mooney, R.A. & Smith, H.C. Metabolic regulation of APOBEC-1 complementation factor trafficking in mouse models of obesity and its positive correlation with the expression of ApoB protein in hepatocytes. *Biochim. Biophys. Acta* **1802**, 976–985 (2010).
28. Galanello, R. & Origa, R. Beta-thalassemia. *Orphanet J. Rare Dis.* **5**, 11 (2010).
29. Fessas, P., Stamatoyannopoulos, G. & Keys, A. Serum-cholesterol and thalassemia trait. *Lancet* **1**, 1182–1183 (1963).
30. Sidore, C. *et al.* Genome sequencing elucidates Sardinian genetic architecture and augments association analyses for lipid and blood inflammatory markers. *Nat. Genet.* **47**, 1272–1281 (2015).
31. Chen, W. *et al.* Genetic variants near TIMP3 and high-density lipoprotein-associated loci influence susceptibility to age-related macular degeneration. *Proc. Natl. Acad. Sci. USA* **107**, 7401–7406 (2010).
32. Neale, B.M. *et al.* Genome-wide association study of advanced age-related macular degeneration identifies a role of the hepatic lipase gene (LIPC). *Proc. Natl. Acad. Sci. USA* **107**, 7395–7400 (2010).
33. Wang, Y.F. *et al.* CETP/LPL/LIPC gene polymorphisms and susceptibility to age-related macular degeneration. *Sci. Rep.* **5**, 15711 (2015).
34. Cheng, C.Y. *et al.* New loci and coding variants confer risk for age-related macular degeneration in East Asians. *Nat. Commun.* **6**, 6063 (2015).
35. Momozawa, Y. *et al.* Low-frequency coding variants in CETP and CFB are associated with susceptibility of exudative age-related macular degeneration in the Japanese population. *Hum. Mol. Genet.* **25**, 5027–5034 (2016).
36. Lotta, L.A. *et al.* Association between low-density lipoprotein cholesterol-lowering genetic variants and risk of type 2 diabetes: A meta-analysis. *J. Am. Med. Assoc.* **316**, 1383–1391 (2016).
37. Schmidt, A.F. *et al.* PCSK9 genetic variants and risk of type 2 diabetes: a mendelian randomisation study. *Lancet Diabetes Endocrinol.* **5**, 97–105 (2017).
38. Ference, B.A. *et al.* Variation in PCSK9 and HMGCR and risk of cardiovascular disease and diabetes. *N. Engl. J. Med.* **375**, 2144–2153 (2016).
39. Mahajan, A. *et al.* Refining the accuracy of validated target identification through coding variant fine-mapping in type 2 diabetes. Preprint at https://www.biorxiv.org/content/early/2017/05/31/144410.1/ (2017).
40. Myocardial Infarction Genetics & CARDIoGRAM Exome Consortia Investigators. *et al.* Coding variation in ANGPTL4, LPL, and SVEP1 and the risk of coronary disease. *N. Engl. J. Med.* **374**, 1134–1144 (2016).

Dajiang J Liu[1,181], Gina M Peloso[2,3,181], Haojie Yu[4,181], Adam S Butterworth[5,6,181], Xiao Wang[7,181], Anubha Mahajan[8,181], Danish Saleheen[5,9,10,181], Connor Emdin[3,11,181], Dewan Alam[12], Alexessander Couto Alves[13], Philippe Amouyel[14], Emanuele Di Angelantonio[5,6], Dominique Arveiler[15], Themistocles L Assimes[16,17], Paul L Auer[18], Usman Baber[19], Christie M Ballantyne[20], Lia E Bang[21], Marianne Benn[22,23], Joshua C Bis[24], Michael Boehnke[25], Eric Boerwinkle[26,27], Jette Bork-Jensen[28], Erwin P Bottinger[29], Ivan Brandslund[30,31], Morris Brown[32], Fabio Busonero[33], Mark J Caulfield[32,34,35], John C Chambers[36–38], Daniel I Chasman[39,40], Y Eugene Chen[41], Yii-Der Ida Chen[42], Rajiv Chowdhury[5], Cramer Christensen[43], Audrey Y Chu[39,44], John M Connell[45], Francesco Cucca[33,46], L Adrienne Cupples[2,44], Scott M Damrauer[47,48], Gail Davies[49,50], Ian J Deary[49,50], George Dedoussis[51], Joshua C Denny[52,53], Anna Dominiczak[54], Marie-Pierre Dubé[55–57], Tapani Ebeling[58], Gudny Eiriksdottir[59], Tõnu Esko[3,60], Aliki-Eleni Farmaki[51], Mary F Feitosa[61], Marco Ferrario[62], Jean Ferrieres[63], Ian Ford[64], Myriam Fornage[65], Paul W Franks[66–68], Timothy M Frayling[69], Ruth Frikke-Schmidt[22,23], Lars G Fritsche[25], Philippe Frossard[10], Valentin Fuster[19,70], Santhi K Ganesh[41,71],

Wei Gao[72], Melissa E Garcia[73], Christian Gieger[74–76], Franco Giulianini[39], Mark O Goodarzi[77,78], Harald Grallert[74–76], Niels Grarup[28], Leif Groop[79], Megan L Grove[26], Vilmundur Gudnason[59,80], Torben Hansen[28,81], Tamara B Harris[82], Caroline Hayward[83], Joel N Hirschhorn[3,84], Oddgeir L Holmen[85,86], Jennifer Huffman[83], Yong Huo[87], Kristian Hveem[88], Sehrish Jabeen[10], Anne U Jackson[25], Johanna Jakobsdottir[59,89], Marjo-Riitta Jarvelin[13], Gorm B Jensen[90], Marit E Jørgensen[91,92], J Wouter Jukema[93,94], Johanne M Justesen[28], Pia R Kamstrup[95], Stavroula Kanoni[96], Fredrik Karpe[97,98], Frank Kee[99], Amit V Khera[3,11], Derek Klarin[3,11,100], Heikki A Koistinen[101–103], Jaspal S Kooner[37,38,104], Charles Kooperberg[105], Kari Kuulasmaa[101], Johanna Kuusisto[106], Markku Laakso[106], Timo Lakka[107–109], Claudia Langenberg[110], Anne Langsted[95,111], Lenore J Launer[82], Torsten Lauritzen[112], David C M Liewald[49,50], Li An Lin[65], Allan Linneberg[113–115], Ruth J F Loos[29,116], Yingchang Lu[29], Xiangfeng Lu[41,117], Reedik Mägi[60], Anders Malarstig[118,119], Ani Manichaikul[120], Alisa K Manning[3,11,121], Pekka Mäntyselkä[122], Eirini Marouli[96], Nicholas G D Masca[123,124], Andrea Maschio[33], James B Meigs[3,121,125], Olle Melander[126], Andres Metspalu[60], Andrew P Morris[8,127], Alanna C Morrison[26], Antonella Mulas[33], Martina Müller-Nurasyid[76,128,129], Patricia B Munroe[32,35], Matt J Neville[97], Jonas B Nielsen[41], Sune F Nielsen[95,111], Børge G Nordestgaard[95,111], Jose M Ordovas[130–132], Roxana Mehran[19], Christoper J O'Donnell[100,133], Marju Orho-Melander[126], Cliona M Molony[134], Pieter Muntendam[135], Sandosh Padmanabhan[54], Colin N A Palmer[45], Dorota Pasko[69], Aniruddh P Patel[3,11,133,136], Oluf Pedersen[28], Markus Perola[101,137], Annette Peters[74,76,129], Charlotta Pisinger[115], Giorgio Pistis[33], Ozren Polasek[138,139], Neil Poulter[140], Bruce M Psaty[24,141,142], Daniel J Rader[143], Asif Rasheed[10], Rainer Rauramaa[108,109], Dermot F Reilly[134], Alex P Reiner[105,144], Frida Renström[66,145], Stephen S Rich[120], Paul M Ridker[39], John D Rioux[55], Neil R Robertson[8,97], Dan M Roden[52,53], Jerome I Rotter[42], Igor Rudan[139], Veikko Salomaa[101], Nilesh J Samani[123,124], Serena Sanna[33], Naveed Sattar[54,97], Ellen M Schmidt[146], Robert A Scott[110], Peter Sever[140], Raquel S Sevilla[147], Christian M Shaffer[53], Xueling Sim[25,148], Suthesh Sivapalaratnam[149], Kerrin S Small[150], Albert V Smith[59,80], Blair H Smith[151,152], Sangeetha Somayajula[153], Lorraine Southam[8,154], Timothy D Spector[150], Elizabeth K Speliotes[146,155], John M Starr[49,156], Kathleen E Stirrups[96,157], Nathan Stitziel[158,159], Konstantin Strauch[160,161], Heather M Stringham[25], Praveen Surendran[5], Hayato Tada[162], Alan R Tall[163], Hua Tang[164], Jean-Claude Tardif[55,57], Kent D Taylor[42], Stella Trompet[93,165], Philip S Tsao[16,17], Jaakko Tuomilehto[166–169], Anne Tybjaerg-Hansen[22,23], Natalie R van Zuydam[8,45], Anette Varbo[95,111], Tibor V Varga[66], Jarmo Virtamo[101], Melanie Waldenberger[75,76,129], Nan Wang[163], Nick J Wareham[110], Helen R Warren[32,35], Peter E Weeke[53,170], Joshua Weinstock[25], Jennifer Wessel[171,172], James G Wilson[173], Peter W F Wilson[174,175], Ming Xu[176], Hanieh Yaghootkar[69], Robin Young[5], Eleftheria Zeggini[154], He Zhang[41], Neil S Zheng[177], Weihua Zhang[36], Yan Zhang[87], Wei Zhou[146], Yanhua Zhou[2], Magdalena Zoledziewska[33], Charge Diabetes Working Group[178], The EPIC-InterAct Consortium[178], EPIC-CVD Consortium[178], GOLD Consortium[178], VA Million Veteran Program[178], Joanna M M Howson[5,182], John Danesh[5,6,154,182], Mark I McCarthy[8,97,98,182], Chad A Cowan[4,179,182], Goncalo Abecasis[25,182], Panos Deloukas[96,180,182], Kiran Musunuru[7,182], Cristen J Willer[41,71,146,181,182] & Sekar Kathiresan[3,11,133,136,181,182]

[1]Department of Public Health Sciences, Institute of Personalized Medicine, Penn State College of Medicine, Hershey, Pennsylvania, USA. [2]Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts, USA. [3]Program in Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts, USA. [4]Department of Stem Cell and Regenerative Biology, Harvard Stem Cell Institute, Harvard University, Cambridge, Massachusetts, USA. [5]MRC/BHF Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. [6]The National Institute for Health Research Blood and Transplant Research Unit (NIHR BTRU) in Donor Health and Genomics at the University of Cambridge, Cambridge, UK. [7]Cardiovascular Institute, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA. [8]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK. [9]Department of Biostatistics and Epidemiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA. [10]Center for Non-Communicable Diseases, Karachi, Pakistan. [11]Center for Genomic Medicine, Massachusetts General Hospital, Boston, Massachusetts, USA. [12]ICDDR, B, Dhaka, Bangladesh. [13]Imperial College London, London, UK. [14]Université Lille, INSERM, CHU Lille, Institut Pasteur de Lille, U1167–RID–AGE–Risk Factors and Molecular Determinants of Aging-related Diseases, Lille, France. [15]Department of Epidemiology and Public Health, EA 3430, University of Strasbourg, Strasbourg, France. [16]VA Palo Alto Health Care System, Palo Alto, California, USA. [17]Department of Medicine, Stanford University School of Medicine, Stanford, California, USA. [18]Zilber School of Public Health, University of Wisconsin–Milwaukee, Milwaukee, Wisconsin, USA. [19]Cardiovascular Institute, Mount Sinai Medical Center, Icahn School of Medicine at Mount Sinai, New York, New York, USA. [20]Department of Medicine, Baylor College of Medicine, Houston, Texas, USA. [21]Department of Cardiology, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark. [22]Department of Clinical Biochemistry, Rigshospitalet, Copenhagen, Denmark. [23]Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. [24]Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, Washington, USA. [25]Center for Statistical Genetics, Department of Biostatistics, University of Michigan School of Public Health, Ann Arbor, Michigan, USA. [26]Human Genetics Center, Department of Epidemiology, Human Genetics, and Environmental Sciences, School of Public Health, The University of Texas Health Science Center at Houston, Houston, Texas, USA. [27]Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas, USA. [28]The Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. [29]The Charles Bronfman Institute for Personalized Medicine, Ichan School of Medicine at Mount Sinai, New York, New York, USA. [30]Department of Clinical Biochemistry, Lillebaelt Hospital, Vejle, Denmark. [31]Institute of Regional Health Research, University of Southern Denmark, Odense, Denmark. [32]Clinical Pharmacology, William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK.

[33]Istituto di Ricerca Genetica e Biomedica, Consiglio Nazionale delle Ricerche (CNR), Monserrato, Cagliari, Italy. [34]The Barts Heart Centre, William Harvey Research Institute, Queen Mary University of London, London, UK. [35]NIHR Barts Cardiovascular Biomedical Research Unit, Queen Mary University of London, London, UK. [36]Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, UK. [37]Department of Cardiology, Ealing Hospital NHS Trust, Southall, UK. [38]Imperial College Healthcare NHS Trust, London, UK. [39]Division of Preventive Medicine, Brigham and Women's Hospital, Boston, Massachusetts, USA. [40]Department of Medicine, Harvard Medical School, Boston, Massachusetts, USA. [41]Department of Internal Medicine, Division of Cardiovascular Medicine, University of Michigan, Ann Arbor, Michigan, USA. [42]The Institute for Translational Genomics and Population Sciences, LABioMed at Harbor–UCLA Medical Center, Departments of Pediatrics and Medicine, Los Angeles, California, USA. [43]Medical Department, Lillebaelt Hospital, Vejle, Denmark. [44]NHLBI Framingham Heart Study, Framingham, Massachusetts, USA. [45]Medical Research Institute, Ninewells Hospital and Medical School, University of Dundee, Dundee, UK. [46]Dipartimento di Scienze Biomediche, Università degli Studi di Sassari, Sassari, Italy. [47]Corporal Michael Crescenz VA Medical Center, Philadelphia, Pennsylvania, USA. [48]Department of Surgery, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA. [49]Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, UK. [50]Department of Psychology, University of Edinburgh, Edinburgh, UK. [51]Department of Nutrition and Dietetics, School of Health Science and Education, Harokopio University, Athens, Greece. [52]Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, Tennessee, USA. [53]Department of Medicine, Vanderbilt University Medical Center, Nashville, Tennessee, USA. [54]British Heart Foundation Glasgow Cardiovascular Research Centre, Institute of Cardiovascular and Medical Sciences, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, UK. [55]Montreal Heart Institute, Montreal, Quebec, Canada. [56]Université de Montréal Beaulieu-Saucier Pharmacogenomics Center, Montreal, Quebec, Canada. [57]Université de Montréal, Montreal, Quebec, Canada. [58]Department of Medicine, Oulu University Hospital and University of Oulu, Oulu, Finland. [59]The Icelandic Heart Association, Kopavogur, Iceland. [60]Estonian Genome Center, University of Tartu, Tartu, Estonia. [61]Division of Statistical Genomics, Department of Genetics, Washington University School of Medicine, St. Louis, Missouri, USA. [62]Research Centre in Epidemiology and Preventive Medicine–EPIMED, Department of Medicine and Surgery, University of Insubria, Varese, Italy. [63]Department of Epidemiology, UMR 1027–INSERM, Toulouse University–CHU Toulouse, Toulouse, France. [64]Robertson Centre for Biostatistics, University of Glasgow, Glasgow, UK. [65]Institute of Molecular Medicine, The University of Texas Health Science Center at Houston, Houston, Texas, USA. [66]Department of Clinical Sciences, Genetic and Molecular Epidemiology Unit, Lund University, Malmö, Sweden. [67]Department of Public Health & Clinical Medicine, Umeå University, Umeå, Sweden. [68]Department of Nutrition, Harvard T. H. Chan School of Public Health, Boston, Massachusetts, USA. [69]Genetics of Complex Traits, University of Exeter Medical School, University of Exeter, Exeter, UK. [70]Centro Nacional de Investigaciones Cardiovasculares Carlos III (CNIC), Madrid, Spain. [71]Department of Human Genetics, University of Michigan, Ann Arbor, Michigan, USA. [72]Department of Cardiology, Peking University Third Hospital, Key Laboratory of Cardiovascular Molecular Biology and Regulatory Peptides, Ministry of Health, Beijing, China. [73]National Heart, Lung, and Blood Institute, Bethesda, Maryland, USA. [74]German Center for Diabetes Research (DZD e.V.), Neuherberg, Germany. [75]Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany. [76]Institute of Epidemiology II, Helmholtz Zentrum München–German Research Center for Environmental Health, Neuherberg, Germany. [77]Departments of Medicine and of Biomedical Sciences, Cedars-Sinai Medical Center, Los Angeles, California, USA. [78]Division of Endocrinology, Diabetes and Metabolism, Cedars-Sinai Medical Center, Los Angeles, California, USA. [79]Department of Clinical Sciences, Diabetes and Endocrinology, Clinical Research Centre, Lund University, Malmö, Sweden. [80]Faculty of Medicine, University of Iceland, Reykjavik, Iceland. [81]Faculty of Health Sciences, University of Southern Denmark, Odense, Denmark. [82]Laboratory of Epidemiology and Population Sciences, National Institute on Aging, Bethesda, Maryland, USA. [83]Medical Research Council Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK. [84]Division of Endocrinology and Center for Basic and Translational Obesity Research, Boston Children's Hospital, Boston, Massachusetts, USA. [85]Department of Public Health and General Practice, HUNT Research Centre, Norwegian University of Science and Technology, Levanger, Norway. [86]St. Olavs Hospital, Trondheim University Hospital, Trondheim, Norway. [87]Department of Cardiology, Peking University First Hospital, Beijing, China. [88]K. G. Jebsen Center for Genetic Epidemiology, Department of Public Health and Nursing, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology (NTNU), Trondheim, Norway. [89]Department of Health Sciences, University of Iceland, Reykjavik, Iceland. [90]The Copenhagen City Heart Study, Frederiksberg Hospital, Copenhagen, Denmark. [91]Steno Diabetes Center, Gentofte, Denmark. [92]National Institute of Public Health, Southern Denmark University, Copenhagen, Denmark. [93]Department of Cardiology, Leiden University Medical Center, Leiden, the Netherlands. [94]The Interuniversity Cardiology Institute of the Netherlands, Utrecht, the Netherlands. [95]Department of Clinical Biochemistry and the Copenhagen General Population Study, Herlev and Gentofte Hospital, Copenhagen University Hospital, Copenhagen, Denmark. [96]William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK. [97]Oxford Centre for Diabetes, Endocrinology and Metabolism, Radcliffe Department of Medicine, University of Oxford, Oxford, UK. [98]Oxford NIHR Biomedical Research Centre, Oxford University Hospitals Trust, Oxford, UK. [99]UKCRC Centre of Excellence for Public Health, Queens University, Belfast, UK. [100]Massachusetts Veterans Epidemiology Research and Information Center (MAVERIC), VA Boston Healthcare System, Boston, Massachusetts, USA. [101]Department of Health, National Institute for Health and Welfare, Helsinki, Finland. [102]Department of Medicine and Abdominal Center: Endocrinology, University of Helsinki and Helsinki University Central Hospital, Helsinki, Finland. [103]Minerva Foundation Institute for Medical Research, Helsinki, Finland. [104]National Heart and Lung Institute, Imperial College London, Hammersmith Hospital, London, UK. [105]Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA. [106]Institute of Clinical Medicine, Internal Medicine, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland. [107]Department of Physiology, Institute of Biomedicine, University of Eastern Finland, Kuopio Campus, Kuopio, Finland. [108]Kuopio Research Institute of Exercise Medicine, Kuopio, Finland. [109]Department of Clinical Physiology and Nuclear Medicine, Kuopio University Hospital, Kuopio, Finland. [110]MRC Epidemiology Unit, Institute of Metabolic Science, University of Cambridge School of Clinical Medicine, Cambridge, UK. [111]Faculty of Health and Medical Sciences, University of Denmark, Copenhagen, Denmark. [112]Department of Public Health, Section of General Practice, University of Aarhus, Aarhus, Denmark. [113]Department of Clinical Experimental Research, Rigshospitalet, Glostrup, Denmark. [114]Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. [115]Research Center for Prevention and Health, Copenhagen, Denmark. [116]The Mindich Child Health and Development Institute, Ichan School of Medicine at Mount Sinai, New York, New York, USA. [117]State Key Laboratory of Cardiovascular Disease, Fuwai Hospital, National Center for Cardiovascular Diseases, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China. [118]Cardiovascular Genetics and Genomics Group, Cardiovascular Medicine Unit, Department of Medicine, Solna, Karolinska Institutet, Stockholm, Sweden. [119]Pharmatherapeutics Clinical Research, Pfizer Worldwide R&D, Sollentuna, Sweden. [120]Center for Public Health Genomics, University of Virginia, Charlottesville, Virginia, USA. [121]Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts, USA. [122]Unit of Primary Health Care, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland. [123]Department of Cardiovascular Sciences, University of Leicester, Leicester, UK. [124]NIHR Leicester Biomedical Research Centre, Glenfield Hospital, Leicester, UK. [125]Division of General Internal Medicine, Massachusetts General Hospital, Boston, Massachusetts, USA. [126]Department of Clinical Sciences, University Hospital Malmo Clinical Research Center, Lund University, Malmo, Sweden. [127]Department of Biostatistics, University of Liverpool, Liverpool, UK. [128]Department of Medicine I, Ludwig-Maximilians-University, Munich, Germany. [129]DZHK German Centre for Cardiovascular Research, Munich Heart Alliance, Munich, Germany. [130]Department of Cardiovascular Epidemiology and Population Genetics, National Center for Cardiovascular Investigation, Madrid, Spain. [131]IMDEA-Alimentacion, Madrid, Spain. [132]Nutrition and Genomics Laboratory, Jean Mayer–USDA Human Nutrition Research Center on Aging at Tufts University, Boston, Massachusetts, USA. [133]Department of Medicine, Harvard Medical School, Boston, Massachusetts, USA. [134]Genetics, Merck Sharp & Dohme Corporation, Kenilworth, New Jersey, USA. [135]G3 Pharmaceuticals, Lexington, Massachusetts, USA. [136]Cardiovascular Research Center, Massachusetts General Hospital, Boston, Massachusetts, USA. [137]Institute of Molecular Medicine FIMM, University of Helsinki, Finland. [138]Faculty of Medicine, University of Split, Split, Croatia. [139]Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK. [140]International Centre for Circulatory Health, Imperial College London, London, UK. [141]Kaiser Permanente Washington Health Research Institute, Seattle, Washington, USA. [142]Departments of Epidemiology and Health Services, University of Washington, Seattle, Washington, USA. [143]Departments of Genetics, Medicine, and Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA. [144]Department of Epidemiology, University of Washington, Seattle, Washington, USA. [145]Department of Biobank Research, Umeå University, Umeå, Sweden. [146]Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, Michigan, USA. [147]Imaging, Merck Sharp & Dohme Corporation, Kenilworth, New Jersey, USA. [148]Saw Swee Hock School of Public Health, National University of Singapore, Singapore. [149]Department of Vascular Medicine, Academic Medical Center, University of Amsterdam, Amsterdam, the Netherlands. [150]Department of Twin Research and Genetic Epidemiology, King's College London, London, UK. [151]Division of Population Health Sciences, Ninewells Hospital and Medical School, University of Dundee, Dundee, UK. [152]Generation Scotland, Centre for Genomic and Experimental Medicine, University of Edinburgh, Edinburgh, UK. [153]Scientific Informatics, Merck Sharp & Dohme Corporation, Kenilworth, New Jersey, USA. [154]Wellcome Trust Sanger Institute, Genome Campus, Hinxton, UK. [155]Department of Internal Medicine, Division of Gastroenterology, University of Michigan, Ann Arbor, Michigan, USA. [156]Alzheimer Scotland Dementia

Research Centre, University of Edinburgh, Edinburgh, UK. [157]Department of Haematology, University of Cambridge, Cambridge, UK. [158]Cardiovascular Division, Departments of Medicine and Genetics, Washington University School of Medicine, St. Louis, Missouri, USA. [159]The McDonnell Genome Institute, Washington University School of Medicine, St. Louis, Missouri, USA. [160]IBE, Faculty of Medicine, Ludwig-Maximilians-Universität Munich, Germany. [161]Institute of Genetic Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany. [162]Division of Cardiovascular Medicine, Kanazawa University Graduate School of Medicine, Kanazawa, Japan. [163]Department of Medicine, Division of Molecular Medicine, Columbia University, New York, New York, USA. [164]Department of Genetics, Stanford University School of Medicine, Stanford, California, USA. [165]Department of Gerontology and Geriatrics, Leiden University Medical Center, Leiden, the Netherlands. [166]Chronic Disease Prevention Unit, National Institute for Health and Welfare, Helsinki, Finland. [167]Dasman Diabetes Institute, Dasman, Kuwait. [168]Centre for Vascular Prevention, Danube-University Krems, Krems, Austria. [169]Saudi Diabetes Research Group, King Abdulaziz University, Fahd Medical Research Center, Jeddah, Saudi Arabia. [170]The Heart Centre, Department of Cardiology, Copenhagen University Hospital, Rigshospitalet, Copenhagen, Denmark. [171]Department of Epidemiology, Indiana University Fairbanks School of Public Health, Indianapolis, Indiana, USA. [172]Department of Medicine, Indiana University School of Medicine, Indianapolis, Indiana, USA. [173]Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, Mississippi, USA. [174]Atlanta VA Medical Center, Decatur, Georgia, USA. [175]Emory Clinical Cardiovascular Research Institute, Atlanta, Georgia, USA. [176]Department of Cardiology, Institute of Vascular Medicine, Peking University Third Hospital, Key Laboratory of Molecular Cardiovascular Sciences, Ministry of Education, Beijing, China. [177]Yale College, Yale University, New Haven, Connecticut, USA. [178]A full list of members and affiliations appears in the **Supplementary Note**. [179]Center for Regenerative Medicine, Massachusetts General Hospital, Boston, Massachusetts, USA. [180]Princess Al-Jawhara Al-Brahim Centre of Excellence in Research of Hereditary Disorders (PACER-HD), King Abdulaziz University, Jeddah, Saudi Arabia. [181]These authors contributed equally to this work. [182]These authors jointly directed this work. Correspondence should be addressed to C.J.W. (cristen@umich.edu) or S. Kathiresan (skathiresan1@mgh.harvard.edu).

ONLINE METHODS

Study samples and phenotypes. Seventy-three studies contributed association results for exome chip genotypes and plasma lipid levels. The outcomes were fasting lipid values in mg/dl (TC, HDL-C, LDL-C, and TG) from the baseline, or an earlier exam with fasting measures. If a study had only nonfasting levels, then it contributed to only the TC and HDL-C analyses. LDL-C and TG analyses were performed on only fasting lipid values. Lipid-lowering therapy with statins was not routinely used before the publication of the 4S study in 1994, which demonstrated the clinical benefits of statin therapy. Therefore, for data collected before 1994, no lipid-lowering-medication adjustment was applied. For data collected after 1994, we adjusted the TC values for individuals taking lipid-lowering medication by replacing their TC values by TC/0.8; this adjustment estimates the effect of statins on TC values. No adjustment was made on HDL-C or TG. LDL-C was calculated with the Friedewald equation for subjects with TG <400 mg/dl (LDL-C = TC – HDL-C – (TG/5)). If TC was modified as described above for medication use after 1994, then modified TC was used in this formula. If only measured LDL-C was available in a study, we used LDL/0.7 for subjects on lipid-lowering medication when data were collected after 1994. TG values were natural-log transformed. For each phenotype, residuals were obtained after accounting for age, age squared, sex, and principal components (as needed by each study, up to four), and inverse-normal-transform residuals were created for analysis.

Genotyping and quality control. All studies used Illumina or Affymetrix Human Exome array v1 or v1.1. Genotypes were determined from Zcall[41] or joint calling[42]. Individual studies used the following quality controls: call rate, heterozygosity, sex discordance, GWAS discordance (if GWAS data were available), fingerprint concordance (if available), and PCA outliers.

Association analyses. For each contributing cohort, ancestries were analyzed within the cohorts separately, and studies collected on case–control status analyzed cases separately from controls. We performed both single-variant and gene-level association tests. In the association analysis, we obtained residuals after controlling for sex, age, age squared, and up to four principal components as covariates. Studies that had related samples analyzed the association by using linear mixed models with relatedness estimated from genome-wide SNPs or from pedigrees.

From each study, we collected single-variant score statistics and their covariance matrix for variants in sliding windows across the genome. Summary association test statistics were generated with RAREMETALWORKER or RVTESTS. Using summary association statistics collected from each study, we performed meta-analysis of single-variant association tests by using the Mantel–Haenszel test and constructed burden, SKAT, and variable threshold tests with the approach by Liu *et al.*[14]. For burden and SKAT, we used MAF thresholds of 1% and 5%, and for VT, we applied a MAF threshold of 5%. In the SKAT test, variants are weighted according to their MAFs, using the beta kernel $\beta(1,25)$.

Using covariance matrices between single-variant association statistics, we were also able to perform conditional association analyses centrally, thereby distinguishing genuine signals from 'shadows' of known loci. Details of the methods can be found in Liu *et al.*[14].

We centrally performed quality control for the data. We aligned study-reported reference and alternative alleles with alleles reported in the NHLBI Exome Sequencing Project[43] and removed mislabeled variant sites that could be strand ambiguous. For variant sites in each study, we removed variants that had calls rated <0.9 or had Hardy–Weinberg P values $<1 \times 10^{-7}$. Finally, in additional checks, for each study we visually inspected the scatter plot of variant-allele frequency against frequencies from ancestry-matched populations in the 1000 Genomes Project[44], and we made sure that the strand and allele labels were well calibrated among studies.

Single-variant associations with $P < 2.1 \times 10^{-7}$ (0.05/242,289 variants analyzed) and gene-based associations with $P < 4.2 \times 10^{-7}$ (0.05/[20,000 genes × 6 tests]) were considered significant. Novel loci were defined as being not within 1 Mb of a known lipid GWAS SNP. Additionally, linkage-disequilibrium information was used to determine independent SNPs in which a locus extended beyond 1 Mb. All novel loci reported in this manuscript were >1 Mb from any previously reported locus and independent (r^2 <0.2 was required for variants within 3 Mb).

Sequential forward selection. To identify independently associated variants for each known and newly identified locus, we performed sequential forward selection. We initialized the set of independently associated variants (denoted Φ), starting with the top association signal in the locus. For each iteration, conditioning on variants in Φ, we performed conditional association analyses for all remaining variants. If the top association signal after the conditional analysis remained significant, we added the top variant to the set Φ, then repeated the conditional association analysis. If the top variant after the conditional analysis was no longer significant, we stopped and reported variants in the set Φ as the final set of independent variants for that locus. We used the same single-variant significance threshold ($P < 2.1 \times 10^{-7}$) to determine statistical significance with the sequential forward-selection results (**Supplementary Fig. 6**).

Annotation. Sequence variants were annotated according to refSeq version 1.9, with SEQMINER software (version 5.7)[45]. Transcript-level annotations were obtained and prioritized. When multiple transcript-level annotations were available, they were prioritized according to their functionality and deleteriousness. To implement gene-level association tests, the annotation with the highest priority was used (along with other filtering criteria such as MAFs) to determine the set of variants that were included.

Heritability and proportion of variance-explained estimates. We estimated the proportion of variance explained by the set of 444 independently associated variants. The joint effects of variants in a locus were approximated by $\vec{\beta}_{JOINT} = \mathrm{V}_{\mathrm{META}}^{-1}\vec{U}_{\mathrm{META}}$, where $\vec{U}_{\mathrm{META}}$ is the single-variant score statistic, and $\mathrm{V}_{\mathrm{META}}^{-1}$ is the covariance matrix between them. The covariance between single-variant genetic effects was approximated by the inverse of the variance–covariance matrix of score statistics, i.e., $\mathrm{V}_{\mathrm{META}}^{-1}$. The phenotypic variance explained by the independently associated variants in a locus is given by $\hat{\vec{\beta}}_{joint}^{T}\,\mathrm{cov}(\mathrm{G})\hat{\vec{\beta}}_{JOINT}$, where G represents the genotypes of the analyzed variants.

Refinement of genome-wide association signals. We sought to quantify what proportion of GWAS loci might be due to a protein-altering variant and therefore to directly identify a functional gene. We assumed that a protein-altering variant was the most likely causal variant for each region if it was the top signal, explained the signal, or was independent of the original signal. To identify putative functional coding variants accounting for the effects at known lipid loci, we performed reciprocal conditional analyses to control for the effects of known lipid GWAS or coding variants within 500 kb, because this was the maximum distance for variants within the covariance matrix. Loci where coding variants were the most significant signals were considered as 'coding as top'. Loci where the initial GWAS variants had conditional P >0.01 were considered to be explained by the coding variants. Loci where the coding variants had conditional $P < 2.1 \times 10^{-7}$ were considered to be independent of the initial GWAS signals.

Jak2 p.Val617Phe and plasma cholesterol in a mouse model. Jak2 p.Val617Phe MxCre mice were created, as reported previously[24]. Bone-marrow cells from the WT or Jak2 p.Val617Phe MxCre mice, both treated with poly I:C, were transplanted into irradiated *Ldlr*$^{-/-}$ recipients. After 4 weeks of recovery, the *Ldlr*$^{-/-}$ recipient mice were fed a Western diet (TD88137, Harlan Teklad) for 8 weeks. Plasma was collected, and 250 µl of pooled plasma from seven WT→*Ldlr*$^{-/-}$ or seven *Jak2* Val617Phe→*Ldlr*$^{-/-}$ recipients was subjected to fast protein liquid chromatography on a Sepharose CL-6B size-exclusion column. The TC content in each fraction was assessed with a Cholesterol E kit (Wako Diagnostics).

Validation of *A1CF* with CRISPR–Cas9 in human cells. To knock out *A1CF* in Huh7 and HepG2 human hepatoma cells, three CRISPRs (**Supplementary Table 21**) targeting exon 4 of *A1CF* were constructed by using the lentiviral vector lentiGuide-Puro. Packaged viruses were used to transduce the Cas9-expressing cells for 16 h. Subsequently, cells were cultured in the presence of 5 µg/ml puromycin for 5 d before splitting for assays. Cells for the APOB secretion assay were cultured for 18 h in serum-free medium, then the amount of APOB100 in the medium was measured with an ELISA kit (MABTECH) according to the manufacturer's instructions.

In a rescue experiment, to avoid cutting of the *A1CF* coding region on the recombinant plasmids by previously designed exon-targeting CRISPRs, four new CRISPRs targeting introns flanking exon 4 were applied to deplete endogenous *A1CF*. The sequences for those sgRNAs are available in **Supplementary Table 21**. The A1CF p.Gly398Ser variant was generated by using overlapping PCR and was confirmed by Sanger sequencing. Constructs encoding both the wild type and the A1CF p.Gly398Ser variant were incorporated into lentiviral plasmids, respectively. After transduction, cells were cultured for 48 h in the presence of 100 ng/ml doxycycline to induce recombinant expression of A1CF or the p.Gly398Ser variant before different assays were performed.

A1cf p.Gly390Ser knock-in mice. All procedures used for animal studies were approved by Harvard University's Faculty of Arts and Sciences Institutional Animal Care and Use Committee and were consistent with local, state, and federal regulations, as applicable. Knock-in mice were generated with a guide RNA designed to target the orthologous site of the A1CF p.Gly390Ser variant. *In vitro*–transcribed *cas9* mRNA (100 ng/μL; TriLink BioTechnologies) and guide RNA (50 ng/μL) were co-injected with 100 ng/μL single-stranded DNA oligonucleotide (Integrated DNA Technologies) (**Supplementary Table 21**) into the cytoplasm in fertilized oocytes from C57BL/6J mice. Genomic-DNA samples from founder mice were screened for knock-in mutations by PCR, and the results were confirmed by Sanger sequencing. Positive mice were bred with C57BL/6J mice to generate wild-type and homozygous knock-in mice. Male colony mates at 12 weeks of age were used for lipid measurements. Blood samples were collected from the lateral tail veins after an overnight fast. Plasma triglyceride levels were measured with Infinity Triglycerides Reagent (Thermo Fisher) according to the manufacturer's instructions.

Intersection of lipid-association signals with AMD, CAD, and T2D. To estimate the association of loss-of-function variants in *HBB* with cholesterol levels, participants from the following two consortia were studied: the Global Lipids Genetics Consortium and the Myocardial Infarction Genetics Consortium (MIGen, 27,939 participants in 12 cohorts). A rare loss-of-function variant in *HBB* (c.92+1G>A, rs33971440) was genotyped in participants from the Global Lipids Genetics Consortium Exome Consortium. This variant was pooled with sequence data for the *HBB* gene in MIGen, available for 19,434 participants with blood cholesterol measurements. The association of loss-of-function variants with cholesterol was estimated with linear regression with adjustment for age, sex, and up to five principal components of ancestry. Estimates from genotype and sequence data were pooled by using inverse-variance-weighted fixed-effects meta-analysis.

To estimate the association of loss-of-function variants in *HBB* with CAD, participants from the following two consortia were studied: the CARDIoGRAM Exome Consortium (69,087 participants from 20 studies) and MIGen (12,384 CAD cases and 15,547 controls from 12 studies). 69,086 individuals who were genotyped for the c.92+1G>A variant in the CARDIoGRAM Exome data were pooled with sequence data for *HBB* from 27,931 individuals in MIGen. The association of loss-of-function variants with CAD was estimated through logistic regression with adjustment for age, sex, and up to five principal components of ancestry. Estimates were pooled with inverse-variance-weighted fixed-effects meta-analysis. To estimate the association of loss-of-function variants in *HBB* with hemoglobin and hematocrit levels, estimates from an exome chip analysis of red-blood-cell traits (24,814 individuals) were used.

For 168 variants independently and significantly associated with HDL-C and a MAF >1%, we looked up the association evidence in 16,144 AMD cases and 17,832 controls with exome chip genotypes[46].

For 132 independently and significantly associated LDL-C variants and MAF >1%, we looked up the association evidence in (i) up to 120,575 individuals with or without CAD and exome chip genotypes (42,335 cases and 78,240 controls)[40] and (ii) up to 69,870 individuals with or without T2D. Only 113 of the 132 LDL variants were available in the T2D results. We used a Bonferroni correction for 132 variants to determine the significance of the results ($\alpha = 4.0 \times 10^{-4}$).

Association of *PCSK9* p.Arg46Leu with type 2 diabetes. For evaluating the association of *PCSK9* p.Arg46Leu with risk of T2D, we considered a total of 42,011 T2D cases and 180,834 controls from 30 studies from populations of European ancestry (**Supplementary Table 18**). The variant was directly genotyped in all studies by using the Metabochip or the Exome array. Sample and variant quality control was performed within each study, as described previously[47–50]. Within each study, the variant was tested for association with T2D under an additive model after adjustment for study-specific covariates, including principal components to adjust for population structure. Association summary statistics for the variant for each study was corrected for residual population structure by using the genomic-control inflation factor, as described previously[47–49]. We then combined association summary statistics for the variant across studies via fixed-effects inverse-variance-weighted meta-analysis.

TG variants, lipids, fatty liver, type 2 diabetes, and CAD. Exome chip results for four variants (*LPL* [rs328], p.Ser474Ter; *ANGPTL4* [rs116843064], p.Glu40Lys; *PNPLA3* [rs738409], p.Ile148Met; and *TM6SF2* [rs58542926], p.Glu167Lys) were obtained from the following sources. (i) Lipids: results were from the current analysis. (ii) Fatty liver: between 2002 and 2005, 1,400 individuals from the Framingham Offspring Study and 2,011 individuals from the third generation underwent multidetector computed tomography, through which we evaluated liver attenuation, as previously described[51]. We tested the association of TG variants with CT liver fat after inverse normal transformation. Covariates in the regression models included age, age squared, and sex. A similar analysis was conducted in 3,293 participants of European ancestry from the BioImage study[52]. Association results for liver attenuation from the Framingham and BioImage studies were combined through fixed-effects inverse-variance-weighted meta-analysis. (iii) T2D: data were from the ExTexT2D Consortium[39]. (iv) CAD: results were from the published Myocardial Infarction Genetics and CARDIoGRAM Exome Consortia study[40] and analysis of the UK Biobank, combined through meta-analysis.

Data availability. The meta-analysis summary association statistics that support the findings of this study are available from http://csg.sph.umich.edu/abecasis/public/lipids2017/. A **Life Sciences Reporting Summary** is available.

41. Goldstein, J.I. *et al.* zCall: a rare variant caller for array-based genotyping: genetics and population analysis. *Bioinformatics* **28**, 2543–2545 (2012).
42. Grove, M.L. *et al.* Best practices and joint calling of the HumanExome BeadChip: the CHARGE Consortium. *PLoS One* **8**, e68095 (2013).
43. Tennessen, J.A. *et al.* Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science* **337**, 64–69 (2012).
44. 1000 Genomes Project Consortium. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
45. Zhan, X. & Liu, D.J. SEQMINER: an R-package to facilitate the functional interpretation of sequence-based associations. *Genet. Epidemiol.* **39**, 619–623 (2015).
46. Fritsche, L.G. *et al.* A large genome-wide association study of age-related macular degeneration highlights contributions of rare and common variants. *Nat. Genet.* **48**, 134–143 (2016).
47. Morris, A.P. *et al.* Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat. Genet.* **44**, 981–990 (2012).
48. Gaulton, K.J. *et al.* Genetic fine mapping and genomic annotation defines causal mechanisms at type 2 diabetes susceptibility loci. *Nat. Genet.* **47**, 1415–1425 (2015).
49. Fuchsberger, C. *et al.* The genetic architecture of type 2 diabetes. *Nature* **536**, 41–47 (2016).
50. The UK Biobank. Genotyping and quality control of UK Biobank, a large-scale, extensively phenotyped prospective resource http://www.ukbiobank.ac.uk/wp-content/uploads/2014/04/UKBiobank_genotyping_QC_documentation-web.pdf (2015).
51. Speliotes, E.K. *et al.* Genome-wide association analysis identifies variants associated with nonalcoholic fatty liver disease that have distinct effects on metabolic traits. *PLoS Genet.* **7**, e1001324 (2011).
52. Baber, U. *et al.* Prevalence, impact, and predictive value of detecting subclinical coronary and carotid atherosclerosis in asymptomatic adults: the BioImage study. *J. Am. Coll. Cardiol.* **65**, 1065–1074 (2015).

doi:10.1038/ng.3977

Corresponding Author: Sekar Kathiresan, Cristen Willer

Date: Sep 1, 2017

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work we publish. This form is published with all life science papers and is intended to promote consistency and transparency in reporting. All life sciences submissions use this form; while some list items might not apply to an individual manuscript, all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size

   Describe how sample size was determined.

   Our sample size is ~300,000 individuals with lipids and exome chip genotypes. This sample size is the largest sample that has been analyzed for plasma lipids against a large set of genetic markers.

2. Data exclusions

   Describe any data exclusions.

   Individuals were excluded on a study-specific basis based on quality control of the genotypes. Study-specific exclusion can be found in Table S3.

3. Replication

   Describe whether the experimental findings were reliably reproduced.

   We tested the new single variant association findings for replication in an independent set of up to 286,268 participants from three studies – Nord-Trøndelag Health Study, (HUNT; max n = 62,168), Michigan Genomics Initiative (MGI; www.michigangenomics.org; max n = 6,411) and the Million Veteran Program18 (MVP; max n = 218,117). Of the novel primary trait associations, 73/73 associations were directionally consistent (Table S10); two SNPs were not available for replication (rs201148465, rs75862065). Furthermore, we were able to replicate the associations of 66/73 (90%) at an alpha of 0.05.

4. Randomization

   Describe how samples/organisms/participants were allocated into experimental groups.

   There was no randomization of the participants as this is an observational study.

5. Blinding

   Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

   Blinding was not relevant to the study. It is an observational study.

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

Nature Genetics: doi:10.1038/ng.3977

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or the Methods section if additional space is needed).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒	A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly.
☐	☒	A statement indicating how many times each experiment was replicated
☐	☒	The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☐	☒	A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒	The test results (e.g. *p* values) given as exact values whenever possible and with confidence intervals noted
☐	☒	A summary of the descriptive statistics, including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒	Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.

Single-variant association statistics and inter-marker linkage disequilibrium information summarized across 1 megabase sliding windows were generated from each cohort using RAREMETALWORKER or RVTESTS13,14 software. Meta-analyses of genetic associations were performed using the R-package rareMETALS (version 6.0).

For all studies, we encourage code deposition in a community repository (e.g. GitHub). Authors must make computer code available to editors and reviewers upon request. The *Nature Methods* guidance for providing algorithms and software for publication may be useful for any submission.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.

No unique materials were used.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

1. A1CF, mouse polyclonal (ab89050, Abcam)
2. β-Actin, mouse monoclonal (A5316, Sigma)
3. APOB, mouse monoclonal (sc-393636, Santa Cruz).

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

Huh7 and HepG2 human hepatoma cells

b. Describe the method of cell line authentication used.

Standard authentication was done.

c. Report whether the cell lines were tested for mycoplasma contamination.

Standard testing for mycoplasma contamination.

d. If any of the cell lines used in the paper are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.

No commonly misidentified cell lines were used.

Nature Genetics: doi:10.1038/ng.3977

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.	Jak2 p.Val617Phe MxCre mice were created and reported previously. C57BL/6J mice were used for the A1CF knock-ins.

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.	Covariate-relevant population characteristics of the included studies are available in Table S1 and S2.

Nature Genetics: doi:10.1038/ng.3977