

Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans

Sekar Kathiresan¹⁻³, Olle Melander⁴, Candace Guiducci², Aarti Surti², Noël P Burt², Mark J Rieder⁵, Gregory M Cooper⁵, Charlotta Roos⁶, Benjamin F Voight^{2,7,8}, Aki S Havulinna⁹, Björn Wahlstrand¹⁰, Thomas Hedner¹⁰, Dolores Corella¹¹, E Shyong Tai¹², Jose M Ordovas¹³, Göran Berglund¹⁴, Erkki Vartiainen⁹, Pekka Jousilahti⁹, Bo Hedblad¹⁵, Marja-Riitta Taskinen¹⁶, Christopher Newton-Cheh¹⁻³, Veikko Salomaa⁹, Leena Peltonen^{2,9,17,18}, Leif Groop^{6,19}, David M Altshuler^{2,3,7,8,20} & Marju Orho-Melander⁶

Blood concentrations of lipoproteins and lipids are heritable¹ risk factors for cardiovascular disease^{2,3}. Using genome-wide association data from three studies ($n = 8,816$ that included 2,758 individuals from the Diabetes Genetics Initiative specific to the current paper as well as 1,874 individuals from the FUSION study of type 2 diabetes and 4,184 individuals from the SardiNIA study of aging-associated variables reported in a companion paper in this issue⁴) and targeted replication association analyses in up to 18,554 independent participants, we show that common SNPs at 18 loci are reproducibly associated with concentrations of low-density lipoprotein (LDL) cholesterol, high-density lipoprotein (HDL) cholesterol, and/or triglycerides. Six of these loci are new ($P < 5 \times 10^{-8}$ for each new locus). Of the six newly identified chromosomal regions, two were associated with LDL cholesterol (1p13 near *CELSR2*, *PSRC1* and *SORT1* and 19p13 near *CILP2* and *PBX4*), one with HDL cholesterol (1q42 in *GALNT2*) and five with triglycerides (7q11 near *TBL2* and *MLXIPL*, 8q24 near *TRIB1*, 1q42 in *GALNT2*, 19p13 near *CILP2* and *PBX4* and 1p31 near *ANGPTL3*). At 1p13, the LDL-associated SNP was also strongly correlated with *CELSR2*, *PSRC1*, and *SORT1* transcript levels in human liver, and a proxy for this SNP was recently shown to affect risk for coronary artery

disease⁵. Understanding the molecular, cellular and clinical consequences of the newly identified loci may inform therapy and clinical care.

We recently conducted the Diabetes Genetics Initiative (DGI) genome-wide association study for type 2 diabetes and 18 other traits, including blood lipoprotein and lipid concentrations⁶. Here, we focus on replication analyses related to three traits—concentrations of LDL cholesterol, HDL cholesterol and triglycerides. In DGI, we analyzed the association of 389,878 markers with blood lipoproteins and lipids in 2,758 individuals. From these results, we selected an initial 196 SNPs for replication on the basis of the strength of statistical evidence. We then combined the DGI results with those from two other genome-wide association studies—the Finland–United States Investigation of NIDDM Genetics (FUSION) and the SardiNIA Study of Aging (see companion manuscript for meta-analytic methods⁴)—and selected an additional 30 SNPs for replication on the basis of the combined evidence (see **Supplementary Fig. 1** online for study design). The 226 SNPs selected for replication were tested in up to 18,554 separate participants from three studies (**Table 1**). Statistical evidence from the DGI genome-wide association study and the three replication studies was summarized using a variance-weighted meta-analysis⁷. We

¹Cardiology Division, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. ²Program in Medical and Population Genetics, Broad Institute of the Massachusetts Institute of Technology and Harvard University, Cambridge, Massachusetts 02142, USA. ³Department of Medicine, Harvard Medical School, Boston, Massachusetts 02115, USA. ⁴Department of Clinical Sciences, Hypertension and Cardiovascular Diseases, University Hospital Malmö, Lund University, Malmö 20502, Sweden. ⁵Department of Genome Sciences, University of Washington, Seattle, Washington 98195, USA. ⁶Department of Clinical Sciences, Diabetes and Endocrinology, University Hospital Malmö, Lund University, Malmö 20502, Sweden. ⁷Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. ⁸Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. ⁹National Public Health Institute Helsinki 00300, Finland. ¹⁰Department of Clinical Pharmacology, Sahlgrenska University Hospital, Göteborg 41345, Sweden. ¹¹Department of Preventive Medicine, School of Medicine, University of Valencia, Valencia 46010, Spain. ¹²Department of Endocrinology, Singapore General Hospital, 169608 Singapore. ¹³Nutrition and Genomics Laboratory, Jean Mayer United States Department of Agriculture Human Nutrition Research Center on Aging at Tufts University, Boston, Massachusetts 02111, USA. ¹⁴Department of Internal Medicine, University Hospital Malmö, Lund University, Malmö 20502, Sweden. ¹⁵Department of Epidemiological Research, University Hospital Malmö, Lund University, Malmö 20502, Sweden. ¹⁶Department of Medicine, University of Helsinki, Helsinki 00029, Finland. ¹⁷Department of Medical Genetics, University of Helsinki, Helsinki 00029, Finland. ¹⁸Wellcome Trust Sanger Institute, Cambridge CB10 1SA, UK. ¹⁹Department of Medicine, Helsinki University Hospital, Helsinki 00029, Finland. ²⁰Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. Correspondence should be addressed to S.K. (skathiresan@partners.org) or M.O.-M. (marju.orho-melander@med.lu.se).

Table 1 Sample characteristics

Characteristic	GWAS: DGI <i>n</i> = 2,758	Replication cohort 1: MDC-CC <i>n</i> = 5,519	Replication cohort 2: FINRISK97 <i>n</i> = 7,940	Replication cohort 3: NORDIL <i>n</i> = 5,095
Ascertainment scheme	Cases with type 2 diabetes, controls free of diabetes	Community-based, prospective cohort	Population-based, prospective cohort	Randomized clinical trial of antihypertensive therapy
Mean age (y)	61.5 ± 10.5	57.6 ± 5.9	50.0 ± 13.4	60.3 ± 6.6
Female gender (%)	50.8	58.6	50.3	50.2
Body mass index (kg/m ²)	27.6 ± 4.2	25.8 ± 3.9	26.7 ± 4.5	28.1 ± 4.4
Total cholesterol (mg/dl)	227 ± 44	239 ± 42	214 ± 21	244 ± 45
Low-density lipoprotein cholesterol (mg/dl)	151 ± 40	161 ± 38	135 ± 36	160 ± 43
High-density lipoprotein cholesterol (mg/dl)	50 ± 13	53 ± 14	54 ± 14	53 ± 21
Triglycerides (mg/dl)	146 ± 104	122 ± 71	133 ± 93	159 ± 108
Systolic blood pressure (mmHg)	143 ± 22	142 ± 19	138 ± 41	173 ± 19
Individuals with diabetes mellitus (%)	1,359 (49.3)	464 (8.4)	280 (3.5)	442 (8.7)

Values with '±' are means ± s.d. To convert values for cholesterol to millimoles per liter, multiply by 0.02586. To convert the values for triglycerides to millimoles per liter, multiply by 0.01129. The body-mass index is the weight in kilograms divided by the square of the height in meters. GWAS, genome-wide association study; DGI, Diabetes Genetics Initiative; MDC-CC, Malmö Diet and Cancer Study—Cardiovascular Cohort; NORDIL, Nordic Diitiazem Study.

pre-specified $P < 5 \times 10^{-8}$ as the statistical significance threshold for each new locus.

At 18 independent genomic loci, common DNA sequence variants were reproducibly related to at least one blood lipoprotein or lipid trait (Table 2). Previous studies have identified nine loci with compelling evidence for association between common variants and lipoprotein or lipid concentrations (those of *APOB*, *APOE-APOC1-APOC4-APOC2*, *PCSK9*, *ABCA1*, *APOA1-APOC3-APOA4-APOA5*, *CETP*, *LIPC*, *LPL* and *ANGPTL4*)^{8–16}, and we confirmed eight of these loci (Table 2). Additionally, we recently identified *GCKR*⁶ as a locus associated with triglyceride concentration.

Prior work has provided suggestive but not definitive evidence for common variants at three loci (*LDLR*, *HMGCR* and *LIPG*)^{17,18}. We found compelling evidence for common variants at each of the three loci (Table 2).

Six of the 18 loci have not been previously reported to relate to lipoprotein or lipid traits in humans. For these six newly identified loci, the statistical evidence for association was robust, ranging from $P = 3 \times 10^{-8}$ to $P = 5 \times 10^{-42}$ in the combined analysis of the DGI genome-wide association study and three replication cohorts (Table 2).

For LDL cholesterol, we identified two new loci and confirmed five loci with prior evidence (Table 3). The first new locus for LDL cholesterol is located on chromosome 1p13. SNPs rs599839 and rs646776 were robustly associated with LDL cholesterol (combined $P = 3 \times 10^{-21}$ and 3×10^{-29} , respectively; Table 3). Both SNPs are located in a 97-kb region of linkage disequilibrium containing four genes—*CELSR2*, *PSRC1*, *MYBPHL* and *SORT1*. The two SNPs are highly correlated with one another ($r^2 = 0.89$ in the HapMap population of European ancestry). Each copy of the minor allele at either SNP (24% frequency) decreased LDL cholesterol concentrations by ~5–8 mg/dl (Supplementary Table 1 online).

The second new locus for LDL cholesterol is located on chromosome 19p13 in an intergenic region between *CILP2* and *PBX4*. SNP rs16996148 replicated for association with LDL cholesterol (combined $P = 3 \times 10^{-8}$, Table 3). Two copies of the minor allele at SNP rs16996148 decreased LDL cholesterol concentrations by ~16 mg/dl (Supplementary Table 1).

Besides the 1p13 and 19p13 loci, we confirmed five loci with prior evidence for association with LDL cholesterol concentrations (*APOB*, *APOE-APOC1-APOC4-APOC2*, *LDLR*, *HMGCR* and *PCSK9*, Table 3).

We found that an intronic *LDLR* SNP strongly related to LDL cholesterol. In the cardiovascular cohort of the Malmö Diet and Cancer Study, LDL cholesterol values varied by ~7 mg/dl per copy of the minor allele at the *LDLR* SNP (combined $P = 2 \times 10^{-51}$, Table 3 and Supplementary Table 1). Similarly, an intronic SNP at *HMGCR* was associated with LDL cholesterol (combined $P = 1 \times 10^{-20}$, Table 3).

For HDL cholesterol, we identified one new locus and confirmed six loci for which there was prior evidence of association (Table 4). The new locus for HDL cholesterol is located at 1q42 in an intron of *GALNT2* (SNP rs4846914, combined $P = 2 \times 10^{-13}$ for association, Table 4). Each copy of the minor allele decreased HDL cholesterol concentrations by ~1.5 mg/dl (Supplementary Table 1). In addition, we confirmed six loci with prior evidence (*ABCA1*, *APOA1-APOC3-APOA4-APOA5*, *CETP*, *LIPC*, *LIPG* and *LPL*).

For triglycerides, we identified five new loci (Table 5). The five replicated SNPs are located at 7q11 near *TBL2* and *MLXIPL*, 8q24 near *TRIB1*, 1q42 in *GALNT2*, 19p13 near *CILP2-PBX4* and 1p31 near *ANGPTL3* (combined $P < 5 \times 10^{-8}$ for each SNP, Table 5). Of these, the SNP at 7q11 near *TBL2* and *MLXIPL* had the strongest effect size, with each copy of the minor allele increasing triglyceride concentrations by ~8 mg/dl (Supplementary Table 1). In addition, we confirmed four loci with prior evidence (*APOA1-APOC3-APOA4-APOA5*, *APOB*, *GCKR* and *LPL*).

We observed that SNPs at four of the newly identified loci—19p13 near *CILP2* and *PBX4*, 1q42 in *GALNT2*, 7q11 near *TBL2* and *MLXIPL* and 8q24 near *TRIB1*—were associated with multiple lipoprotein or lipid traits (Table 6 and Supplementary Table 2 online). We did not require the associations with the second and/or third trait to meet a genome-wide association threshold of $P < 5 \times 10^{-8}$. We find these secondary associations to be of interest, as the patterns of association may provide clues to how the locus affects lipoprotein metabolism.

The minor allele at SNP rs16996148 on 19p13 near *CILP2* and *PBX4* was associated with lower concentrations of both LDL cholesterol and triglycerides (Table 6). This pattern of association is similar to that of *APOB* coding SNP rs693, in which a variant allele is associated with both LDL cholesterol and triglycerides in the same direction.

The minor alleles of *GALNT2* SNP rs4846914 as well as SNP rs17145738 on 7q11 near *TBL2* and *MLXIPL* were associated with

Table 2 Genetic loci where common SNPs are associated with blood lipoproteins or lipids

Trait	Unique loci	SNP	Locus	SNP type	Nearest gene(s)	Allele (frequency) ^a	Combined analysis: DGI GWAS & ≤3 replications ^{Pb}	Effect size (s.e.m.) ^c
Newly identified loci								
LDL cholesterol	1	rs646776 or rs599838 ^d	1p13	Intergenic	<i>CELSR2, PSRC1, SORT1</i>	C (0.24) G (0.24)	5×10^{-42}	-0.16 (0.01)
LDL cholesterol	2	rs16996148	19p13	Intergenic	<i>CILP2, PBX4</i>	T (0.10)	3×10^{-8}	-0.10 (0.02)
Triglycerides		rs16996148	19p13	Intergenic	<i>CILP2, PBX4</i>	T (0.10)	4×10^{-9}	-0.10 (0.02)
HDL cholesterol	3	rs4846914	1q42	Intronic	<i>GALNT2</i>	G (0.40)	2×10^{-13}	-0.07 (0.01)
Triglycerides		rs4846914	1q42	Intronic	<i>GALNT2</i>	G (0.40)	7×10^{-15}	0.08 (0.01)
Triglycerides	4	rs17145738	7q11	Intergenic	<i>BCL7B, TBL2, MLXIPL</i>	T (0.13)	7×10^{-22}	-0.14 (0.02)
Triglycerides	5	rs17321515	8q24	3' downstream	<i>TRIB1</i>	G (0.49)	4×10^{-17}	-0.08 (0.01)
Triglycerides	6	rs12130333	1p31	Intergenic	<i>ANGPTL3, DOCK7, ATG4C</i>	T (0.22)	2×10^{-8}	-0.11 (0.02)
Loci with prior evidence								
LDL cholesterol	7	rs693	2p24	Coding	<i>APOB</i>	A (0.48)	1×10^{-21}	0.12 (0.01)
Triglycerides		rs693	2p24	Coding	<i>APOB</i>	A (0.48)	2×10^{-7}	0.08 (0.02)
LDL cholesterol	8	rs4420638	19q13	5' upstream	<i>APOE-C1-C4-C2</i>	G (0.20)	1×10^{-60}	0.19 (0.02)
LDL cholesterol	9	rs12654264	5q13	Intronic	<i>HMGCR</i>	T (0.39)	1×10^{-20}	0.10 (0.01)
LDL cholesterol	10	rs6511720	19p13	Intronic	<i>LDLR</i>	T (0.10)	2×10^{-51}	-0.26 (0.02)
LDL cholesterol	11	rs11591147 ^e	1p32	Coding	<i>PCSK9</i>	T (0.01)	2×10^{-44}	-0.47 (0.03)
HDL cholesterol	12	rs3890182	9q31	Intronic	<i>ABCA1</i>	A (0.13)	3×10^{-10}	-0.10 (0.02)
HDL cholesterol	13	rs28927680 ^f	11q23	5' upstream	<i>APOA1-C3-A4-A5, ZNF259, BUD13</i>	G (0.07)	2×10^{-5}	-0.13 (0.03)
Triglycerides		rs28927680	11q23	5' upstream	<i>APOA1-C3-A4-A5, ZNF259, BUD13</i>	G (0.07)	2×10^{-17}	0.26 (0.03)
HDL cholesterol	14	rs1800775	16q13	5' upstream	<i>CETP</i>	C (0.51)	1×10^{-73}	-0.18 (0.01)
HDL cholesterol	15	rs1800588	15q21	5' upstream	<i>LIPC</i>	T (0.21)	2×10^{-32}	0.14 (0.01)
HDL cholesterol	16	rs2156552	18q21	Intergenic	<i>LIPG, ACAA2</i>	A (0.18)	2×10^{-7}	-0.07 (0.01)
HDL cholesterol	17	rs328	8p21	Coding	<i>LPL</i>	G (0.09)	9×10^{-23}	0.17 (0.02)
Triglycerides		rs328	8p21	Coding	<i>LPL</i>	G (0.09)	2×10^{-28}	-0.19 (0.02)
Triglycerides	18	rs780094 ^g	2p23	Intronic	<i>GCKR</i>	T (0.34)	3×10^{-14}	0.13 (0.02)

^aAlleles for the SNP on the forward strand of human genome reference sequence (National Center for Biotechnology Information Build 35) were modeled. Allele frequency in the MDC-CC replication sample is presented, except in the case of rs599838, where the frequency in NORDIL is shown. ^bVariance-weighted meta-analysis performed using data from up to four samples (DGI, MDC-CC, FINRISK and NORDIL) as described in Methods. ^cEffect size shown is beta-coefficient (β), which represents the proportion of 1 s.d. change (in standardized blood lipid residual with mean = 0 and s.d. = 1 after adjustment for age, age², gender and diabetes status) per copy of the allele modeled. ^dData shown are for a combination of rs646776 results from MDC-CC and FINRISK97 and results for a proxy (rs599839, $r^2 = 0.89$ between two SNPs in HapMap population of European ancestry) from DGI and NORDIL. Data for each SNP are separately shown in **Table 3**. ^ePCSK9 SNP rs11206510 (17% minor allele frequency) was associated with LDL cholesterol in the combined analysis of DGI/FUSION/Sardinia ($P = 7.5 \times 10^{-6}$); however, to confirm this locus, we selected rs11591147 for replication because of prior reports suggesting a very strong effect on LDL cholesterol²⁶. ^frs28927680 is highly correlated with *APOA1-APOC3-APOA4-APOA5* coding SNP rs3133506 ($r^2 = 0.98$ in MDC-CC); SNP rs3133506 (also known as 56C>G) has been reproducibly related to HDL cholesterol and triglyceride concentrations. ^gData in this row on *GCKR* has been previously reported by our group⁶ and is presented here for completeness.

both triglyceride and HDL cholesterol concentrations: rs4846914 is associated with lower HDL concentrations and higher triglyceride concentrations, and rs17145738 is associated with higher HDL concentrations and lower triglyceride concentrations (**Table 6**). These patterns of association are similar to that of the common *LPL* nonsense mutation S447X (rs328).

SNP rs17321515 at 8q24 near *TRIB1* was strongly associated with triglycerides and was also associated with LDL cholesterol and HDL cholesterol (**Table 6**). The minor G allele at this SNP was associated with lower triglycerides, lower LDL cholesterol and higher HDL cholesterol. This pattern of association has not been previously described for any lipid-modulating SNP.

Of note, for the 23 associations from 18 common alleles in this study, we found that the effect size of an allele varied inversely with allele frequency ($r = -0.49$, $P = 0.01$). For example, lower-frequency alleles, such as the 1% frequency allele at *PCSK9*, affected LDL cholesterol concentrations by ~ 0.5 s.d. units, whereas a 48% frequency allele at *APOB* affected LDL cholesterol by ~ 0.1 s.d. units (**Table 2**). Such an inverse relationship is predicted if alleles perturbing physiology are deleterious during evolution, as such alleles would not rise to a high frequency in the population.

Having observed that common variants at 18 loci are convincingly associated with lipoprotein- or lipid-related traits, we next addressed the extent to which these alleles explain inter-individual variability in

lipoprotein or lipid concentrations. In the cardiovascular cohort of the Malmö Diet and Cancer Study, after accounting for age, age², gender and diabetes status, we found that, in sum, seven SNPs explained an additional 5.7% of the residual LDL cholesterol variance (**Table 3**). Meanwhile, seven SNPs explained an additional 5.2% of the residual HDL cholesterol variance (**Table 4**) and nine SNPs explained an additional 4.5% of the residual triglyceride level variance (**Table 5**).

Though these common alleles explain an appreciable fraction of variance, it is likely that these values are underestimates of the impact of each validated locus. As nine of the loci with common variants (*ABCA1*, *APOA1-APOC3-APOA4-APOA5*, *APOB*, *APOE*, *CETP*, *LDLR*, *LIPC*, *LPL* and *PCSK9*) have also been shown to cause mendelian syndromes or harbor multiple rare alleles that contribute to trait variation¹⁹, sequencing of each validated locus will be required to discover all common and rare variants and determine the full impact of each locus.

It is not yet clear what the causal variants or even the causal genes are at the new loci. Each of the six associated SNPs is noncoding. The genes nearest to the associated SNPs are annotated in **Table 2**.

However, the linkage disequilibrium pattern and the genes in the associated intervals suggest functional hypotheses. At 19p13, the variant associated with LDL cholesterol (located between *CILP2* and *PBX4*) is in high linkage disequilibrium with a nonsynonymous

Table 3 Genetic loci where common SNPs are associated with blood low-density lipoprotein cholesterol

SNP(s)	Locus	SNP type	Nearest gene(s)	Allele (frequency) ^a	Combined analysis:	GWAS:	Replication 1:	Replication 2:	Replication 3:
					DGI GWAS & ≤3 replications	DGI	MDC-CC	FINRISK 97	NORDIL
					<i>P</i>	<i>P</i>	<i>P</i>	<i>P</i>	<i>P</i>
					β	β	β	β	β
					(s.e.m.) ^b	(s.e.m.)	(s.e.m.)	(s.e.m.)	(s.e.m.)
Newly identified loci									
rs646776 ^c	1p13	Intergenic	<i>CELSR2</i> , <i>PSRC1</i> , <i>SORT1</i>	C (0.24)	3×10^{-29} -0.16 (0.01)	8×10^{-8d} -0.18 (0.03)	2×10^{-10} -0.15 (0.02)	1×10^{-13} -0.15 (0.02)	-
rs599839 ^c				G (0.24)	3×10^{-21} -0.18 (0.02)	9×10^{-8} -0.18 (0.03)	-	-	1×10^{-14} -0.18 (0.02)
rs16996148	19p13	Intergenic	<i>CILP2</i> , <i>PBX4</i>	T (0.10)	3×10^{-8} -0.10 (0.02)	0.04 -0.10 (0.05)	0.01 -0.09 (0.03)	0.12 -0.05 (0.03)	3×10^{-6} -0.15 (0.03)
Loci with prior evidence									
rs693	2p24	Coding	<i>APOB</i>	A (0.48)	1×10^{-21} 0.12 (0.01)	7×10^{-7} 0.14 (0.03)	2×10^{-11} 0.13 (0.02)	-	1×10^{-6} 0.09 (0.02)
rs4420638	19q13	5' upstream	<i>APOE-C1-C4-</i> <i>C2</i>	G (0.20)	1×10^{-60} 0.19 (0.02)	3×10^{-13} 0.25 (0.03)	3×10^{-21} 0.23 (0.02)	8×10^{-19} 0.16 (0.02)	6×10^{-13} 0.18 (0.03)
rs12654264	5q13	Intronic	<i>HMGCR</i>	T (0.39)	1×10^{-20} 0.10 (0.01)	0.0004 0.10 (0.03)	0.002 0.06 (0.02)	9×10^{-11} 0.11 (0.02)	5×10^{-7} 0.11 (0.02)
rs6511720	19p13	Intronic	<i>LDLR</i>	T (0.10)	2×10^{-51} -0.26 (0.02)	9×10^{-7d} -0.27 (0.05)	8×10^{-14} -0.25 (0.03)	9×10^{-22} -0.26 (0.03)	3×10^{-13} -0.25 (0.03)
rs11591147 ^e	1p32	Coding	<i>PCSK9</i>	T (0.01)	2×10^{-44} -0.47 (0.03)	-	7×10^{-7} -0.45 (0.09)	2×10^{-31} -0.48 (0.04)	8×10^{-9} -0.47 (0.08)

A dash (-) indicates that we did not genotype that SNP in the sample. Beta-coefficient (β) represents the proportion of 1 s.d. change in standardized LDL cholesterol residual (mean = 0, s.d. = 1 after adjustment for age, age², gender, and diabetes status) per copy of the allele modeled.

^aAlleles for the SNP on the forward strand of human genome reference sequence (National Center for Biotechnology Information Build 35) were modeled. Allele frequency in the MDC-CC sample is presented, except for rs599839, where allele frequency in DGI is presented. ^bVariance-weighted meta-analysis performed using data from up to four samples (DGI, MDC-CC, FINRISK and NORDIL), as described in Methods. ^crs646776 and rs599839 are in high linkage disequilibrium with $r^2 = 0.89$ in the HapMap population of European ancestry. ^dAssociation data from imputed SNP genotypes. ^e*PCSK9* SNP rs11206510 (17% minor allele frequency) was associated with LDL cholesterol in the combined analysis of DGI/FUSION/Sardinia ($P = 7.5 \times 10^{-6}$); however, to confirm this locus, we elected to take forward the previously reported SNP rs11591147 for replication because of its reported strong effect on LDL cholesterol²⁶.

coding SNP in the *CSPG3* gene encoding neurocan (rs2228603, 329 kb upstream, $r^2 = 0.85$ in HapMap population of European ancestry), suggesting that *CSPG3* may be the causal gene at the locus. At the 1q42 locus for HDL cholesterol and triglycerides, *GALNT2* encodes polypeptide *N*-acetylgalactosaminyltransferase 2, an enzyme involved in O-linked glycosylation and transfer of *N*-acetylgalactosamine to the serine or threonine residues on proteins. O-linked glycosylation has a regulatory role for many proteins²⁰. This suggests the hypothesis that enzymatic glycosylation of any of a number of proteins involved in HDL cholesterol and triglyceride metabolism may lead to the observed pattern of association. At the 7q11 locus for triglycerides, the associated interval includes *MLXIPL*, encoding a transcription factor recently described to connect carbohydrate flux with fatty-acid synthesis in the liver (also called carbohydrate response element binding protein or ChREBP)²¹. Finally, inactivating mutations in *ANGPTL3* (encoding angiopoietin-like 3) have already been demonstrated to lead to low triglycerides in mice²².

We next considered one mechanism by which SNPs (and particularly noncoding SNPs) may relate to traits, namely, the regulation of local gene expression. We analyzed the correlation of lipid-associated

SNPs with mRNA transcript levels of nearby genes in 60 human liver samples. At five of the six newly identified loci, lipid-associated SNPs showed no effect on expression of local genes ($P > 0.05$).

However, SNP rs646776 at the 1p13 locus was strongly associated with transcript concentrations of not only a single gene, but three neighboring genes: *SORT1* ($P = 3 \times 10^{-26}$), *CELSR2* ($P = 2 \times 10^{-12}$) and *PSRC1* ($P = 3 \times 10^{-12}$) (**Supplementary Fig. 2** online). SNP rs646776 explained 86%, 58%, and 58% of the inter-individual variability in *SORT1*, *CELSR2* and *PSRC1* transcript concentrations, respectively. In analyses conditioning on either the *CELSR2* or *PSRC1* transcript levels, rs646776 remained associated with *SORT1* transcript concentration ($P = 1 \times 10^{-5}$ and 1×10^{-5} , respectively). Conversely, after *SORT1* transcript level was accounted for, rs646776 was weakly or not associated with *PSRC1* or *CELSR2* ($P = 0.04$ and 0.81 , respectively). Overall, our results suggest that variation at the 1p13 interval may have a regional effect on gene expression.

SORT1, or sortilin, functions both as a sorting protein and as a cell-surface receptor, and it is abundant in skeletal muscle and adipocytes^{23,24}. As a sorting protein, sortilin enables insulin-mediated glucose uptake by catalyzing the biogenesis of insulin-sensitive vesicles

Table 4 Genetic loci where common SNPs are associated with blood high-density lipoprotein cholesterol

SNP	Locus	SNP type	Nearest gene(s)	Allele (frequency) ^a	Combined analysis:	GWAS:	Replication 1:	Replication 2:	Replication 3:				
					GWAS & ≤3	DGI	MDC-CC	FINRISK97	NORDIL				
					Replications					<i>P</i>	<i>P</i>	<i>P</i>	<i>P</i>
					β	β	β	β	β				
Newly identified locus													
rs4846914	1q42	Intronic	<i>GALNT2</i>	G (0.40)	2×10^{-13} -0.07 (0.01)	3×10^{-4} -0.10 (0.03)	1×10^{-6} -0.10 (0.02)	1×10^{-4} -0.06 (0.02)	0.01 -0.05 (0.02)				
Loci with prior evidence													
rs3890182	9q31	Intronic	<i>ABCA1</i>	A (0.13)	3×10^{-10} -0.10 (0.02)	3×10^{-5} -0.17 (0.04)	0.003 -0.09 (0.03)	0.001 -0.09 (0.03)	0.02 -0.08 (0.03)				
rs28927680 ^c	11q23	5' upstream	<i>APOA1-C3-A4-A5</i> , <i>ZNF259</i> , <i>BUD13</i>	G (0.07)	2×10^{-5} -0.13 (0.03)	0.31 -0.06 (0.05)	6×10^{-6} -0.17 (0.04)	-	-				
rs1800775	16q13	5' upstream	<i>CETP</i>	C (0.51)	1×10^{-73} -0.18 (0.01)	3×10^{-13} -0.20 (0.03)	2×10^{-29} -0.22 (0.02)	1×10^{-23} -0.17 (0.02)	2×10^{-12} -0.14 (0.02)				
rs1800588	15q21	5' upstream	<i>LIPC</i>	T (0.21)	2×10^{-32} 0.14 (0.01)	3×10^{-5d} 0.15 (0.03)	4×10^{-10} 0.15 (0.02)	8×10^{-17} 0.16 (0.02)	3×10^{-5} 0.10 (0.02)				
rs2156552	18q21	Intergenic	<i>LIPG</i> , <i>ACAA2</i>	A (0.18)	2×10^{-7} -0.07 (0.01)	0.02 -0.09 (0.03)	0.02 -0.06 (0.03)	0.001 -0.07 (0.02)	0.03 -0.06 (0.03)				
rs328	8p21	Coding	<i>LPL</i>	G (0.09)	9×10^{-23} 0.17 (0.02)	3×10^{-4d} 0.17 (0.05)	3×10^{-12} 0.23 (0.03)	1×10^{-6} 0.14 (0.03)	6×10^{-5} 0.14 (0.04)				

A dash (-) indicates that we did not genotype that SNP in sample. Beta-coefficient (β) represents the proportion of 1 s.d. change in standardized HDL cholesterol residual (mean = 0, s.d. = 1 after adjustment for age, age², gender and diabetes status) per copy of the allele modeled.

^aAlleles for the SNP on the forward strand of human genome reference sequence (National Center for Biotechnology Information Build 35) were modeled. Allele frequency in the MDC-CC sample is presented. ^bVariance-weighted meta-analysis performed using data from up to four samples (DGI, MDC-CC, FINRISK and NORDIL), as described in Methods. ^crs28927680 is highly correlated with *APOA1-APOC3-APOA4-APOA5* coding SNP rs3133506 ($r^2 = 0.98$ in MDC-CC); SNP rs3133506 (also known as 56C>G) has been reproducibly related to HDL cholesterol and triglyceride concentrations. ^dAssociation data from imputed SNP genotypes.

that transport the glucose transporter GLUT4 to the plasma membrane. In addition, as a multiligand receptor, sortilin can bind several proteins, including lipoprotein lipase, and potentially facilitate lipoprotein uptake. Overall, these observations suggest a mechanism by which increased sortilin expression seen with the C allele (at SNP rs646776) could lead to lower circulating LDL cholesterol concentrations.

Notably, a proxy for SNP rs646776 at the 1p13 locus, SNP rs599839, was recently reported to affect risk of coronary artery disease⁵. SNP rs599839 was also related to LDL cholesterol (Table 3) in our study, and the allele associated with lower LDL cholesterol (G allele, 24% frequency) was the same as that correlated with lower risk of coronary artery disease (odds ratio 0.78; $P = 4.0 \times 10^{-9}$)⁵.

As participants in the initial and replication studies were of European ancestry, it remains to be shown whether the new loci will be associated with lipid-related traits in individuals of other ancestries. In a pilot study, we tested whether the six SNPs from the six new loci identified in those of European ancestry would be associated with lipoprotein or lipid traits in a multiethnic sample. We studied 4,259 participants from the Singapore National Health Survey 98 comprising ethnic Chinese, Indians and Malays²⁵. SNPs at two of the six loci (1p13 near *CELSR2-PSRC1-SORT1* associated with LDL cholesterol;

7q11 near *TBL2-MLXIPL* associated with triglycerides) replicated for association in each of the three ethnic groups (Supplementary Table 3 online). Because of well-known differences in linkage disequilibrium structure and allele frequencies across populations of different ancestries, a comprehensive testing of genetic variation at each new locus is needed for each ethnic group.

We have obtained definitive evidence for six new independent loci at which common genetic variation influences one or more lipoprotein or lipid traits. By establishing these loci as relevant to lipoprotein metabolism in humans, we nominate these as high-priority targets for further investigation. Before considering these loci as targets for pharmacological therapy, it will be critical to assess whether causal alleles at each locus affect risk for cardiovascular disease. If alleles are convincingly associated with risk of cardiovascular disease (as has been shown for *PCSK9* (ref. 26)), this would give *in vivo* human proof for the locus as a valid target and support a path forward.

METHODS

Study samples. DGI is a genome-wide association study of type 2 diabetes mellitus and 18 other traits, including blood lipoproteins and lipids. Details of the study design and ascertainment scheme have been recently described⁶. Of



Table 5 Genetic loci where common SNPs are associated with blood triglycerides

SNP	Locus	SNP type	Nearest gene(s)	Allele (frequency) ^a	Combined analysis:	GWAS:	Replication 1:	Replication 2:	Replication 3:				
					GWAS & ≤3	DGI	MDC-CC	FINRISK97	NORDIL				
					Replications					<i>P</i>	<i>P</i>	<i>P</i>	<i>P</i>
					<i>P</i>	<i>P</i>	<i>P</i>	<i>P</i>	<i>P</i>				
					β	β	β	β	β				
					(s.e.m.) ^b	(s.e.m.)	(s.e.m.)	(s.e.m.)	(s.e.m.)				
Newly identified loci													
rs17145738	7q11	Intergenic	<i>BCL7B</i> , <i>TBL2</i> , <i>MLXIPL</i>	T (0.13)	7×10^{-22} -0.14 (0.02)	0.003 -0.12 (0.04)	3×10^{-8} -0.17 (0.03)	2×10^{-7} -0.13 (0.02)	3×10^{-7} -0.16 (0.03)				
rs17321515	8q24	3' downstream	<i>TRIB1</i>	G (0.49)	4×10^{-17} -0.08 (0.01)	7×10^{-4} -0.10 (0.03)	1×10^{-5} -0.09 (0.02)	7×10^{-5} -0.07 (0.02)	8×10^{-7} -0.10 (0.02)				
rs4846914	1q42	Intronic	<i>GALNT2</i>	G (0.40)	7×10^{-15} 0.08 (0.01)	9×10^{-5} 0.11 (0.03)	0.001 0.07 (0.02)	2×10^{-8} 0.09 (0.02)	0.01 0.05 (0.02)				
rs16996148	19p13	Intergenic	<i>CILP2</i> , <i>PBX4</i>	T (0.10)	4×10^{-9} -0.10 (0.02)	0.05 -0.09 (0.05)	0.23 -0.04 (0.03)	2×10^{-5} -0.14 (0.03)	3×10^{-5} -0.13 (0.03)				
rs12130333	1p31	Intergenic	<i>ANGPTL3</i> , <i>DOCK7</i> , <i>ATG4C</i>	T (0.22)	2×10^{-8} -0.11 (0.02)	0.0006 -0.12 (0.03)	2×10^{-5} -0.10 (0.02)	- - -	- - -				
Loci with prior evidence													
rs28927680 ^c	11q23	5' upstream	<i>APOA1-C3</i> - <i>A4-A5</i> , <i>ZNF259</i> , <i>BUD13</i>	G (0.07)	2×10^{-17} 0.26 (0.03)	6×10^{-5} 0.22 (0.05)	7×10^{-14} 0.29 (0.04)	- - -	- - -				
rs693	2p24	Coding	<i>APOB</i>	A (0.48)	2×10^{-7} 0.08 (0.02)	7×10^{-4} 0.09 (0.03)	7×10^{-5} 0.08 (0.02)	- - -	- - -				
rs780094 ^d	2p23	Intronic	<i>GCKR</i>	T (0.34)	3×10^{-14} 0.13 (0.02)	4×10^{-8} 0.16 (0.03)	5×10^{-8} 0.11 (0.02)	- - -	- - -				
rs328	8p21	Coding	<i>LPL</i>	G (0.09)	2×10^{-28} -0.19 (0.02)	4×10^{-7e} -0.24 (0.05)	9×10^{-9} -0.20 (0.03)	2×10^{-10} -0.18 (0.03)	2×10^{-6} -0.16 (0.03)				

A dash (-) indicates that we did not genotype that SNP in sample. Beta-coefficient (β) represents the proportion of 1 s.d. change in standardized log triglycerides residual (mean = 0, s.d. = 1 after adjustment for age, age², gender and diabetes status) per copy of the allele modeled.

^aAlleles for the SNP on the forward strand of human genome reference sequence (National Center for Biotechnology Information Build 35) were modeled. Allele frequency in the MDC-CC sample is presented. ^bVariance-weighted meta-analysis performed using data from up to four samples (DGI, MDC-CC, FINRISK and NORDIL), as described in Methods. ^crs28927680 is highly correlated with *APOA1-APOC3-APOA4-APOA5* coding SNP rs3133506 ($r^2 = 0.98$ in MDC-CC); SNP rs3133506 (also known as 56C>G) has been reproducibly related to triglyceride and HDL cholesterol concentrations. ^dData in this row on *GCKR* has been previously published⁶ and is presented here for completeness. ^eAssociation data from imputed SNP genotypes.

the 2,931 participants, 2,758 had DNA samples and data available for at least one lipoprotein or lipid phenotype.

Replication cohort 1 consisted of the Malmö Diet and Cancer Study, a community-based prospective epidemiologic cohort of 28,449 persons recruited for a baseline examination between 1991 and 1996 (ref. 27). From this cohort, 6,103 persons were randomly selected to participate in the cardiovascular cohort, which sought to investigate risk factors for cardiovascular disease. All participants underwent a medical history assessment and a physical examination. Of the participants in the cardiovascular cohort, 5,519 had DNA samples and data available for at least one lipoprotein or lipid phenotype.

Replication cohort 2 consisted of FINRISK97, a population-based cross-sectional survey designed to study the prevalence of cardiovascular risk factors in Finland²⁸. Surveys are conducted every 5 years, and the 1997 survey included 8,389 Finnish men and women aged 25–74. Participants underwent a

physical examination and completed a questionnaire regarding cardiovascular risk factors. Of these FINRISK97 participants, 7,940 had DNA samples and data available for at least one lipoprotein or lipid phenotype.

Replication cohort 3 consisted of NORDIL, a randomized controlled trial that compared calcium antagonists with diuretics and beta-blockers on cardiovascular morbidity and mortality in hypertension²⁹. Participants from Norway and Sweden were enrolled between 1992 and 1999. Of the 10,881 participants in this clinical trial, the 5,152 participants from Sweden form the sample for this report. Of these participants, 5,095 had DNA samples and data available from the baseline examination for at least one lipoprotein or lipid phenotype.

All participants in the DGI genome-wide association study and the three replication studies were of self-reported European ancestry.

We attempted to extend our replicated findings to a multiethnic sample from Singapore, the Singapore National Health Survey 98 (NHS98). Singapore



Table 6 Effects of loci on multiple lipoprotein or lipid traits

SNP	Locus	Nearest gene(s)	Allele (frequency) ^a	LDL cholesterol ^b	HDL cholesterol	Triglycerides
Newly identified loci						
rs646776	1p13	<i>CELSR2</i> , <i>PSRC1</i> , <i>SORT1</i>	C (0.24)	↑		
rs16996148	19p13	<i>CILP2</i> , <i>PBX4</i> ,	T (0.10)	↓		↓
rs4846914	1q42	<i>GALNT2</i>	G (0.40)		↓	↑
rs17145738	7q11	<i>BCL7B</i> , <i>TBL2</i> , <i>MLXIPL</i>	T (0.13)		↑	↓
rs17321515	8q24	<i>TRIB1</i>	G (0.49)	↓	↑	↓
rs12130333	1p31	<i>ANGPTL3</i> , <i>DOCK7</i> , <i>ATG4C</i>	T (0.22)			↓
Selected loci with prior evidence						
rs693	2p24	<i>APOB</i>	A (0.48)	↑		↑
rs28927680	11q23	<i>APOA1-C3-A4-A5</i> , <i>ZNF259</i> , <i>BUD13</i>	G (0.07)		↓	↑
rs328	8p21	<i>LPL</i>	G (0.09)		↑	↓

We required each new locus to be associated with at least one lipoprotein or lipid trait at a genome-wide association threshold ($P < 5 \times 10^{-8}$). We did not require the associations with the second and/or third trait to be at this threshold. ^aAlleles for the SNP on the forward strand of human genome reference sequence (National Center for Biotechnology Information Build 35) were modeled. Allele frequency in the MDC-CC replication sample is presented. ^b↑ indicates that the modeled allele increases trait value; ↓, lowers trait value. Allele shown in column 4 is modeled. Data derived from a combined analysis of DGI and three replication studies, as described in **Table 2**.

NHS98 was an initiative to determine the risk factors for the major noncommunicable diseases in Singapore²⁵. The sampling scheme captured the three major Singaporean ethnic groups—Chinese, Malays and Asian Indians—and has been previously described²⁵. Of the 4,723 individuals who participated, we studied a sample of 4,259 subjects who had DNA samples and data available for lipoprotein or lipid phenotypes.

All participants in the studies provided written informed consent. The individual study protocols were approved by local ethical committees at each participating institution. The conduct of the genome-wide association study was approved by the Massachusetts Institute of Technology Review Board.

Determination of blood lipid concentration. In DGI, the Malmö Diet and Cancer Study, NORDIL and Singapore NHS98, we measured total cholesterol, HDL cholesterol and triglyceride concentrations in fasting blood samples drawn at the baseline examination for each study. In FINRISK97, individuals were instructed to fast for at least 4 h with a mean fasting time of 6 ± 4 h. We measured lipid concentrations according to standard enzymatic methods. LDL cholesterol concentrations were calculated according to Friedewald's formula, and missing values were assigned to individuals with triglycerides > 400 mg/dl.

SNP selection and genotyping. The overall SNP selection procedure is summarized in **Supplementary Figure 1**. First, we attempted to validate 196 SNPs from a single genome-wide association study for blood lipid traits—the DGI—for which the full association results are publicly available (see URLs section below)⁶. From these association results, we selected 196 SNPs representing the top 40 independent loci from an interim and a final analysis of the

DGI scan. These 196 SNPs were genotyped in a single replication cohort, the cardiovascular cohort of the Malmö Diet and Cancer Study.

Second, during the replication effort described above, we were contacted by two groups—FUSION and SardiNIA—who had performed genome-wide association studies for lipid traits⁴. We agreed to share data and collaboratively carried out a meta-analysis of all three studies, as described in the companion manuscript⁴. For each of three traits—LDL cholesterol, HDL cholesterol and triglycerides—the final meta-analysis involved the association testing of ~ 2.2 million SNPs in 8,816 individuals.

From this meta-analysis, we selected an additional 30 SNPs for replication testing. These 30 SNPs represented independent loci with $P < 10^{-4}$ in the meta-analysis. SNPs at several loci that met this criteria (for example, *GRIN3A*, *LCAT* and *FARS2* (ref. 4)) were not taken forward into replication because of failure to design primers or probes within one Sequenom genotyping assay. At *PCSK9*, SNP rs11206510 (17% minor allele frequency) was associated with LDL cholesterol in the combined analysis of DGI, FUSION and SardiNIA ($P = 7.5 \times 10^{-6}$); however, to confirm this locus, we substituted rs11591147 for replication because of prior reports suggesting a very strong effect of rs11591147 on LDL cholesterol²⁶. Replication genotyping for 30 SNPs was conducted in up to 18,554 participants from all three replication cohorts: cardiovascular cohort of the Malmö Diet and Cancer Study, FINRISK97 and NORDIL. Genotyping was conducted for six SNPs at six new loci in Singapore NHS98.

We carried out genotyping using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry on the Sequenom MassARRAY platform or using allelic discrimination method on the ABI 7900 instrument (Applied Biosystems). We excluded SNPs with a Hardy-Weinberg equilibrium $P < 0.001$. SNPs with a genotyping call rate $< 90\%$ were also excluded from analysis. The average genotyping call rate exceeded 97% in each of the replication samples. We estimated a genotyping error rate for the Sequenom platform using 15 samples placed in quadruplicate and found this rate to be $< 0.7\%$.

mRNA expression and genotyping in human liver samples. Total RNA and DNA were extracted from 60 human liver tissue samples from the University of Washington School of Pharmacy Human Liver Bank. Full details of transcript level measurement, SNP genotyping and association analyses of SNPs with transcript concentrations are described in the **Supplementary Methods** online.

Statistical analysis. For each participant in the three replication cohorts and the one multiethnic cohort, we log transformed triglyceride values. We adjusted LDL cholesterol, HDL cholesterol and log triglyceride values for the same variables as in DGI (age, age², gender, diabetes status and enrolling center, if applicable). A participant's multivariable-adjusted residual lipid concentration served as the phenotype in genotype-phenotype association analyses. A small proportion of individuals were on lipid lowering therapy (4.9% in DGI, 1.8% in the cardiovascular cohort of the Malmö Diet and Cancer Study and 3.3% in FINRISK97), and these subjects were excluded.

For genotype-lipid association analyses, we assumed an additive model of inheritance. We conducted multivariable linear regression analyses to test the null hypothesis that LDL cholesterol, HDL cholesterol or triglyceride residuals did not differ by increasing minor allele copy number. We were able to standardize the analyses across the discovery and replication cohorts in the

manner described above with two exceptions: in FINRISK97, age² was not considered and outlier individuals in the top 0.5% and the lower 0.5% of the lipid distributions were excluded, and in NORDIL and Singapore NHS98, information on lipid-lowering therapy was not available and thus not considered. Association analyses were conducted in either SAS, SPSS or PLINK³⁰.

To summarize the statistical evidence across the discovery and three replication cohorts, we conducted a fixed-effects variance-weighted meta-analysis⁷, as described in **Supplementary Methods**.

URLS. DGI association results, <http://www.broad.mit.edu/diabetes>.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We acknowledge R. Saxena, P.I. de Bakker, V. Lyssenko, M. Daly, J. Hirschhorn, S. Gabriel, H. Chen, T. Hughes, the entire Diabetes Genetics Initiative study team, and the Botnia Study team for their roles in sample collection, phenotyping, design and conduct of the DGI study. We thank our colleagues from the FUSION study and the Sardinia study for sharing data before publication. We thank J. Carlson for the isolation and preparation of the DNA samples from the Cardiovascular Cohort of the Malmö Diet and Cancer Study. We thank M. Svenson and L. Rosberg for technical assistance in Malmö and the Region Scania Competence Centrum 2 genotyping facility for help with genotyping in the NORDIL sample. We thank D. Ausiello and K. Chien for support and E. Lander for comments on the manuscript.

S.K. is supported by a Doris Duke Charitable Foundation Clinical Scientist Development Award, a charitable gift from the Fannie E. Rippel Foundation, the Donovan Family Foundation, and a career development award from the United States National Institutes of Health (NIH). C.N.-C. is supported by a Doris Duke Charitable Foundation Clinical Scientist Development Award, a career development award from the US NIH and a Burroughs Wellcome Fund Career Award for Medical Scientists. D.M.A. is a Burroughs Wellcome Fund Clinical Scholar in Translational Research and a Distinguished Clinical Scholar of the Doris Duke Charitable Foundation. V.S. is supported by the Sigrid Juselius Foundation. L.G. is supported by the Sigrid Juselius Foundation, the Finnish Diabetes Research Foundation, The Folkhalsan Research Foundation and Clinical Research Institute HUCH Ltd. His work in Malmö, Sweden was also funded by a Linné grant from the Swedish Medical Research Council. M.O.-M. is supported by the Diabetes programme at the Lund University, the Novo Nordic Foundation and the Linne grant from the Swedish Research Council. M.O.-M. and O.M. are supported by the Swedish Medical Research Council, the Swedish Heart and Lung Foundation, the Medical Faculty of Lund University, Malmö University Hospital, the Albert Pahlsson Research Foundation and the Crafoord Foundation. O.M. is supported by the Swedish Medical Society, the Ernhold Lundströms Research Foundation, the Mossfelt Foundation, the King Gustav V and Queen Victoria Foundation and the Region Skane. J.M.O. is supported by the United States Department of Agriculture, the National Heart, Lung, and Blood Institute, and the National Institute of Diabetes and Digestive and Kidney Diseases. G.B. is supported by grants from The Swedish Medical Research Council, the Region Skane, the Ernhold Lundstrom Research Foundation and the European Council (InterAct study). B.H. is supported by the Swedish Medical Research Council, the Swedish Heart and Lung Foundation, the Region Skane, the Ernhold Lundstrom Research Foundation and the Malmö University Hospital Foundation. M.-R.T. is supported by the Sigrid Juselius Foundation, the Clinical Research Institute HUCH Ltd. and the Finnish Heart Foundation. G.M.C. is a fellow of the Jane Coffin Childs Memorial Fund for Medical Research. M.J.R. is supported by a grant from the US NIH. The University of Washington School of Pharmacy Human Liver Bank is supported in part by a US NIH grant. Liver expression studies were supported in part by the University of Washington School of Pharmacy Drug Metabolism, Transporter and Pharmacogenetics Research program (funded by unrestricted gifts from Abbott, Allergan, Amgen, Bend Research, Bristol-Myers Squibb, Eli Lilly, Johnson & Johnson, Merck, Roche and Pfizer).

AUTHOR CONTRIBUTIONS

S.K., O.M. and M.O.-M. designed the study. C.G. and A.S. performed genotyping and laboratory work. N.P.B. and M.O.-M. supervised the laboratory work. L.G., N.P.B. and D.M.A. designed and conducted the DGI genome-wide association study. M.-R.T. conducted the lipoprotein and lipid phenotype measurements in the DGI samples. G.B., B.H., and O.M. collected and phenotyped the Malmö Diet and Cancer Study sample. A.S.H., E.V., P.J., V.S. and L.P. collected and phenotyped the FINRISK97 sample. B.W., T.H. and O.M. collected and

phenotyped the NORDIL sample. E.S.T., D.C. and J.M.O. conducted the replication study in the Singaporean sample. M.J.R. and G.M.C. conducted the liver expression studies. S.K., M.J.R., G.M.C., C.R., B.F.V. and M.O.-M. conducted the analyses. S.K. wrote the first draft of the paper. O.M., M.J.R., G.M.C., J.M.O., G.B., M.-R.T., C.N.-C., V.S., L.P., L.G., D.M.A. and M.O.-M. revised the manuscript for important intellectual content.

Published online at <http://www.nature.com/naturegenetics>

Reprints and permissions information is available online at <http://npg.nature.com/reprintsandpermissions>

- Namboodiri, K.K. *et al.* The Collaborative Lipid Research Clinics Family Study: biological and cultural determinants of familial resemblance for plasma lipids and lipoproteins. *Genet. Epidemiol.* **2**, 227–254 (1985).
- Kannel, W.B., Dawber, T.R., Kagan, A., Revotskie, N. & Stokes, J. III. Factors of risk in the development of coronary heart disease—six year follow-up experience. The Framingham Study. *Ann. Intern. Med.* **55**, 33–50 (1961).
- Miller, N.E. & Miller, G.J. Letter: high-density lipoprotein and atherosclerosis. *Lancet* **1**, 1033 (1975).
- Willer, C.J. *et al.* Newly identified loci that influence lipid concentrations and risk of coronary artery disease. *Nat. Genet.* advance online publication 13 January 2008; doi:10.1038/ng.76.
- Samani, N.J. *et al.* Genomewide association analysis of coronary artery disease. *N. Engl. J. Med.* **357**, 443–453 (2007).
- Diabetes Genetics Initiative of Broad Institute of Harvard and MIT, Lund University, and Novartis Institutes for BioMedical Research. *et al.* Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* **316**, 1331–1336 (2007).
- Deeks, J.J., Higgins, J.P.T. & Altman, D.G. *Cochrane Handbook for Systematic Reviews of Interventions 4.2.6 [Updated September 2006]* (eds. Higgins, J. & Green, S.) Analysing and presenting results. 126–128 (John Wiley & Sons, Chichester, UK, 2007).
- Benn, M. *et al.* Polymorphism in APOB associated with increased low-density lipoprotein levels in both genders in the general population. *J. Clin. Endocrinol. Metab.* **90**, 5797–5803 (2005).
- Sing, C.F. & Davignon, J. Role of the apolipoprotein E polymorphism in determining normal plasma lipid and lipoprotein variation. *Am. J. Hum. Genet.* **37**, 268–285 (1985).
- Kotowski, I.K. *et al.* A spectrum of PCSK9 alleles contributes to plasma levels of low-density lipoprotein cholesterol. *Am. J. Hum. Genet.* **78**, 410–422 (2006).
- Frikke-Schmidt, R., Nordestgaard, B.G., Jensen, G.B. & Tybjaerg-Hansen, A. Genetic variation in ABC transporter A1 contributes to HDL cholesterol in the general population. *J. Clin. Invest.* **114**, 1343–1353 (2004).
- Pennacchio, L.A. *et al.* An apolipoprotein influencing triglycerides in humans and mice revealed by comparative sequencing. *Science* **294**, 169–173 (2001).
- Boekholdt, S.M. *et al.* Cholesteryl ester transfer protein TaqIB variant, high-density lipoprotein cholesterol levels, cardiovascular risk, and efficacy of pravastatin treatment: individual patient meta-analysis of 13,677 subjects. *Circulation* **111**, 278–287 (2005).
- Guerra, R., Wang, J., Grundy, S.M. & Cohen, J.C. A hepatic lipase (LIPC) allele associated with high plasma concentrations of high density lipoprotein cholesterol. *Proc. Natl. Acad. Sci. USA* **94**, 4532–4537 (1997).
- Rip, J. *et al.* Lipoprotein lipase S447X: a naturally occurring gain-of-function mutation. *Arterioscler. Thromb. Vasc. Biol.* **26**, 1236–1245 (2006).
- Romeo, S. *et al.* Population-based resequencing of ANGPTL4 uncovers variations that reduce triglycerides and increase HDL. *Nat. Genet.* **39**, 513–516 (2007).
- Knoblauch, H. *et al.* Haplotypes and SNPs in 13 lipid-relevant genes explain most of the genetic variance in high-density lipoprotein and low-density lipoprotein cholesterol. *Hum. Mol. Genet.* **13**, 993–1004 (2004).
- Hutter, C.M. *et al.* Association of endothelial lipase gene (LIPG) haplotypes with high-density lipoprotein cholesterol subfractions and apolipoprotein A1 plasma levels in Japanese Americans. *Atherosclerosis* **185**, 78–86 (2006).
- Cohen, J.C. *et al.* Multiple rare alleles contribute to low plasma levels of HDL cholesterol. *Science* **305**, 869–872 (2004).
- Kingsley, D.M., Kozarsky, K.F., Hobbie, L. & Krieger, M. Reversible defects in O-linked glycosylation and LDL receptor expression in a UDP-Gal/UDP-GalNAc 4-epimerase deficient mutant. *Cell* **44**, 749–759 (1986).
- Uyeda, K. & Repa, J.J. Carbohydrate response element binding protein, ChREBP, a transcription factor coupling hepatic glucose utilization and lipid synthesis. *Cell Metab.* **4**, 107–110 (2006).
- Koishi, R. *et al.* Angptl3 regulates lipid metabolism in mice. *Nat. Genet.* **30**, 151–157 (2002).
- Nielsen, M.S., Jacobsen, C., Olivecrona, G., Gliemann, J. & Petersen, C.M. Sortilin/neurotensin receptor-3 binds and mediates degradation of lipoprotein lipase. *J. Biol. Chem.* **274**, 8832–8836 (1999).
- Shi, J. & Kandror, K.V. Sortilin is essential and sufficient for the formation of Glut4 storage vesicles in 3T3-L1 adipocytes. *Dev. Cell* **9**, 99–108 (2005).

25. Cutter, J., Tan, B.Y. & Chew, S.K. Levels of cardiovascular disease risk factors in Singapore following a national intervention programme. *Bull. World Health Organ.* **79**, 908–915 (2001).
26. Cohen, J.C., Boerwinkle, E., Mosley, T.H. Jr. & Hobbs, H.H. Sequence variations in PCSK9, low LDL, and protection against coronary heart disease. *N. Engl. J. Med.* **354**, 1264–1272 (2006).
27. Berglund, G., Elmstahl, S., Janzon, L. & Larsson, S.A. The Malmo Diet and Cancer Study. Design and feasibility. *J. Intern. Med.* **233**, 45–51 (1993).
28. Vartiainen, E. *et al.* Cardiovascular risk factor changes in Finland, 1972–1997. *Int. J. Epidemiol.* **29**, 49–56 (2000).
29. Hansson, L. *et al.* Randomised trial of effects of calcium antagonists compared with diuretics and beta-blockers on cardiovascular morbidity and mortality in hypertension: the Nordic Diltiazem (NORDIL) study. *Lancet* **356**, 359–365 (2000).
30. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analysis. *Am. J. Hum. Genet.* **81**, 559–575 (2007).