Genome-wide association analysis identifies six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol, or triglycerides in humans

Sekar Kathiresan[1,2,3], Olle Melander[4], Candace Guiducci[2], Aarti Surti[2], Noël P. Burtt[2], Mark J. Rieder[8], Gregory M. Cooper[8], Charlotta Roos[5], Benjamin F. Voight[2,18,19], Aki S. Havulinna[9], Björn Wahlstrand[10], Thomas Hedner[10], Dolores Corella[11], E. Shyong Tai[12], Jose M. Ordovas[13], Göran Berglund[6], Erkki Vartiainen[9], Pekka Jousilahti[9], Bo Hedblad[7], Marja-Riitta Taskinen[14], Christopher Newton-Cheh[1,2,3], Veikko Salomaa[9], Leena Peltonen[2,9,15,16], Leif Groop[5,17], David M. Altshuler[2,3,18,19,20], Marju Orho-Melander[5]

[1]Cardiology Division, Massachusetts General Hospital, Boston, U.S.A.
[2]Program in Medical and Population Genetics, Broad Institute of the Massachusetts Institute of Technology and Harvard University, Cambridge, U.S.A.
[3]Department of Medicine, Harvard Medical School, Boston, U.S.A.
[4]Department of Clinical Sciences, Hypertension and Cardiovascular Diseases, University Hospital Malmö, Lund University, Malmö, Sweden.
[5]Department of Diabetes and Endocrinology, University Hospital Malmö, Lund University, Malmö, Sweden.
[6]Department of Internal Medicine, University Hospital Malmö, Lund University, Malmö, Sweden.
[7]Department of Epidemiological Research, University Hospital Malmö, Lund University, Malmö, Sweden.
[8]Department of Genome Sciences, University of Washington, Seattle, U.S.A.
[9]KTL/National Public Health Institute, Helsinki, Finland.
[10]Department of Clinical Pharmacology, Sahlgrenska University Hospital, Göteborg, Sweden
[11]Department of Preventive Medicine, School of Medicine, University of Valencia, Valencia, Spain.
[12]Department of Endocrinology, Singapore General Hospital, Singapore.
[13]Nutrition and Genomics Laboratory, Jean Mayer USDA Human Nutrition Research Center on Aging at Tufts University, Boston, U.S.A.
[14]Department of Medicine, University of Helsinki, Helsinki, Finland.
[15]Department of Genetics, University of Helsinki, Helsinki, Finland.
[16]Wellcome Trust Sanger Institute, Cambridge, U.K.
[17]Department of Medicine, Helsinki University Hospital, Helsinki, Finland.
[18]Center for Human Genetic Research, Massachusetts General Hospital, Boston, U.S.A.
[19]Department of Molecular Biology, Massachusetts General Hospital, Boston, U.S.A.
[20]Department of Genetics, Harvard Medical School, Boston, U.S.A.

SUPPLEMENTARY METHODS

Informed Consent. All participants in the four studies gave written informed consent. The DGI and MDC-CC study protocols were approved by the ethics committee of Lund University and the NORDIL study protocol was approved by the ethics committee of Gothenburg University. The FINRISK97 study protocol was approved by the ethics committee of the National Public Health Institute of Finland. The Singapore NHS98 was approved with the Ethics Committee of the Ministry of Health of Singapore.

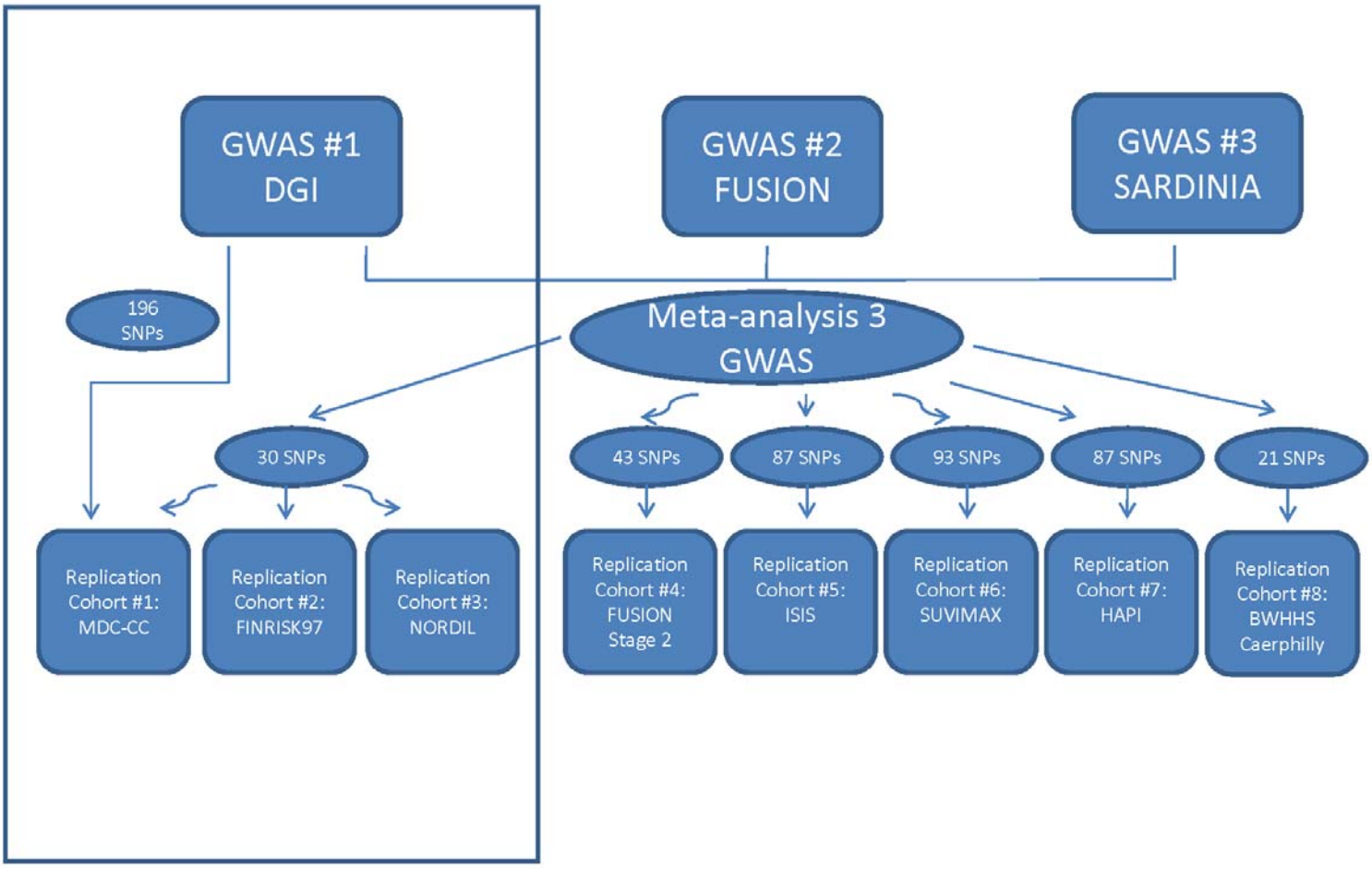
Messenger RNA expression and genotyping in human liver samples. Total RNA and DNA were extracted from 60 human liver tissue samples, primarily of European ancestry (n =56, 93%), from the University of Washington School of Pharmacy Human Liver Bank as previously described[1]. Genome wide expression analysis was performed using 250 ng of total RNA on the Illumina HumanRef-8 v.2 platform. Liver expression was assayed in duplicate, with each replicate randomized between processed batches of 32 arrays done on different days. Raw signal intensity measurements from each array were processed using the Illumina BeadStudio software v. 2.3.41 using the 'average' normalization function. Replicate data was averaged and log transformed prior to statistical analysis. Whole genome genotyping was performed on each liver sample using the Illumina HumanHap550 Beadchip platform. To assess the integrity of the samples, we evaluated the following positive control: the association of variation at *VKORC1* (SNP rs10874514) and *VKORC1* liver transcript levels. This association has been replicated by others (E. Schadt, personal communication). Using the transcript level data from the Illumina expression array, we observe an association between cis-acting genetic variation at *VKORC1* (SNP rs10874514) and *VKORC1* transcript levels ($P=2x10^{-9}$). The mRNA expression study was approved by the human subjects review committee at the University of Washington.

We analyzed the correlation between genotype and mRNA expression in human liver after log-transformation. For the index associated SNP at each validated locus, we used linear regression analyses (using the R statistical package) to relate SNP genotype with the mRNA level for genes near the index SNP, assuming an additive effect model. In secondary analyses, we excluded the 4 individuals of non-European ancestry and this did not alter the magnitude or significance of the associations; thus, we present the results for all 60 available samples.

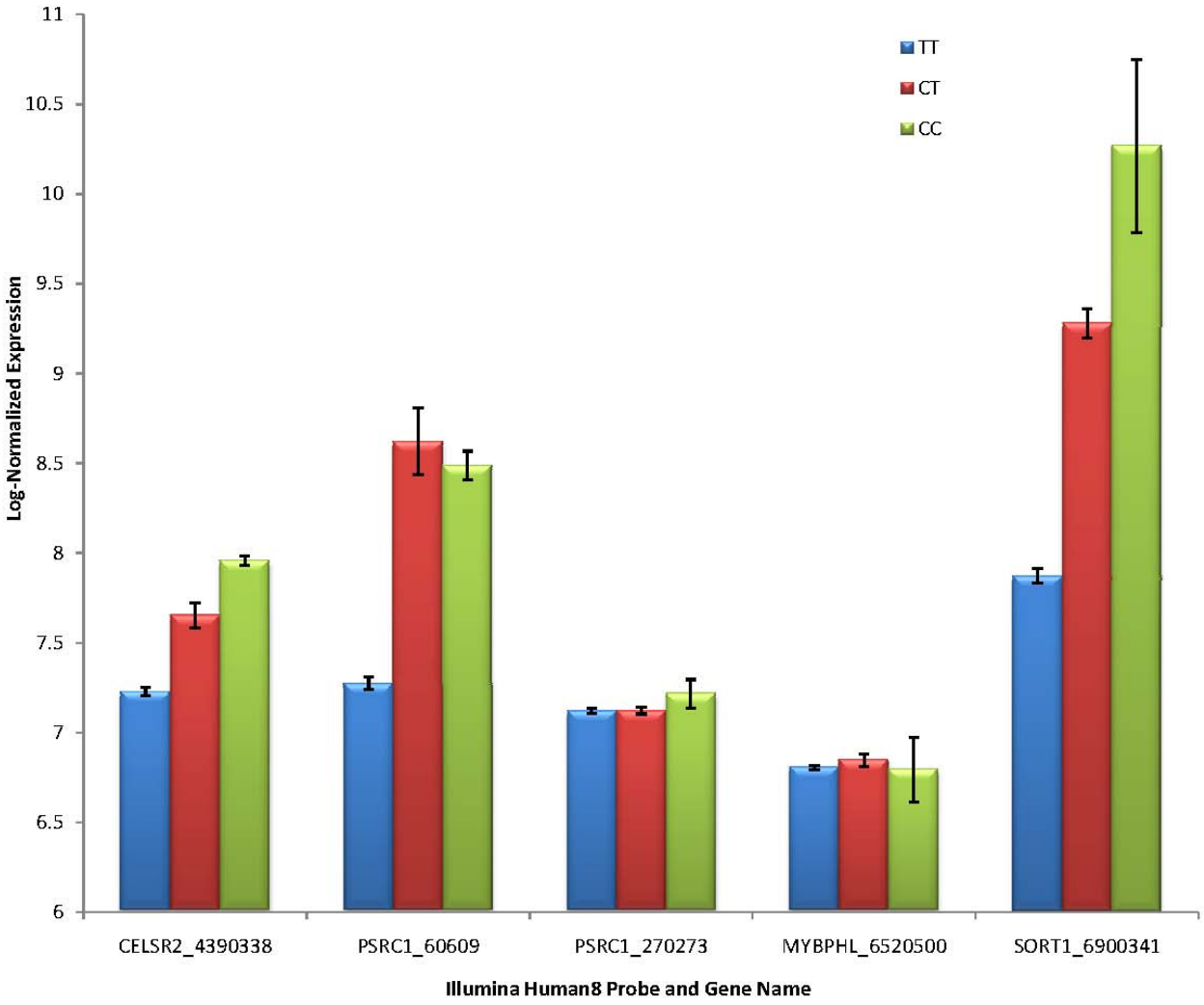
Statistical Analysis. To summarize the statistical evidence across the discovery and replication cohorts, we conducted a fixed-effects variance-weighted meta-analysis[2]. We computed a weighted average of the beta-coefficient estimates and standard errors (from the linear regression models above) for the DGI genome-wide association study and replication cohorts (MDC-CC, FINRISK97, and NORDIL), using the inverse of the variance in each cohort as weights. We set $P < 5 \times 10^{-8}$ as the threshold for statistical significance as it has been estimated that testing all common DNA sequence variants in the genome (≥5% frequency) involves the testing of ~1,000,000 independent hypotheses[3]. Association analyses were conducted in either SAS, SPSS or PLINK[4].

1. Rieder, M.J. et al. Effect of VKORC1 haplotypes on transcriptional regulation and warfarin dose. *N Engl J Med* **352**, 2285-2293 (2005).
2. Higgins, J., Green S., editors. Analysing and presenting results. in *Cochrane Handbook for Systematic Reviews of Interventions 4.2.6 [Updated September 2006]* (John Wiley & Sons, Ltd., Chichester, UK).
3. A haplotype map of the human genome. *Nature* **437**, 1299-1320 (2005).
4. 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).

Supplementary Figure 1: Overall study design. Single-nucleotide polymorphisms (SNPs) were selected for replication based on either the single Diabetes Genetics Initiative (DGI) genome-wide association study or a meta-analysis of three genome-wide association studies, namely DGI, the Finland-United States Investigation of NIDDM Genetics (FUSION), and SardiNIA Study of Aging (SardiNIA). The data presented in the current study is denoted by a box.


GWAS #1
DGI
GWAS #2
FUSION
GWAS #3
SARDINIA
196
SNPs
Meta-analysis 3
GWAS
30 SNPs
43 SNPs
87 SNPs
93 SNPs
87 SNPs
21 SNPs
Replication
Cohort #1:
MDC-CC
Replication
Cohort #2:
FINRISK97
Replication
Cohort #3:
NORDIL
Replication
Cohort #4:
FUSION
Stage 2
Replication
Cohort #5:
ISIS
Replication
Cohort #6:
SUVIMAX
Replication
Cohort #7:
HAPI
Replication
Cohort #8:
BWHHS
Caerphilly

Supplementary Figure 2. Human liver transcript level of four genes on 1p32 stratified by genotype at the LDL cholesterol-associated SNP rs646676. SNP rs646676 is associated with transcript levels of *SORT1* ($P=2x10^{-26}$), *CELSR2* ($2x10^{-12}$), and *PSRC1*_60609 ($P=1x10^{-12}$). Sample size for each genotype class was n=17, TT; n=41, CT; n=2, CC. Data is shown as mean ± standard error.


11
10.5
10
9.5
9
8.5
8
7.5
7
6.5
6
Log-Normalized Expression
TT
CT
CC
CELSR2_4390338
PSRC1_60609
PSRC1_270273
MYBPHL_6520500
SORT1_6900341
Illumina Human8 Probe and Gene Name

Supplementary Table 1. Lipid level by genotype in the Malmö Diet and Cancer Study – Cardiovascular Cohort for SNPs at newly-discovered genetic loci[*]

SNP	Locus	Nearest Gene(s)	M/M	M/m	m/m	% of Residual Variance Explained	P[‡]
LDL cholesterol							
rs646776	1p13	*CELSR2* *PSRC1* *SORT1*	TT[†] 164 ± 39 n=2876	TC 158 ± 37 n=1702	CC 154 ± 34 n=290	0.8	2×10^{-10}
rs599839	1p13	*CELSR2* *PSRC1* *SORT1*	AA 164 ± 42 n=2691	AG 157 ± 45 n=1771	GG 148 ± 39 n=296	1.2	1×10^{-14}
rs16996148	19p13	*CILP2* *PBX4*	GG 162 ± 38 n=3926	GT 160 ± 37 n=897	TT 146 ± 37 n=46	0.1	0.01
rs17321515	8q24	*TRIB1*	AA 165 ± 38 n=1220	AG 161 ± 39 n=2435	GG 159 ± 37 N=1197	0.3	1×10^{-4}
HDL cholesterol							
rs4846914	1q42	*GALNT2*	AA 55 ± 14 n=1842	AG 53 ± 14 n=2299	GG 52 ± 14 n=796	0.5	1×10^{-6}
rs17145738	7q11	*BCL7B* *TBL2* *MLXIPL*	CC 53 ± 14 n=3785	CT 55 ± 15 n=1075	TT 55 ± 16 n=81	0.2	8×10^{-4}
rs17321515	8q24	*TRIB1*	AA 53 ± 14 n=1244	AG 54± 14 n=2469	GG 55± 15 N=1208	0.2	0.002
Triglycerides							
rs17145738	7q11	*BCL7B* *TBL2* *MLXIPL*	CC 122 ± 71 n=3816	CT 113 ± 67 n=1093	TT 107 ± 57 n=82	0.6	3×10^{-8}
rs17321515	8q24	*TRIB1*	AA 126 ± 80 n=1262	AG 121 ± 70 n=2491	GG 114 ± 59 N=1218	0.4	1×10^{-5}
rs4846914	1q42	*GALNT2*	AA 118 ± 74 n=1858	AG 120 ± 68 n=2327	GG 126 ± 68 n=802	0.2	0.001
rs16996148	19p13	*CILP2* *PBX4*	GG 120 ± 72 n=4024	GT 120 ± 64 n=919	TT 106 ± 59 n=46	--	0.23
rs12130333	1p31	*ANGPTL3* *DOCK7* *ATG4C*	CC 123 ± 73 n=3070	CT 116 ± 67 n=1687	TT 112 ± 63 n=253	0.4	2×10^{-5}

[*]Data are from Replication Sample #1 – the MDC-CC – except for rs599839 where data from the NORDIL study is presented.

Plus-minus values are means (in mg/dL) ± SD. To convert values for cholesterol to millimoles per liter, multiply by 0.02586. To convert values for triglycerides to millimoles per liter, multiply by 0.01129.
SNP refers to single nucleotide polymorphism; MAF, minor allele frequency; M/M, major allele homozygote; M/m, heterozygote; m/m, minor allele homozygote
[†]Within each cell are the alleles for the SNP on the forward strand of the human genome reference sequence (from National Center for Biotechnology Information Build 35), the mean unadjusted cholesterol or triglyceride value in mg/dL plus minus standard deviation in mg/dL, and the number of individuals of that genotype class.
[‡]Association analyses were conducted using multivariable-adjusted lipid concentration (adjusted for age, age^2, sex, and diabetes status) as the phenotype.

Supplementary Table 2. New genetic loci where common SNPs are associated with multiple lipoprotein/lipid traits*

Phenotype	Locus	SNP Type	Nearest Gene(s)	Allele (Frequency) †	Meta-Analysis: GWAS & ≤3 Replications P β (SE) ‡	GWAS: DGI P β (SE)	Replication #1: MDC-CC P β (SE)	Replication #2: FINRISK 97 P β (SE)	Replication #3: NORDIL P β (SE)
19p13 near CILP2/PBX4 (rs16996148)									
LDL cholesterol	19p13	Intergenic	*CILP2* *PBX4*	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)
Triglycerides					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)
1q42 in GALNT2 (rs4846914)									
HDL cholesterol	1q42	Intronic	*GALNT2*	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)
Triglycerides					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)
7q11 near TBL2/MLXIPL (rs17145738)									
Triglycerides	7q11	Intergenic	*BCL7B* *TBL2* *MLXIPL*	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)
HDL cholesterol					4×10^{-6} 0.07 (0.02)	0.03 0.09 (0.04)	8×10^{-4} 0.10 (0.03)	0.004 0.07 (0.02)	0.47 0.02 (0.03)
8q24 near TRIB1 (rs17321515)									
Triglycerides	8q24	3' Downstream	*TRIB1*	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)
LDL cholesterol					2×10^{-7} -0.05 (0.01)	0.083 -0.05 (0.03)	1×10^{-4} -0.08 (0.02)	0.003 -0.05 (0.02)	0.12 -0.03 (0.02)
HDL Cholesterol					1×10^{-5} 0.05 (0.01)	0.41 0.02 (0.03)	0.002 0.06 (0.02)	0.03 0.04 (0.02)	0.01 0.05 (0.02)

SNP refers to single nucleotide polymorphism; GWAS, genome-wide association study; SE, standard error; -- indicates that SNP was not genotyped in sample.

*Beta-coefficient (β) represents the proportion of 1 SD change in standardized LDL cholesterol residual (mean=0, SD=1 after adjustment for age, age^2, sex, and diabetes status) per copy of the allele modeled.

†Alleles 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.

‡Variance-weighted meta-analysis performed using data from up to four samples - DGI, MDC-CC, FINRISK, and NORDIL - as described in the methods.

Supplementary Table 3. Association of SNPs from novel loci in a multi-ethnic sample – Singapore National Health Survey 98

SNP	Locus	SNP Type	Nearest Gene(s)	Phenotype	Allele† Frequency in (Chinese) (Asian Indians) (Malays)	Chinese: N P β (SE)	Asian Indians: N P β (SE)	Malays: N P β (SE)
rs646776	1p13	Intergenic	*CELSR2* *PSRC1* *SORT1*	LDL cholesterol	C (0.06) (0.07) (0.27)	2891 <0.001 -0.20 (0.05)	587 0.003 -0.19 (0.06)	781 0.004 -0.29 (0.10)
rs16996148	19p13	Intergenic	*CILP2* *PBX4*	LDL cholesterol	T (0.07) (0.14) (0.15)	2835 0.16 0.07 (0.05)	577 0.13 -0.13 (0.09)	765 0.92 0.01 (0.07)
rs4846914	1q42	Intronic	*GALNT2*	HDL cholesterol	A (0.22) (0.40) (0.31)	2844 0.06 -0.06 (0.03)	579 0.46 -0.04 (0.06)	766 0.70 0.02 (0.05)
rs17145738	7q11	Intergenic	*BCL7B* *TBL2* *MLXIPL*	Triglycerides	T (0.09) (0.06) (0.10)	2836 0.001 -0.15 (0.05)	581 0.001 -0.41 (0.13)	772 0.04 -0.18 (0.09)
rs17321515	8q24	3' Downstream	*TRIB1*	Triglycerides	G (0.54) (0.39) (0.49)	2830 0.08 -0.05 (0.03)	598 0.92 0.01 (0.06)	770 0.48 0.04 (0.05)
rs4846914	1q42	Intronic	*GALNT2*	Triglycerides	A (0.22) (0.40) (0.31)	2844 0.92 0.003 (0.003)	579 0.87 -0.01 (0.06)	766 0.35 -0.05 (0.05)
rs16996148	19p13	Intergenic	*CILP2* *PBX4*	Triglycerides	T (0.07) (0.14) (0.15)	2835 0.40 -0.04 (0.05)	577 0.14 -0.13 (0.09)	765 0.06 -0.13 (0.07)
rs12130333	1p31	Intergenic	*ANGPTL3* *DOCK7* *ATG4C*	Triglycerides	T (0.01) (0.16) (0.03)	2869 0.37 -0.11 (0.13)	587 0.30 -0.08 (0.08)	776 0.32 -0.15 (0.15)

SNP refers to single nucleotide polymorphism; GWAS, genome-wide association study; SE, standard error

*Beta-coefficient (β) represents the proportion of 1 SD change in standardized residual (mean=0, SD=1 after adjustment for age, age^2, sex, and diabetes status) per copy of the allele modeled.

†Alleles for the SNP on the forward strand of human genome reference sequence (National Center for Biotechnology Information Build 35) were modeled.