Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Arterioscler Thromb Vasc Biol. Author manuscript; available in PMC 2022 September 01.

Published in final edited form as:

Arterioscler Thromb Vasc Biol. 2021 September ; 41(9): 2494–2508. doi:10.1161/ATVBAHA.120.315391.

Genome wide association study identifies a functional *SIDT2* variant associated with HDL-C levels and premature coronary artery disease.

Paola León-Mimila[*], **Hugo Villamil-Ramírez**[*], **Luis R. Macias-Kauffer**[*], **Leonor Jacobo-Albavera**[*], **Blanca E. López-Contreras**, **Rosalinda Posadas-Sánchez**, **Carlos Posadas-Romero**, **Sandra Romero-Hidalgo**, **Sofía Morán-Ramos**, **Mayra Domínguez-Pérez**, **Marisol Olivares-Arevalo**, **Priscilla Lopez-Montoya**, **Roberto Nieto-Guerra**, **Víctor Acuña-Alonzo**, **Gastón Macín-Pérez**, **Rodrigo Barquera-Lozano**, **Blanca E. Del-Río-Navarro**, **Israel González-González**, **Francisco Campos-Pérez**, **Francisco Gómez-Pérez**, **Victor J. Valdés**, **Alicia Sampieri**, **Juan G. Reyes-García**, **Miriam del C. Carrasco-Portugal**, **Francisco J. Flores-Murrieta**, **Carlos A. Aguilar-Salinas**, **Gilberto Vargas-Alarcón**, **Diana Shih**, **Peter J. Meikle**, **Anna C. Calkin**, **Brian G. Drew**, **Luis Vaca**, **Aldons J. Lusis**, **Adriana Huertas-Vazquez**, **Teresa Villarreal-Molina**, **Samuel Canizales-Quinteros**

Unidad de Genómica de Poblaciones Aplicada a la Salud, Facultad de Química, Universidad Nacional Autónoma de México (UNAM) /Instituto Nacional de Medicina Genómica (INMEGEN), Mexico City, Mexico (P.L.-M., H.V.-R., L.R.M.-K., B.E.L.-C., S.M.-R., M.O.-A., P.L.-M., R.N.-G., S.C.-Q.), Laboratorio de Enfermedades Cardiovasculares, INMEGEN, Mexico City, Mexico (L.J.-A., M.D.-P., T.V.-M.), Departamento de Endocrinología, Instituto Nacional de Cardiología Ignacio Chávez, Mexico City, México (R.P.-S., C.P.-R.), Departamento de Genómica Computacional, INMEGEN, Mexico City, Mexico (S.R.-H.), Consejo Nacional de Ciencia y Tecnología (CONACyT), Mexico City, México (S.M.-R.), Escuela Nacional de Antropología e Historia, Mexico City, Mexico (V.A.-A., G.M.-P.), Max Planck Institute for the Science of Human History, Jena, Germany (R.B.-L.), Hospital Infantil de México Federico Gómez, Mexico City, Mexico (B.E.D.-R.-N.), Hospital General Rubén Leñero, Mexico City, Mexico (I.G.-G., F.C.-P.), Unidad de Investigación en Enfermedades Metabólicas and Departamento de Endocrinología y Metabolismo, Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, Mexico City, Mexico (F.G.-P., C.A.A.-S.), Instituto de Fisiología Celular, UNAM, Mexico City, Mexico (V.J.V., A.S., L.V.), Sección de Estudios de Posgrado e Investigación, Escuela Superior de Medicina, Instituto Politécnico Nacional, Mexico City, México (J.G.R.-G., F.J.F.-M.), Unidad de Investigación en Farmacología, Instituto Nacional de Enfermedades Respiratorias Ismael Cosío Villegas, Mexico City, Mexico (M.C.-P., F.J.F.-M.), Tecnológico de Monterrey, Escuela de Medicina y Ciencias de la Salud. Monterrey, N.L. Mexico (C.A.A.-S.), Departamento de Biología Molecular, Instituto Nacional de Cardiología Ignacio Chávez, Mexico City, Mexico (G.V.-A.), Division of

Correspondence to: Teresa Villarreal-Molina MD., PhD. Laboratorio de Enfermedades Cardiovasculares. Instituto Nacional de Medicina Genómica. Periférico Sur No. 4809, Tlalpan, 14610. México City, Mexico. Phone: +52 55 53501900 Ext. 1161, Fax: +52 55 53501999, mvillareal@inmegen.gob.mx, Samuel Canizales-Quinteros PhD. Unidad de Genómica de Poblaciones Aplicada a la Salud, Facultad de Química, UNAM/Instituto Nacional de Medicina Genómica. Periférico Sur No. 4809, Tlalpan, 14610. México City, Mexico. Phone: +52 55 53501900 Ext. 1186, Fax: +52 55 53501999, cani@unam.mx.

[*]These authors contributed equally to this work.

Disclosures

The authors declare no competing interests.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Cardiology, Department of Medicine, David Geffen School of Medicine, University of California, Los Angeles, CA, USA (D.S., A.J.L., A.H.-V.), Head Metabolomics Laboratory, Baker Heart and Diabetes Institute, Melbourne, VIC, Australia (P.J.M.), Baker Heart and Diabetes Institute, Melbourne, VIC, Australia; Central Clinical School, Monash University, Melbourne, VIC, Australia; Baker Department of Cardiometabolic Health, University of Melbourne, Parkville, VIC, Australia (A.C.C., B.G.D.).

Abstract

Objective: Low HDL-C is the most frequent dyslipidemia in Mexicans, but few studies have examined the underlying genetic basis. Our purpose was to identify genetic variants associated with HDL-C levels and cardiovascular risk in the Mexican population.

Approach and Results: A GWAS for HDL-C levels in 2,335 Mexicans, identified four loci associated with genome-wide significance: *CETP*, *ABCA1*, *LIPC* and *SIDT2*. The *SIDT2* missense Val636Ile variant was associated with HDL-C levels and was replicated in 3 independent cohorts (P=5.9×10^{-18} in the conjoint analysis). The *SIDT2*/Val636Ile variant is more frequent in Native American and derived populations than in other ethnic groups. This variant was also associated with increased ApoA1 and glycerophospholipid serum levels, decreased LDL-C and ApoB levels, and a lower risk of premature CAD. Because SIDT2 was previously identified as a protein involved in sterol transport, we tested whether the SIDT2/Ile636 protein affected this function using an *in vitro* site-directed mutagenesis approach. The SIDT2/Ile636 protein showed increased uptake of the cholesterol analog dehydroergosterol, suggesting this variant affects function. Finally, liver transcriptome data from humans and the Hybrid Mouse Diversity Panel (HMDP) are consistent with the involvement of *SIDT2* in lipid and lipoprotein metabolism.

Conclusions: This is the first genome wide association study for HDL-C levels seeking associations with coronary artery disease in the Mexican population. Our findings provide new insight into the genetic architecture of HDL-C and highlight *SIDT2* as a new player in cholesterol and lipoprotein metabolism in humans.

Keywords

Genome Wide Association Studies; High-density lipoprotein cholesterol; SIDT2; cholesterol; Coronary Artery Disease; Lipids and Cholesterol; Metabolism; Cardiovascular Disease; Race and Ethnicity; Genetic; Association Studies; Coronary Artery Disease

INTRODUCTION

Observational epidemiologic studies have reported that low plasma high density lipoprotein cholesterol (HDL-C) concentrations are an independent risk factor for cardiovascular disease.[1, 2] Heritability of HDL-C serum levels has been estimated as high as 70% in various populations, including Mexican-Americans.[3, 4] Genome wide association studies (GWAS) have successfully identified more than 150 loci associated with lipid levels mainly in European populations,[5–8] while relatively few GWAS have been performed in Mexicans.[9–12] The main HDL-C associated loci identified in Europeans are also associated with HDL-C levels in Mexicans, although novel loci have been reported in the latter group.[10] Notably, a functional variant apparently private to the Americas (*ABCA1* Arg230Cys) was found to

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

be associated with lower HDL-C levels in Mexicans.[13, 14] Although low HDL-C levels are a well-established cardiovascular risk factor, Mendelian randomization studies have shown that most genetic variants associated with this trait are not associated with cardiovascular risk, suggesting that this relationship is not necessarily causal.[15–17] In this regard, it has been postulated that pleiotropic effects of the genetic variants or HDL-C particle functionality rather than HDL-C plasma concentrations may affect cardiovascular risk.[18–21]

Low HDL-C levels are highly prevalent in the Mexican population.[22] This population group has been underrepresented in GWAS, and few lipid-associated variants have been tested for coronary artery disease (CAD) risk in Mexicans.[23] Therefore, we performed a GWAS for HDL-C levels in Mexican individuals, using a multiethnic array that includes rare and common genetic variants for Hispanic populations. We identified four loci associated with HDL-C levels with genome-wide significance, including a missense variant in the *SIDT2* gene (Val636Ile, rs17120425). We then sought to replicate these associations in two independent cohorts: the Genetics of Atherosclerotic Disease (GEA) and the Morbid Obesity Surgery (MOBES) studies, and to test their possible effect on CAD risk. Lastly, we explored the effect of the *SIDT2* gene, a member of a novel family of cholesterol transporters,[24] on lipid metabolism using existing liver transcriptome data from the Hybrid Mouse Diversity Panel, as well as the effect of the *SIDT2* Val636Ile variant on uptake of the cholesterol analog dehydroergosterol (DHE) *in vitro* using a site-directed mutagenesis approach.

MATERIALS AND METHODS

The data that support the findings of this study have not been deposited in a public repository because this is inconsistent with consent documents. The data are also part of other studies in progress, but are available from the corresponding author upon reasonable request.

Study populations

A detailed description of all cohorts is provided in the Data Supplement. The Obesity Research Study for Mexican Children (ORSMEC) and Mexican Adult cohorts were used for the discovery phase. The Genetics of Atherosclerotic Disease (GEA) cohort, the Mexican Obesity Surgery (MOBES) cohort and a group of Native American individuals were used for the replication phase.

Ethics Statement

This study was conducted according to the principles expressed in the Declaration of Helsinki and was approved by the Ethics Committees of participant institutions. All adult participants provided written informed consent prior to inclusion in the study. For children, parents or guardians of each child signed the informed consent and children assented to participate.

GWAS and quality control

Genomic DNA was isolated from peripheral white blood cells using standard methods. A total of 2,335 children and adults included in the discovery phase were genotyped

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

using the Multi-Ethnic Genotyping Array (MEGA, Illumina Inc., San Diego, CA), USA), which included >1600k SNPs. This array includes both common and rare variants in Latin American, African, European and Asian populations. Standard quality control (QC) measures were as previously described (Data supplement).[25] Identity-by-descent (IBD) was estimated using Plink v1.07.[26] A total of 624,242 SNPs remained after QC measures.

Selection of SNPs for the replication analyses

Ten SNPs at 7 loci found to be significantly associated with HDL-C levels in the discovery phase ($P<1\times10^{-5}$) were selected for replication in two independent cohorts (GEA controls and MOBES). These were the lead SNP of 4 loci (rs711975 in *STAG1,* rs6866550 in *LOC107986449,* rs983309 in *LOC157273*, and rs1077834 in *LIPC*), and 6 independent SNPs at 3 loci (rs4149310 and rs9282541 in *ABCA1;* rs948690 near *BUD13* and rs17120425 in *SIDT2*, both near the *APOA5* cluster; and rs12448528 and rs11508026 in *CETP*). SNPs were defined as independent when LD with other variants within the locus was low (r^2 and D' <0.2). Genotypes of selected SNPs were obtained from microarray data, or genotyped using KASP (LGC, U.S. http://www.lgcgroup.com) or Taqman assays (ABI Prism 7900HT Sequence Detection System, Applied Biosystems). The rs17120425 variant was also genotyped in 302 Native Mexicans using Taqman probes. Call rates exceeded 95% and no discordant genotypes were found in 10% of duplicate samples. No SNPs deviated from Hardy–Weinberg equilibrium in any group ($P>0.05$).

Fine mapping of the 11q23 locus

For fine mapping of the 11q23 locus, we increased the density of imputed SNPs using the IMPUTE2 algorithm, and phased genotypes from the 1000 Genomes Project and from NAT trios as separate reference panels. Unphased genotypes from the discovery cohorts were imputed with default parameters for effective population size, Markov chain Monte Carlo iterations and hidden Markov Models. Only genotypes imputed with a probability > 0.9 were included in the analysis. Because of special interest in rs11216230 (*SIK3* gene), which lies within the 11q23 locus and was previously associated with HDL-C in Mexicans,[10] only subjects in whom this SNP was successfully imputed were included in the analysis (1162 children and 981 adults).

To identify distinct association signals arising from multiple causal variants in the same region, we screened for independent index SNPs in the chr11q23 region using conditional and join multiple-SNP (COJO) in GCTA.[27, 28] The GCTA model made use of summary statistics from the random-effects meta-analysis in the discovery cohorts, and genotype data from these cohorts as a reference for LD across the region. We confirmed the index variant identified by GCTA (rs17120425) through conditional regression analysis using this SNP as covariate. We performed an additional conditioned analysis using the TG level lead SNP within the 11q23 region (rs964184), as covariate. Moreover, to identify the credible set of variants with a 99% probability of containing a causal variant at the *SIDT2* locus, we used Bayesian fine-mapping estimating the approximate Bayes factors of association.[29] Afterwards, variants in this locus such that their cumulative posterior probability was greater than or equal to 0.99 were selected by using unscaled variance. Unconditioned statistics from our meta-analysis were used for Bayesian fine-mapping.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Mendelian Randomization

In order to test the causal effect of HDL-C levels on CAD we performed Mendelian Randomization (MR) analyses by using HDL-C associated SNPs as an instrument. In MR analyses, genetic variants act as proxies for HDL-C levels in a manner independent of confounders.[30] We used the inverse-variance weighted (IVW) method[31] which assumes that all genetic variants satisfy the instrumental variable assumptions (including zero pleiotropy). We also performed MR-Egger regression,[32] which allows each variant to exhibit pleiotropy. SNPs found to be associated with HDL-C (n=10) were included in the MR analyses. Both methods were performed with the aid of the Mendelian Randomization R package.[33]

HEK293T cell cultures and wildtype and SIDT2/Ile636 transfection

Human embryonic kidney (HEK293T) cells were purchased from the American Type Culture Collection (ATCC, Manassas, VA). Cells were grown on 35 mm culture dishes using Dulbecco's modified Eagles medium (DMEM) (Invitrogen, Carlsbad, CA) supplemented with 10% of heat inactivated fetal bovine serum (Wisent, premium quality, Canada), penicillin-streptomycin and glutamine (Life Technologies) in an incubator with humidity control at 37 °C and 5% CO_2.

Human *SIDT2* was cloned from human cDNA CGI-40 (AF151999.1) obtained from the Riken Consortium (Japan). The product was cloned in pEGFP-N1 (Clontech, Mountain View, CA. USA). The SIDT2/Ile636 variant was produced using the QuikChange® Site-Directed Mutagenesis Kit (Stratagene, Santa Clara, CA. USA) following manufacturer's instructions. The wildtype SIDT2/Val636 variant will be referred to as SIDT2, and the isoleucine variant will be referred as Ile636/SIDT2. The primers used to change valine for isoleucine were the following: FORWARD 5´- GCGTTCTGGATCATTTTCTCCATCATT-3´ and REVERSE 5´-CGCAAGACCTAGTAAAAGAGGTAGTAA-3´. Constructs were sequenced prior to use.

The plasmids containing SIDT2-GFP and Ile636/SIDT2-GFP were transfected into HEK293T cells grown on 35 mm Petri dishes with a glass bottom (MatTek, Ashland, MA, USA). For transfection we used a mixture of 1 μg of DNA and 6 μl $CaCl_2$ reaching 60 μl of final volume with distilled H_2O. The mixture was added by dropping to HeBS buffer (50mM HEPES, 280mM NaCl, 1.5mM Na_2HPO_4, pH 7.05) and the final mixture was incubated for 30 minutes prior to replacing with DMEM. The cells were incubated overnight with the transfection mixture, which was then replaced with fresh medium to perform the assays 24 hours later.

Fluorescent cholesterol analog dehydroergosterol (DHE) uptake experiments

The naturally occurring blue fluorescent cholesterol analog DHE was purchased from Sigma (Saint Louis, MO). HEK293T cells expressing either SIDT2-GFP or SIDT2/Ile636-GFP were incubated with 5mM DHE solution, adding 300 mM methyl-β-cyclodextrin (MβCD). The solution was carefully resuspended and diluted in PBS to obtain a DHE/MβCD ratio of 1:8 (mol/mol) and 100 μL of this solution were added to the HEK293T cells. Cells were monitored using the scanning confocal microscope (FV1000, Olympus Japan). Focal plane was positioned at the middle of the cells. Confocal images were acquired every 15 seconds

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

with no averaging to reduce photobleaching. Excitation was at 300 nm using a solid-state laser and emission was collected at 535 nm.

Liver transcriptome analysis in the Hybrid Mouse Diversity Panel (HMDP) and humans

HMDP: The HMDP is a collection of approximately 100 well-characterized inbred strains of mice that can be used to analyze the genetic and environmental factors underlying complex traits such as dyslipidemia, obesity, diabetes, atherosclerosis and fatty liver disease. We analyzed the liver transcriptome of strains from this panel carrying the human cholesteryl ester transfer protein (CETP) and the human ApoE3-Leiden transgenes. At the age of about 8 weeks, these mice were placed on a "Western style" synthetic high fat diet supplemented with 1% cholesterol.[34] After 16 weeks on this diet, plasma lipid profiles were measured by colorimetric analysis as previously described[35] and animals were euthanized for the collection of liver tissue. Total RNA was isolated from the left lobe using the Qiagen (Valencia, CA) RNeasy kit (cat# 74104), as described.[36] Genome wide expression profiles were determined by hybridization to Affymetrix HT-MG_430 PM microarrays. Microarray data were filtered as previously described.[37] The ComBat method from the SVA Bioconductor package was used to remove known batch effects.[38] All animal work was conducted according to relevant national and international guidelines and was approved by the UCLA Institutional Animal Care and Use Committee (IACUC).

Humans: Total RNA was extracted from liver biopsies of 144 MOBES participants using Trizol reagent (Invitrogen). Clinical and biochemical characteristics of this subgroup of patients are described in Table SVI in the Data Supplement. RNA sequencing was performed as previously described.[39] Briefly, RNA quality was assessed using the Bioanalyzer RNA chip analysis to ensure that the RNA integrity number was >7. Complementary DNA libraries were prepared using the TruSeq RNA Stranded Total RNA Library Preparation kit (Illumina) and sequenced using an Illumina HiSeq2500 instrument, generating approximately 50 million reads/sample. After data quality control, sequencing reads were mapped to the human reference genome using TopHat software v2.0.1[40] and quantified using Cufflinks software.[41]

Statistical Methods

For the discovery phase, genome-wide association with HDL-C was tested independently in 4 groups (normal-weight and obese children and normal-weight and obese adults) under an additive linear mixed model with sex, age and BMI percentile (children) or BMI (adults) as fixed effects, and the genetic relatedness matrix as a random effect. An inverse variance method was used to perform a meta-analysis of the 4 groups. A Z-score was calculated for each SNP in each study, which summarized the magnitude and direction of the effect relative to the minor allele. The Z-score for the meta-analysis was calculated from the weighted sum of the individual study statistics.[42] Genetic relationship matrices from genome-wide data were considered for the analysis using GCTA software.[27] A *P*-value $<5\times10^{-8}$ was considered genome-wide significant, suggestive significance was defined as a *P*-value $<1\times10^{-5}$. Group heterogeneity in the meta-analysis was evaluated by I^2 and Cochrane's Q, using the R package meta. We used publicly available databases such as the GWAS Catalog (https://www.ebi.ac.uk/gwas/home) to annotate associated SNPs. SNPs

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

within a 1Mb range of the *SIDT2*/Val636Ile variant were included in a locus zoom plot.[43] The Mexican Americans from Los Angeles (MXL) 1000 Genomes sample was used to define LD patterns.

For the validation phase, linear regression under additive models was used to test for genetic associations with log-transformed lipid traits (HDL-C, LDL-C, TC and TG levels) in the GEA control, MOBES and Native American cohorts. In the GEA cohort, associations were also tested for log-transformed apoA1 and apoB levels. In the MOBES cohort, associations were also tested for lipid classes. Genetic associations with premature CAD in the GEA cohort were tested using multiple logistic regression under additive models. All tests were adjusted by age, sex and BMI. Associations were tested using the SPSS Statistics package (IBM SPSS Statistics, version 24, Chicago, IL, USA), and statistical significance was considered at $P<0.05$.

Correlations of normalized *SIDT2* liver expression with serum lipid levels and the normalized liver transcriptome were performed using biweight midcorrelation (bicor) coefficient with the R package WGCNA, a robust alternative to Pearson's correlation coefficient not sensitive to outliers.[44] Genes significantly correlated with *SIDT2* expression in liver ($P\leq1.0\times10^{-4}$) were tested for pathway enrichment analysis using Metascape and ToppGene Suite software.[45, 46] Enrichment *P* values <0.05 after FDR correction were considered significant.[47]

RESULTS

Low HDL-C levels were highly prevalent in Mexican children and adults (24.7 and 57.7%, respectively). This trait was significantly more frequent in obese as compared to lean individuals (40.9% vs 10.5% respectively in children, and 78.1% vs 32.7% respectively in adults; $P<0.001$) (Tables SI and SII in the Data Supplement). To identify loci associated with HDL-C levels, we carried out a GWAS in 2 independent cohorts of Mexican children and adults using a multi-ethnic array. Of note, the same HDL-C associated loci were found in children and adults, regardless of obesity status, and effect sizes were similar and showed the same directionality. There was no significant evidence of heterogeneity (Table SVII), and therefore a fixed-effects meta-analysis was conducted.

In total, we identified 98 variants distributed across 7 chromosomal regions associated with HDL-C levels with genome wide or suggestive significance ($P<1.0\times10^{-5}$) after adjusting for age, sex, BMI and ancestry (Figure 1 and Table SVII in the Data Supplement). Most of the 98 variants were also associated with other lipid parameters. Four loci (*CETP* in chromosome 16, *ABCA1* gene in chromosome 9, *LIPC* in chromosome 15, and a region near the *APOA5* cluster in chromosome 11) showed genome-wide significance ($P<5.0\times10^{-8}$) adjusted by age, sex and BMI (Table 1). The same loci reached genome-wide significance when adjusted only by age and sex (Figure SII). Notably, the lead SNP rs17120425 in 11q23 near the *APOA5* cluster is a missense variant (Val636Ile) within the *SIDT2* gene, which was associated with HDL-C levels for the first time in the present study ($P=4.4\times10^{-12}$). The sex stratified meta-analysis showed that the effect size of

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

SIDT2/Val636Ile on HDL-C levels was similar in females (B=3.45, P=3.6×10^{-7}) and males (B=3.22, P=2.1×10^{-5}) (Table SVIII).

Fine mapping of the 11q23 region

After imputation to increase SNP density within this region, a total of 3010 SNPs spanning 1 Mb in chromosome 11q23 were included in the analysis. Figure 2A shows the locus zoom plot for rs17120425 in the 11q23 region. We identified 96 other SNPs within this region associated with HDL-C levels (P<5×10^{-8}; Table SIX). Five of these SNPs were in LD (r^2>0.6) with rs17120425 (3 within the *SIK3* gene, 1 in *PCSK7* and 1 in *PAFAH1B2*), while two (rs948690 and rs2156121 near *BUD13*) were not in LD with the lead SNP (r^2<0.2). Of note, rs11216230 within the *SIK3* gene, previously associated with HDL-C levels and in high LD with *SIDT2*/Val636Ile (r^2=0.75), was also significantly associated with HDL-C levels (P=1.1×10^{-8}). The GCTA joint regression analysis of the 3010 genotyped and imputed SNPs in the 11q23 locus identified a single distinct association signal, represented by *SIDT2*/Val636Ile (P=1.5×10^{-11}). In addition, when conditioned on *SIDT2*/Val636Ile (rs17120425), only rs948690 and rs2156121 near *BUD13* remained associated with HDL-C levels (P=1.9×10^{-5} and 4.3×10^{-5}, respectively; Figure 2B). Bayesian fine-mapping identified a not well-resolved 99% credible set within this locus (96 variants comprised the credible set), however the index SNP (*SIDT2*/Val636Ile) showed the maximum posterior probability (PP_{max}= 0.302).

Because this region (11q23) is well known for containing gene variation associated with TG levels, we analyzed whether rs964184, robustly associated with TG levels in previous studies, affected the association of the lead SNP with HDL-C levels. This SNP was strongly associated with higher TG levels (P=1.74×10^{-22}) and significantly associated with lower HDL-C levels (P=0.003) in our study, however conditioning on rs964184 did not modify the association of rs17120425 (*SIDT2*/Val636Ile) with HDL-C levels (P=7.3×10^{-11}, Figure SIII).

Notably, minor allele frequencies of *ABCA1*/Arg230Cys and *SIDT2*/Val636Ile variants are highest in populations from the Americas (Table SX). We then genotyped the *SIDT2*/Val636Ile variant in two independent Native American populations from central Mexico (Totonacs and Nahuas), which was also associated with higher HDL-C levels in these indigenous groups (B=2.81 mg/dL; P=0.027; Table SXI). Moreover, the rs17120425 “A” allele was significantly more frequent in Native Mexicans (15%) than in Mexican Mestizos (10.3%, P=0.001), and local ancestry analysis revealed that this allele was found in a block of Native American origin in 98% of individuals. However, according to EHH analysis, LD extension break down was similar in the ancestral and derived rs17120425 alleles, with no evidence of positive selection (Figure SIV, Methods supplement).

Replication of associations with HDL-C Levels and other lipid traits in independent cohorts

We then sought to replicate associations with HDL-C levels in GEA controls. Nine of the 10 variants associated with HDL-C levels in the discovery phase replicated in GEA controls (P<0.02, Table 2). Altogether, these nine variants explained ~25% of HDL-C level

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

variation ($P_{\text{Genetic Risk Score}}=1.6\times10^{-21}$). Table 2 shows associations of these SNPs with other lipid traits in the GEA cohort. *SIDT2*/Val636Ile and rs948690 (near *BUD13*) were both associated with higher HDL-C and lower ApoB levels. SNP rs948690, but not *SIDT2*/Val636Ile, was also significantly associated with lower TG levels. In the MOBES cohort, rs11508026 and rs12448528 (*CETP*, $P=8.8\times10^{-5}$ and 0.013, respectively), rs9282541 (*ABCA1,* $P=4.8\times10^{-4}$) and *SIDT2*/Val636Ile ($P=0.005$) were significantly associated with HDL-C levels (Table 3). The association of *SIDT2*/Val636Ile with HDL-C levels was highly significant in the conjoint analysis including the discovery phase and replication cohorts (B=3.25, $P=5.9\times10^{-18}$) (Table SVIII).

The SIDT2 Val636Ile variant is associated with serum glycerophospholipid levels in the MOBES cohort.

We then sought associations of the 4 HDL-C associated SNPs in the MOBES cohort with lipid classes known to be the main components of HDL-C lipoprotein particles (19 classes of cholesterol esters, glycerophospholipids and TG).[48, 49] *SIDT2/Val636Ile* was significantly associated with higher glycerophospholipid serum levels ($P<0.05$) (Figure 3), particularly with total phosphatidylcholine (PC), phosphatidylethanolamine (PE), phosphatidylglycerol (PG), phosphatidylinositol (PI) and phosphatidylserine (PS) as adjusted by age, sex, BMI and lipid lowering treatment. In contrast, *ABCA1* (rs9282541) and *CETP* variants (rs12448528 and rs11508026) were not significantly associated with any of these lipid classes.

Association with premature CAD and Mendelian randomization

We then tested whether HDL-C variants were also associated with premature CAD in the GEA cohort. Five of the ten variants were significantly associated with CAD risk: rs7119175 (*STAG1*), rs948690 (near *BUD13*) and *SIDT2*/Val636Ile were associated with higher HDL-C levels and lower CAD risk, while rs9282541 (*ABCA1)* and rs1077834/*LIPC* were associated with lower HDL-C levels and lower CAD risk (Table 4). Mendelian randomization analyses were performed including the 10 variants significantly associated with HDL-C levels. There was no evidence of a causal effect of HDL-C levels on CAD (IVW $P=0.585$ and MR-Egger $P= 0.322$; Table SXII).

Fluorescent cholesterol analog uptake is enhanced in cells expressing Ile636/SIDT2.

It was previously suggested that the transmembrane conserved cholesterol binding (CRAC) domain of murine Sidt2 associates with the cholesterol analogue dehydroergosterol in HEK293 cells.[24] Because the *SIDT2*/Val636Ile variant is near the transmembrane CRAC domain in the human SIDT2 protein, we evaluated the effect of this variant on DHE uptake in HEK293T cells. Interestingly, DHE uptake was enhanced in cells expressing the Ile636/SIDT2 protein as compared to cells expressing wildtype SIDT2, reaching the highest difference approximately 1.5 minutes after adding DHE to the culture (6.85 ± 0.42 AU vs 9.40 ± 0.73 AU, $P<0.01$) (Figure 4, Video SI in the Data Supplement).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

SIDT2 liver expression correlates with the expression of genes involved in lipid and lipoprotein metabolism

Because unidentified variants in LD with *SIDT2*/Val636Ile may affect expression of this gene, we tested for differences in hepatic *SIDT2* expression levels according to genotype in the MOBES cohort, but the differences were not significant (P=0.486). Moreover, human *SIDT2* liver expression showed no significant correlation with lipid traits including HDL-C, TC, or TG levels (Figure 5A). In contrast, in HMDP mice fed with an atherogenic diet, hepatic *Sidt2* expression correlated positively with HDL-C levels (r=0.312, P=0.002), and negatively with total cholesterol and TG levels (r=−0.381, $P=1.2\times10^{-4}$; r=−0.304, P=0.002, respectively) (Figure 5B).

To analyze correlations between *SIDT2* expression and the liver transcriptome in mice, we used data from the genetically diverse HMDP where environmental factors are controlled. A total of 2,240 genes showed significant correlations with *Sidt2* expression, and the most significantly enriched pathway was lipid and lipoprotein metabolism ($P_{FDR}=1.5\times10^{-7}$). In human liver tissue *SIDT2* expression correlated with the expression of 2,782 genes. Consistent with observations in mice, the lipid and lipoprotein metabolism pathways were among the most significantly enriched. *SIDT2* expression correlated with the expression of 227 genes that were shared by both mice and humans, and as expected metabolism of lipids and lipoproteins was the most significantly enriched pathway (Figure 5C–D).

DISCUSSION

Low HDL-C levels are the most common dyslipidemia in Mexicans, both in adults and children.[22, 50] Consistently, low HDL-C levels were highly prevalent in children and adults of the present study. Moreover, the prevalence of low HDL-C levels was significantly higher in obese than in lean individuals, in line with the high prevalence of metabolic syndrome observed in the Mexican population.[51]

Here, using a GWAS for HDL-C levels in different cohorts of the Mexican population, we identified four loci genome-wide associated with HDL-C levels at chromosomes 9 (*ABCA1*), 11 (near the *APOA5* cluster), 15 (*LIPC)* and 16 (*CETP*). Of note, the effect and direction of the associations were consistent in the 4 study groups (normal weight and obese children and adults). Variants in *ABCA1, LIPC* and *CETP* have been robustly associated with HDL-C levels across populations.[6, 11, 48] Of note, the signal in *ABCA1* is the Arg230Cys variant (rs9282541) previously reported as associated with lower HDL-C levels by our group, apparently private to Native American and derived populations.[13, 14] The signal in chromosome 11 near the *APOA5* cluster corresponds to a missense variant within the *SIDT2* gene (Val636Ile, rs17120425), associated with higher HDL-C levels. This association was replicated in 3 independent cohorts including Native Americans from Mexico. This is relevant because Native Americans live in rural areas, while the GEA and MOBES cohorts are urban populations, known to have different dietary habits, which may affect HDL-C levels.[52]

To our knowledge, this is the first time that the *SIDT2*/Val636Ile variant has been associated with HDL-C levels in a GWAS. Although intronic *SIDT2* variants (rs2269399, rs7107152,

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

rs1242229 and rs1784042) have been previously associated with lipid traits and the metabolic syndrome in multi-ethnic cohorts, mainly in Koreans,[53, 54] none of these variants are in LD with *SIDT2*/Val636Ile ($r^2<0.2$). It is likely that previous GWAS in Mexicans failed to identify this SNP as associated with HDL-C because *SIDT2*/Val636Ile was not included in the microarray platforms used in these studies, and the paucity of Native American references for imputation.[9–12] Although the *SIDT2*/Val636Ile variant is not private to the Americas, it is more frequent in Native Americans (15%) and Hispanics (6%), than other ethnic groups (internationalgenome.org). Local ancestry analyses revealed that the derived "A" allele was of Native American origin in most individuals. However, the extended haplotype homozygosity (EHH) analysis of *SIDT2*/Val636Ile did not show evidence of recent positive selection.

The 11q23 region contains several lipid-associated genes in close proximity, and fine-mapping showed a cluster of HDL-C associated variants distributed in several genes. It was thus necessary to define whether these signals are independent of *SIDT2*/Val636Ile. Notably, conditional and joint multiple-SNP analysis of this region identified a single distinct signal associated with HDL-C levels represented by *SIDT2*/Val636Ile, suggesting there are no other relevant independent signals associated with HDL-C levels within this locus. A previous GWAS in the Mexican population identified two intronic *SIK3* gene variants within the 11q23 region, rs139961185 associated with TG levels and rs11216230 with higher HDL-C levels.[10] The latter association was replicated in this study and in an independent Hispanic population.[12] After conditioning the analysis on *SIDT2*/Val636Ile, the association of rs11216230 with HDL-C levels lost significance (P=0.500), indicating that at least in our study, this variant is not independent of *SIDT2*/Val636Ile. Moreover, conditioned analysis on rs964184 (*APOA5* cluster), the lead SNP for TG levels in several populations,[9–11] also associated with HDL-C levels in the present study, did not modify the association of *SIDT2*/Val636Ile with HDL-C levels. This suggests that the association of *SIDT2*/Val636Ile with HDL-C levels is independent of the APOA5 locus. The credible set of variants obtained by Bayesian fine-mapping was not well-resolved, however the index SNP (*SIDT2*/Val636Ile) showed the maximum posterior probability (PP_{max}=0.302). Thus, fine mapping results suggest that, although *SIDT2*/Val636Ile can be considered as potentially casual, there may be other yet unidentified variants in this locus that influence serum HDL-C levels.

In Mendelian randomization analyses, genetic variants act as proxies for HDL-C levels in a manner independent of confounders to analyze the causality of HDL-C levels on coronary artery disease. Our MR analysis is consistent with previous studies suggesting that higher HDL-C levels are not causally protective against coronary heart disease.[15–17] This suggests that the effect of individual variants on CAD risk may be mediated by pleiotropic effects on other cardiovascular risk factors, or on HDL-C composition and functionality.[18–21, 55] We thus explored whether *SIDT2*/Val636Ile affects other cardiovascular risk parameters in addition to HDL-C levels. Notably, this variant was also associated with higher ApoA1 levels, and lower LDL-C and ApoB serum levels in the GEA cohort. APOA1 is the major protein component of HDL-C particles, and a Mendelian randomization analysis in Finnish individuals reported that ApoA1 was not associated with risk of CAD.[56] A multivariable Mendelian randomization study examining serum lipid and apolipoprotein levels reported

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

that only ApoB retained a robust relationship with the risk of CAD,[57] and recent Mendelian randomization studies suggest that ApoB is the primary lipid determinant of cardiovascular disease risk.[58, 59] Thus, the association of *SIDT2*/Val636Ile with decreased cardiovascular risk observed in the GEA study could be mediated by its effect on ApoB levels. However, the association with premature CAD had suggestive significance (P=0.039) and needs to be confirmed in independent cohorts.

HDL lipidome composition has been associated with HDL-C functional properties,[60, 61] Notably, of the 4 main variants associated with HDL-C levels in the MOBES cohort, only *SIDT2*/Val636Ile was associated with lipid species, specifically with higher serum concentrations of several glycerophospholipid classes including PE, PG, PC, PI and PS. It has been reported that HDL-C particles enriched in phospholipids can increase HDL-C stability,[62] while decreased levels of phospholipids in HDL-C were found to impair cholesterol efflux and decrease the cardiovascular protective effects of HDLs.[49, 61] Particularly, recent studies indicate that phosphatidylserine, a minor component of the monolayer surface of HDL-C, is enriched in small, dense HDL-C particles, which display potent anti-atherosclerotic activities.[62, 63] Although we measured phospholipids in serum and not directly in HDL-C particles, a limitation of the study, the association of *SIDT2*/Val636Ile with higher phospholipid levels is consistent with the lower cardiovascular risk conferred by this variant.

The SIDT2 protein is found mainly in lysosome membranes,[64] is a lysosomal nucleic acid transporter, and is expressed in several tissues, including the liver.[65–67] The mammalian SIDT2 protein has high homology to the *C. elegans* cholesterol uptake protein-1 (CUP-1).[68] SIDT2 has been identified as a sterol-interacting protein[69] and more recently as a cholesterol-binding protein.[24] Moreover, SIDT2 is predicted to contain two CRAC domains (Cholesterol Recognition/interaction Amino Acid Consensus), found in a broad range of proteins involved in cholesterol transport, metabolism and regulation.[24, 70] Specifically, the transmembrane CRAC domain from human SIDT1 and mouse SIDT2 appears to bind cholesterol.[24] The Val636Ile variant and the CRAC domain are located within the same transmembrane segment, and this variant is 19 amino acids upstream the tyrosine CRAC domain residue, suggested to interact with the cholesterol OH-polar group.[24] It is unknown if the Val636Ile variant modifies the interactions of the CRAC domain with cholesterol, thus affecting circulating lipid levels.

In the present study, HEK293T cells expressing the Ile636/SIDT2 protein showed higher cholesterol analog DHE uptake than those expressing the wildtype protein. This increased uptake observed *in vitro*, may affect circulating levels of cholesterol-rich lipoproteins. Although the mechanism by which this variant increases HDL-C serum levels is unknown, ABCA1-mediated cholesterol efflux and HDL-C formation are primarily dependent on autophagy for cholesterol source.[71] *Sidt2*$^{-/-}$ deficient mice show blocked autophagosome maturation, apparently by altering the fusion of autophagosomes with endosomes/lysosomes, the terminal stage of autophagosome maturation, which affects lipid metabolism. [72] It has been suggested that autophagy might simultaneously regulate the uptake and use of cholesterol in cells.[73] It is thus tempting to speculate that the putative gain of function of the Ile636 SIDT2 protein would enhance autophagosome-lysosome

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

fusion, increasing both cholesterol uptake and autophagy-mediated cholesterol efflux which is primarily ABCA1 dependent, thus elevating HDL-C levels. Further studies are required to establish how this variant affects lipid metabolism in humans.

Sidt2 knockout mice show a wide range of metabolic phenotypes including impaired glucose tolerance likely due to compromised NAADP-involved insulin secretion.[72, 74–77] Meng *et al.*,[77] showed that *Sidt2* deficient mice present significantly increased serum total cholesterol, TG and LDL-C levels, and significantly lower HDL-C serum levels as compared to $Sidt2^{+/+}$ mice. These findings are consistent with the inverse correlation between hepatic *Sidt2* expression and total cholesterol and triglycerides levels, and the direct correlation of *Sidt2* expression with HDL-C levels observed in the HMDP. Moreover, $Sidt2^{-/-}$ mice not only have altered lipid serum levels, but also showed impaired liver function and liver steatosis.[72, 76, 77] Consistently, *Sidt2* expression correlated with increased fat liver content in HMDP mice (data not shown). In contrast, in the MOBES cohort *SIDT2* liver expression did not correlate with serum lipid levels or liver steatosis. This discrepancy may be explained by differences in the design of the HMDP and the MOBES cohort. Firstly, HMDP mice had atherosclerotic-promoting transgenes (apoE-Leiden and human CETP) and were fed a "Western" diet. While this provides a unique model to study lipid metabolism, it may also introduce a bias towards certain mechanisms. Moreover, liver biopsies were performed in MOBES participants during bariatric surgery and the patients had previously been submitted to a ketogenic diet. Several studies have reported that diet regulates the hepatic expression of genes involved in lipid metabolism.[78, 79] Whether dietary components affect *SIDT2* expression and its correlation with metabolic traits, requires further study.

Because correlations of the expression of a single gene with metabolic traits may overlook common disease mechanisms among species,[80] we performed pathway enrichment analyses. In this regard, *SIDT2* expression correlations with the liver transcriptome revealed that the most significantly enriched pathways were lipid and lipoprotein metabolism in both species. Altogether, these findings support a role of SIDT2 in lipid metabolism. Further studies are required to better understand the mechanisms by which SIDT2 participates in cholesterol and lipoprotein metabolism at the cellular and systemic levels, and its role in cardiovascular risk.

In conclusion, this is the first study assessing genetic variants contributing to HDL-C levels and coronary artery disease in the Mexican population. Our GWAS revealed for the first time that the *SIDT2*/Val636Ile variant is associated with increased HDL-C and phospholipid levels and decreased risk of CAD. We also provide evidence that the *SIDT2*/Val636Ile variant is functional, increasing the uptake of a cholesterol analog *in vitro*. Our data support a role of SIDT2 in cholesterol and lipid metabolism. The mechanisms by which this protein affects lipid and metabolic parameters in humans require further investigation.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Acknowledgements

We thank Luz E. Guillén for her help with biochemical measurements.

Sources of Funding

This work was supported by grants FOSISS-289699, FOSISS-112547, NIH HL147883; and partially supported by PEI-230129 and PEI-212940 from Mexican National Council for Science and Technology (CONACyT).

Abbreviations

ApoA1	Apolipoprotein A1
ApoB	Apolipoprotein B
CAD	Coronary artery disease
GEA	Genetics of Atherosclerotic Disease
GWAS	Genome wide association studies
HDL-C	High density lipoprotein cholesterol
HMDP	Hybrid Mouse Diversity Panel
LDL-C	Low density lipoprotein cholesterol
MOBES	Morbid Obesity Surgery
TC	Total cholesterol
TG	Triglycerides

REFERENCES

1. Boden WE. High-density lipoprotein cholesterol as an independent risk factor in cardiovascular disease: assessing the data from Framingham to the Veterans Affairs High--Density Lipoprotein Intervention Trial. Am J Cardiol. 2000;86:19L–22L
2. Barter PJ, Puranik R, Rye KA. New insights into the role of HDL as an anti-inflammatory agent in the prevention of cardiovascular disease. Curr Cardiol Rep. 2007;9:493–498 [PubMed: 17999875]
3. Brunham LR, Hayden MR. Human genetics of HDL: Insight into particle metabolism and function. Prog Lipid Res. 2015;58:14–25 [PubMed: 25601427]
4. Gao C, Tabb KL, Dimitrov LM, Taylor KD, Wang N, Guo X, Long J, Rotter JI, Watanabe RM, Curran JE, Blangero J, Langefeld CD, Bowden DW, Palmer ND. Exome Sequencing Identifies Genetic Variants Associated with Circulating Lipid Levels in Mexican Americans: The Insulin Resistance Atherosclerosis Family Study (IRASFS). Sci Rep. 2018;8:5603 [PubMed: 29618726]
5. Teslovich TM, Musunuru K, Smith AV, et al. Biological, clinical and population relevance of 95 loci for blood lipids. Nature. 2010;466:707–713 [PubMed: 20686565]
6. Willer CJ, Schmidt EM, Sengupta S, et al. Discovery and refinement of loci associated with lipid levels. Nat Genet. 2013;45:1274–1283 [PubMed: 24097068]
7. Wu Y, Waite LL, Jackson AU, et al. Trans-ethnic fine-mapping of lipid loci identifies population-specific signals and allelic heterogeneity that increases the trait variance explained. PLoS Genet. 2013;9:e1003379
8. Klarin D, Damrauer SM, Cho K, et al. Genetics of blood lipids among ~300,000 multi-ethnic participants of the Million Veteran Program. Nat Genet. 2018;50:1514–1523 [PubMed: 30275531]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

9. Weissglas-Volkov D, Aguilar-Salinas CA, Nikkola E, et al. Genomic study in Mexicans identifies a new locus for triglycerides and refines European lipid loci. J Med Genet. 2013;50:298–308 [PubMed: 23505323]
10. Ko A, Cantor RM, Weissglas-Volkov D, et al. Amerindian-specific regions under positive selection harbour new lipid variants in Latinos. Nat Commun. 2014;5:3983 [PubMed: 24886709]
11. Below JE, Parra EJ, Gamazon ER, et al. Meta-analysis of lipid-traits in Hispanics identifies novel loci, population-specific effects, and tissue-specific enrichment of eQTLs. Sci Rep. 2016;6:19429 [PubMed: 26780889]
12. Graff M, Emery LS, Justice AE, et al. Genetic architecture of lipid traits in the Hispanic community health study/study of Latinos. Lipids Health Dis. 2017;16:200 [PubMed: 29025430]
13. Villarreal-Molina MT, Aguilar-Salinas CA, Rodriguez-Cruz M, Riano D, Villalobos-Comparan M, Coral-Vazquez R, Menjivar M, Yescas-Gomez P, Konigsoerg-Fainstein M, Romero-Hidalgo S, Tusie-Luna MT, Canizales-Quinteros S, Metabolic Study G. The ATP-binding cassette transporter A1 R230C variant affects HDL cholesterol levels and BMI in the Mexican population: association with obesity and obesity-related comorbidities. Diabetes. 2007;56:1881–1887 [PubMed: 17287470]
14. Acuna-Alonzo V, Flores-Dorantes T, Kruit JK, et al. A functional ABCA1 gene variant is associated with low HDL-cholesterol levels and shows evidence of positive selection in Native Americans. Hum Mol Genet. 2010;19:2877–2885 [PubMed: 20418488]
15. Voight BF, Peloso GM, Orho-Melander M, et al. Plasma HDL cholesterol and risk of myocardial infarction: a mendelian randomisation study. Lancet. 2012;380:572–580 [PubMed: 22607825]
16. Holmes MV, Asselbergs FW, Palmer TM, et al. Mendelian randomization of blood lipids for coronary heart disease. Eur Heart J. 2015;36:539–550 [PubMed: 24474739]
17. Kawashiri MA, Tada H, Nomura A, Yamagishi M. Mendelian randomization: Its impact on cardiovascular disease. J Cardiol. 2018;72:307–313 [PubMed: 29801689]
18. Khera AV, Cuchel M, de la Llera-Moya M, Rodrigues A, Burke MF, Jafri K, French BC, Phillips JA, Mucksavage ML, Wilensky RL, Mohler ER, Rothblat GH, Rader DJ. Cholesterol efflux capacity, high-density lipoprotein function, and atherosclerosis. N Engl J Med. 2011;364:127–135 [PubMed: 21226578]
19. Solovieff N, Cotsapas C, Lee PH, Purcell SM, Smoller JW. Pleiotropy in complex traits: challenges and strategies. Nat Rev Genet. 2013;14:483–495 [PubMed: 23752797]
20. Kosmas CE, Christodoulidis G, Cheng JW, Vittorio TJ, Lerakis S. High-density lipoprotein functionality in coronary artery disease. Am J Med Sci. 2014;347:504–508 [PubMed: 24603157]
21. Burgess S, Foley CN, Allara E, Staley JR, Howson JMM. A robust and efficient method for Mendelian randomization with hundreds of genetic variants. Nat Commun. 2020;11:376 [PubMed: 31953392]
22. Aguilar-Salinas CA, Olaiz G, Valles V, Torres JM, Gomez Perez FJ, Rull JA, Rojas R, Franco A, Sepulveda J. High prevalence of low HDL cholesterol concentrations and mixed hyperlipidemia in a Mexican nationwide survey. J Lipid Res. 2001;42:1298–1307 [PubMed: 11483632]
23. Villarreal-Molina T, Posadas-Romero C, Romero-Hidalgo S, et al. The ABCA1 gene R230C variant is associated with decreased risk of premature coronary artery disease: the genetics of atherosclerotic disease (GEA) study. PLoS One. 2012;7:e49285
24. Mendez-Acevedo KM, Valdes VJ, Asanov A, Vaca L. A novel family of mammalian transmembrane proteins involved in cholesterol transport. Sci Rep. 2017;7:7450 [PubMed: 28785058]
25. Macias-Kauffer LR, Villamil-Ramirez H, Leon-Mimila P, et al. Genetic contributors to serum uric acid levels in Mexicans and their effect on premature coronary artery disease. Int J Cardiol. 2019;279:168–173 [PubMed: 30305239]
26. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81:559–575 [PubMed: 17701901]
27. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet. 2011;88:76–82 [PubMed: 21167468]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

28. Yang J, Ferreira T, Morris AP, et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. Nat Genet. 2012;44:369–375, S361–363 [PubMed: 22426310]
29. Sakornsakolpat P, Prokopenko D, Lamontagne M, et al. Genetic landscape of chronic obstructive pulmonary disease identifies heterogeneous cell-type and phenotype associations. Nat Genet. 2019;51:494–505 [PubMed: 30804561]
30. Corbin LJ, Richmond RC, Wade KH, Burgess S, Bowden J, Smith GD, Timpson NJ. BMI as a Modifiable Risk Factor for Type 2 Diabetes: Refining and Understanding Causal Estimates Using Mendelian Randomization. Diabetes. 2016;65:3002–3007 [PubMed: 27402723]
31. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. Genet Epidemiol. 2013;37:658–665 [PubMed: 24114802]
32. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. Int J Epidemiol. 2015;44:512–525 [PubMed: 26050253]
33. Yavorska OO, Burgess S. MendelianRandomization: an R package for performing Mendelian randomization analyses using summarized data. Int J Epidemiol. 2017;46:1734–1739 [PubMed: 28398548]
34. Bennett BJ, Davis RC, Civelek M, et al. Genetic Architecture of Atherosclerosis in Mice: A Systems Genetics Analysis of Common Inbred Strains. PLoS Genet. 2015;11:e1005711
35. Hedrick CC, Castellani LW, Warden CH, Puppione DL, Lusis AJ. Influence of mouse apolipoprotein A-II on plasma lipoproteins in transgenic mice. J Biol Chem. 1993;268:20676–20682 [PubMed: 8376417]
36. Ghazalpour A, Bennett B, Petyuk VA, et al. Comparative analysis of proteome and transcriptome variation in mouse. PLoS Genet. 2011;7:e1001393
37. Bennett BJ, Farber CR, Orozco L, et al. A high-resolution association mapping panel for the dissection of complex traits in mice. Genome Res. 2010;20:281–290 [PubMed: 20054062]
38. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics. 2012;28:882–883 [PubMed: 22257669]
39. Hui ST, Kurt Z, Tuominen I, et al. The Genetic Architecture of Diet-Induced Hepatic Fibrosis in Mice. Hepatology. 2018;68:2182–2196 [PubMed: 29907965]
40. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 2013;14:R36 [PubMed: 23618408]
41. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol. 2010;28:511–515 [PubMed: 20436464]
42. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010;26:2190–2191 [PubMed: 20616382]
43. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, Boehnke M, Abecasis GR, Willer CJ. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics. 2010;26:2336–2337 [PubMed: 20634204]
44. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9:559 [PubMed: 19114008]
45. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, Benner C, Chanda SK. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat Commun. 2019;10:1523 [PubMed: 30944313]
46. Chen J, Bardes EE, Aronow BJ, Jegga AG. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. Nucleic Acids Res. 2009;37:W305–311 [PubMed: 19465376]
47. Reimand J, Isserlin R, Voisin V, Kucera M, Tannus-Lopes C, Rostamianfar A, Wadi L, Meyer M, Wong J, Xu C, Merico D, Bader GD. Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, Cytoscape and EnrichmentMap. Nat Protoc. 2019;14:482–517 [PubMed: 30664679]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

48. Kontush A, Lhomme M, Chapman MJ. Unraveling the complexities of the HDL lipidome. J Lipid Res. 2013;54:2950–2963 [PubMed: 23543772]
49. Ding M, Rexrode KM. A Review of Lipidomics of Cardiovascular Disease Highlights the Importance of Isolating Lipoproteins. Metabolites. 2020;10 [PubMed: 33375435]
50. Hernandez-Alcaraz C, Aguilar-Salinas CA, Mendoza-Herrera K, Pedroza-Tobias A, Villalpando S, Shamah-Levy T, Rivera-Dommarco J, Hernandez-Avila M, Barquera S. Dyslipidemia prevalence, awareness, treatment and control in Mexico: results of the Ensanut 2012. Salud Publica Mex. 2020;62:137–146 [PubMed: 32237556]
51. Rojas R, Aguilar-Salinas CA, Jimenez-Corona A, Shamah-Levy T, Rauda J, Avila-Burgos L, Villalpando S, Ponce EL. Metabolic syndrome in Mexican adults: results from the National Health and Nutrition Survey 2006. Salud Publica Mex. 2010;52 Suppl 1:S11–18 [PubMed: 20585723]
52. Ponce X, Rodriguez-Ramirez S, Mundo-Rosas V, Shamah T, Barquera S, Gonzalez de Cossio T. Dietary quality indices vary with sociodemographic variables and anthropometric status among Mexican adults: a cross-sectional study. Results from the 2006 National Health and Nutrition Survey. Public Health Nutr. 2014;17:1717–1728 [PubMed: 24124890]
53. Gombojav B, Lee SJ, Kho M, Song YM, Lee K, Sung J. Multiple susceptibility loci at chromosome 11q23.3 are associated with plasma triglyceride in East Asians. J Lipid Res. 2016;57:318–324 [PubMed: 26634697]
54. Moon S, Lee Y, Won S, Lee J. Multiple genotype-phenotype association study reveals intronic variant pair on SIDT2 associated with metabolic syndrome in a Korean population. Hum Genomics. 2018;12:48 [PubMed: 30382898]
55. Prats-Uribe A, Sayols-Baixeras S, Fernandez-Sanles A, Subirana I, Carreras-Torres R, Vilahur G, Civeira F, Marrugat J, Fito M, Hernaez A, Elosua R. High-density lipoprotein characteristics and coronary artery disease: a Mendelian randomization study. Metabolism. 2020;112:154351
56. Karjalainen MK, Holmes MV, Wang Q, Anufrieva O, Kahonen M, Lehtimaki T, Havulinna AS, Kristiansson K, Salomaa V, Perola M, Viikari JS, Raitakari OT, Jarvelin MR, Ala-Korpela M, Kettunen J. Apolipoprotein A-I concentrations and risk of coronary artery disease: A Mendelian randomization study. Atherosclerosis. 2020;299:56–63 [PubMed: 32113648]
57. Richardson TG, Sanderson E, Palmer TM, Ala-Korpela M, Ference BA, Davey Smith G, Holmes MV. Evaluating the relationship between circulating lipoprotein lipids and apolipoproteins with risk of coronary heart disease: A multivariable Mendelian randomisation analysis. PLoS Med. 2020;17:e1003062
58. Ference BA, Kastelein JJP, Ray KK, et al. Association of Triglyceride-Lowering LPL Variants and LDL-C-Lowering LDLR Variants With Risk of Coronary Heart Disease. JAMA. 2019;321:364–373 [PubMed: 30694319]
59. Zuber V, Gill D, Ala-Korpela M, Langenberg C, Butterworth A, Bottolo L, Burgess S. High-throughput multivariable Mendelian randomization analysis prioritizes apolipoprotein B as key lipid risk factor for coronary artery disease. Int J Epidemiol. 2020
60. Marmillot P, Patel S, Lakshman MR. Reverse cholesterol transport is regulated by varying fatty acyl chain saturation and sphingomyelin content in reconstituted high-density lipoproteins. Metabolism. 2007;56:251–259 [PubMed: 17224341]
61. Hancock-Cerutti W, Lhomme M, Dauteuille C, Lecocq S, Chapman MJ, Rader DJ, Kontush A, Cuchel M. Paradoxical coronary artery disease in humans with hyperalphalipoproteinemia is associated with distinct differences in the high-density lipoprotein phosphosphingolipidome. J Clin Lipidol. 2017;11:1192–1200 e1193 [PubMed: 28826666]
62. Camont L, Lhomme M, Rached F, Le Goff W, Negre-Salvayre A, Salvayre R, Calzada C, Lagarde M, Chapman MJ, Kontush A. Small, dense high-density lipoprotein-3 particles are enriched in negatively charged phospholipids: relevance to cellular cholesterol efflux, antioxidative, antithrombotic, anti-inflammatory, and antiapoptotic functionalities. Arterioscler Thromb Vasc Biol. 2013;33:2715–2723 [PubMed: 24092747]
63. Darabi M, Kontush A. Phosphatidylserine in atherosclerosis. Curr Opin Lipidol. 2016;27:414–420 [PubMed: 27070078]
64. Jialin G, Xuefan G, Huiwen Z. SID1 transmembrane family, member 2 (Sidt2): a novel lysosomal membrane protein. Biochem Biophys Res Commun. 2010;402:588–594 [PubMed: 20965152]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

65. Ren R, Xu X, Lin T, Weng S, Liang H, Huang M, Dong C, Luo Y, He J. Cloning, characterization, and biological function analysis of the SidT2 gene from Siniperca chuatsi. Dev Comp Immunol. 2011;35:692–701 [PubMed: 21334374]
66. Li W, Koutmou KS, Leahy DJ, Li M. Systemic RNA Interference Deficiency-1 (SID-1) Extracellular Domain Selectively Binds Long Double-stranded RNA and Is Required for RNA Transport by SID-1. J Biol Chem. 2015;290:18904–18913 [PubMed: 26067272]
67. Aizawa S, Contu VR, Fujiwara Y, Hase K, Kikuchi H, Kabuta C, Wada K, Kabuta T. Lysosomal membrane protein SIDT2 mediates the direct uptake of DNA by lysosomes. Autophagy. 2017;13:218–222 [PubMed: 27846365]
68. Valdes VJ, Athie A, Salinas LS, Navarro RE, Vaca L. CUP-1 is a novel protein involved in dietary cholesterol uptake in Caenorhabditis elegans. PLoS One. 2012;7:e33962
69. Hulce JJ, Cognetta AB, Niphakis MJ, Tully SE, Cravatt BF. Proteome-wide mapping of cholesterol-interacting proteins in mammalian cells. Nat Methods. 2013;10:259–264 [PubMed: 23396283]
70. Sharpe LJ, Rao G, Jones PM, Glancey E, Aleidi SM, George AM, Brown AJ, Gelissen IC. Cholesterol sensing by the ABCG1 lipid transporter: Requirement of a CRAC motif in the final transmembrane domain. Biochim Biophys Acta. 2015;1851:956–964 [PubMed: 25732853]
71. Ouimet M, Franklin V, Mak E, Liao X, Tabas I, Marcel YL. Autophagy regulates cholesterol efflux from macrophage foam cells via lysosomal acid lipase. Cell Metab. 2011;13:655–667 [PubMed: 21641547]
72. Chen X, Gu X, Zhang H. Sidt2 regulates hepatocellular lipid metabolism through autophagy. J Lipid Res. 2018;59:404–415 [PubMed: 29363559]
73. Gao F, Li G, Liu C, et al. Autophagy regulates testosterone synthesis by facilitating cholesterol uptake in Leydig cells. J Cell Biol. 2018;217:2103–2119 [PubMed: 29618492]
74. Gao J, Gu X, Mahuran DJ, Wang Z, Zhang H. Impaired glucose tolerance in a mouse model of sidt2 deficiency. PLoS One. 2013;8:e66139
75. Chang G, Yang R, Cao Y, Nie A, Gu X, Zhang H. SIDT2 is involved in the NAADP-mediated release of calcium from insulin secretory granules. J Mol Endocrinol. 2016;56:249–259 [PubMed: 26744456]
76. Gao J, Zhang Y, Yu C, Tan F, Wang L. Spontaneous nonalcoholic fatty liver disease and ER stress in Sidt2 deficiency mice. Biochem Biophys Res Commun. 2016;476:326–332 [PubMed: 27233614]
77. Meng Y, Wang L, Ling L. Changes of lysosomal membrane permeabilization and lipid metabolism in sidt2 deficient mice. Exp Ther Med. 2018;16:246–252 [PubMed: 29896245]
78. Kennedy AR, Pissios P, Otu H, Roberson R, Xue B, Asakura K, Furukawa N, Marino FE, Liu FF, Kahn BB, Libermann TA, Maratos-Flier E. A high-fat, ketogenic diet induces a unique metabolic state in mice. Am J Physiol Endocrinol Metab. 2007;292:E1724–1739 [PubMed: 17299079]
79. Renaud HJ, Cui JY, Lu H, Klaassen CD. Effect of diet on expression of genes involved in lipid metabolism, oxidative stress, and inflammation in mouse liver-insights into mechanisms of hepatic steatosis. PLoS One. 2014;9:e88584
80. Teufel A, Itzel T, Erhart W, et al. Comparison of Gene Expression Patterns Between Mouse Models of Nonalcoholic Fatty Liver Disease and Liver Tissues From Patients. Gastroenterology. 2016;151:513–525 e510 [PubMed: 27318147]
81. Leon-Mimila P, Villamil-Ramirez H, Villalobos-Comparan M, et al. Contribution of common genetic variants to obesity and obesity-related traits in mexican children and adults. PLoS One. 2013;8:e70640
82. Obesity: preventing and managing the global epidemic. Report of a WHO consultation. World Health Organ Tech Rep Ser. 2000;894:i–xii, 1–253 [PubMed: 11234459]
83. Kuczmarski RJ, Ogden CL, Grummer-Strawn LM, Flegal KM, Guo SS, Wei R, Mei Z, Curtin LR, Roche AF, Johnson CL. CDC growth charts: United States. Adv Data. 2000:1–27
84. Leon-Mimila P, Vega-Badillo J, Gutierrez-Vidal R, et al. A genetic risk score is associated with hepatic triglyceride content and non-alcoholic steatohepatitis in Mexicans with morbid obesity. Exp Mol Pathol. 2015;98:178–183 [PubMed: 25597287]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

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

86. DeLong DM, DeLong ER, Wood PD, Lippel K, Rifkind BM. A comparison of methods for the estimation of plasma low- and very low-density lipoprotein cholesterol. The Lipid Research Clinics Prevalence Study. JAMA. 1986;256:2372–2377 [PubMed: 3464768]

87. National Cholesterol Education Program Expert Panel on Detection E, Treatment of High Blood Cholesterol in A. Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) final report. Circulation. 2002;106:3143–3421 [PubMed: 12485966]

88. Parker BL, Calkin AC, Seldin MM, et al. An integrative systems genetic analysis of mammalian lipid metabolism. Nature. 2019;567:187–193 [PubMed: 30814737]

89. Huynh K, Barlow CK, Jayawardana KS, Weir JM, Mellett NA, Cinel M, Magliano DJ, Shaw JE, Drew BG, Meikle PJ. High-Throughput Plasma Lipidomics: Detailed Mapping of the Associations with Cardiometabolic Risk Factors. Cell Chem Biol. 2019;26:71–84 e74 [PubMed: 30415965]

90. Browning BL, Browning SR. Genotype Imputation with Millions of Reference Samples. Am J Hum Genet. 2016;98:116–126 [PubMed: 26748515]

91. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 2009;19:1655–1664 [PubMed: 19648217]

92. Maples BK, Gravel S, Kenny EE, Bustamante CD. RFMix: a discriminative modeling approach for rapid and robust local-ancestry inference. Am J Hum Genet. 2013;93:278–288 [PubMed: 23910464]

93. Sabeti PC, Reich DE, Higgins JM, et al. Detecting recent positive selection in the human genome from haplotype structure. Nature. 2002;419:832–837 [PubMed: 12397357]

94. Szpiech ZA, Hernandez RD. selscan: an efficient multithreaded program to perform EHH-based scans for positive selection. Mol Biol Evol. 2014;31:2824–2827 [PubMed: 25015648]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

HIGHLIGHTS

- A GWAS in Mexican individuals identified the *SIDT2*/Val636Ile variant associated with higher HDL-C levels.
- *SIDT2*/Val636Ile was associated with higher HDL-C, ApoA1 and glycerophospholipids, and with lower LDL-C and ApoB levels in independent cohorts.
- The *SIDT2*/Val636Ile variant was also associated with a lower risk of premature CAD.
- Transcriptomic data in mice and humans identified lipid and lipoprotein metabolism pathways as the most significantly correlated with *SIDT2* liver expression.
- Cholesterol analog dehydroergosterol uptake was increased in HEK293T cells transfected with Ile636/*SIDT2* as compared to wildtype *SIDT2*.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

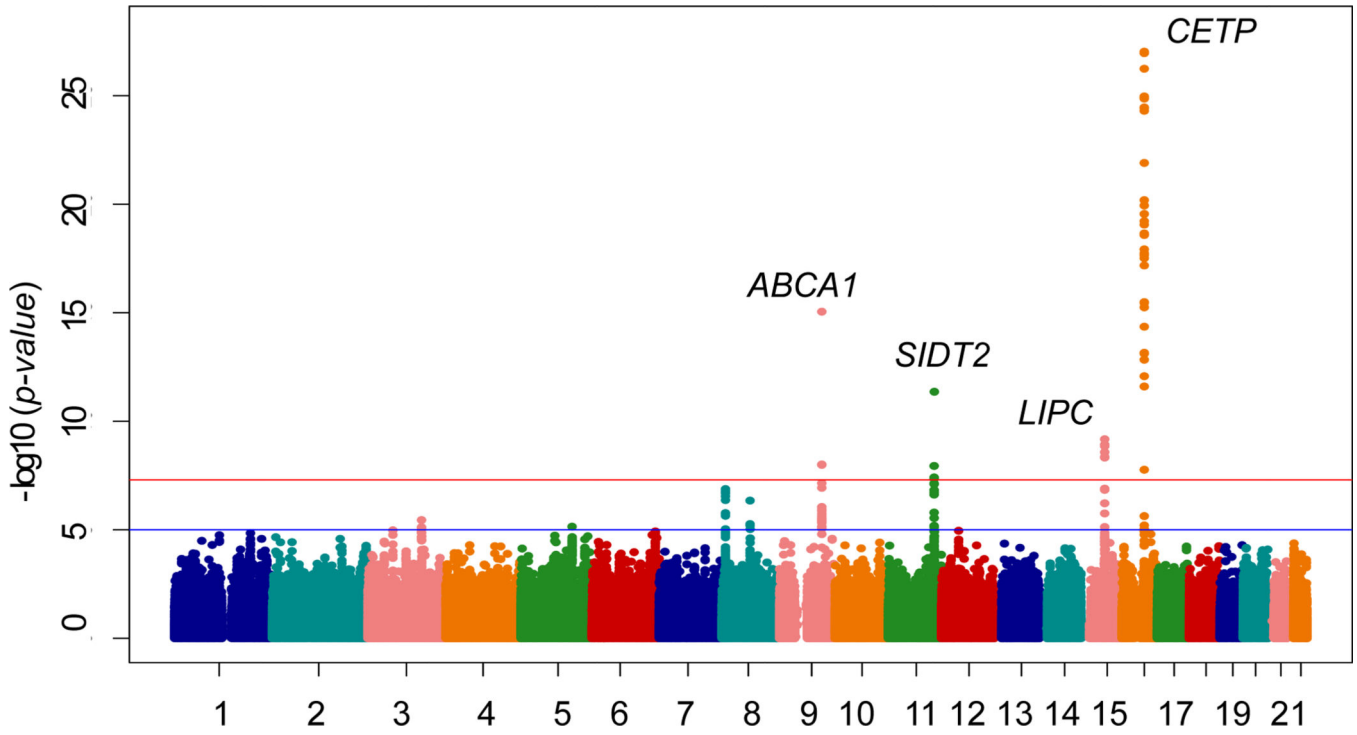

Figure. 1. Manhattan plot for HDL-C levels in the discovery phase.

Plot showing the -Log10 transformed *P*-value of SNPs for 2,335 Mexican children and adults. The red line indicates the genome-wide significance level (P=5×10^{-8}). Genes closest to the SNP with lowest *P*-value at each locus are indicated.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

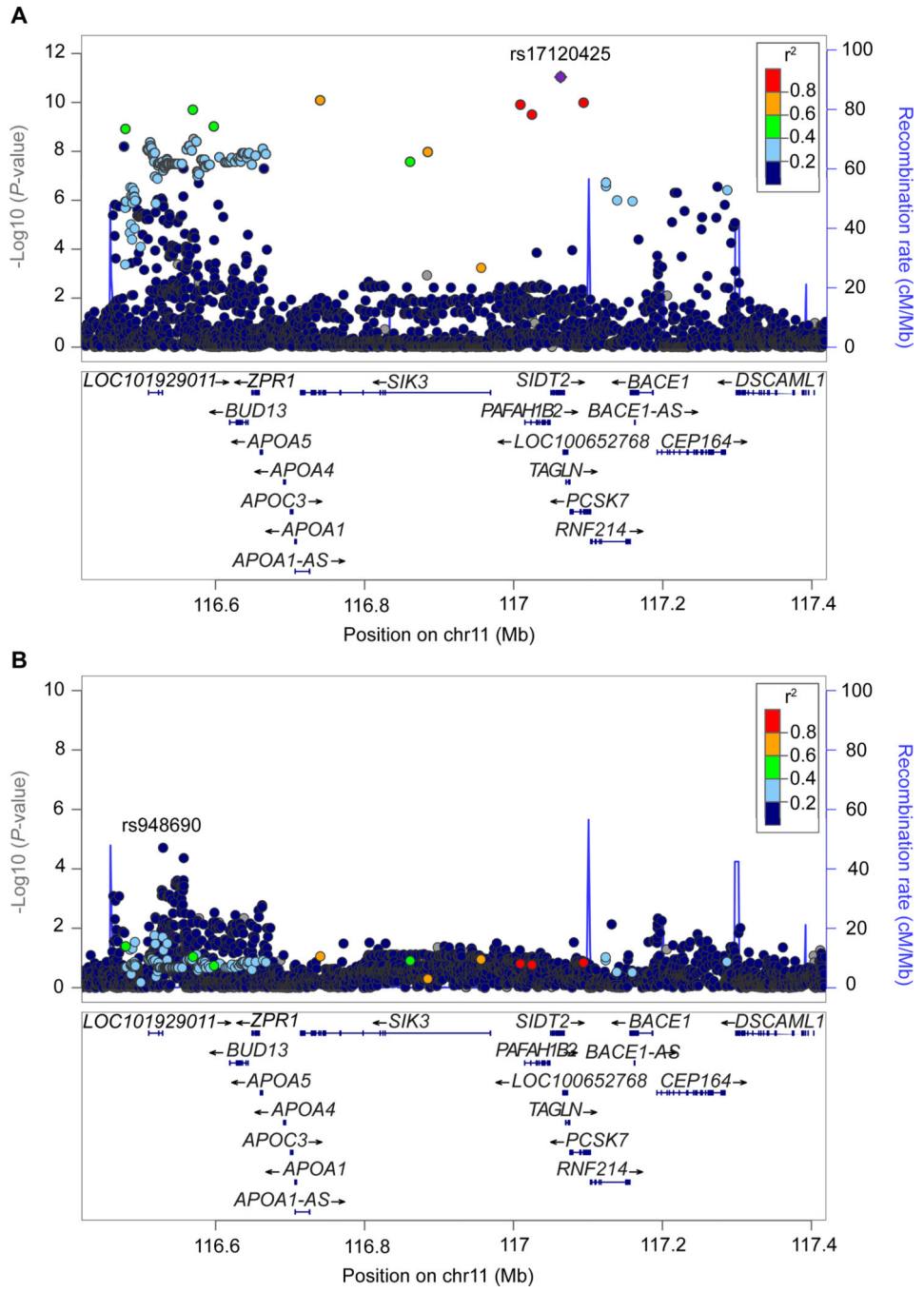

Figure 2. Locus zoom plots for associations with HDL-C levels in the 11q23 region, unconditioned and conditioning on rs17120425 (*SIDT2*/Val636Ile).

Plots show the regional associations before (A) and after (B) conditioning on rs17120425 (*SIDT2*/Val636Ile), which showed the strongest association with HDL-C levels (P=9.2×10^{-12}). SNPs are colored based on their linkage disequilibrium (r^2) with rs17120425 (purple diamond) using AMR subjects from 1000 Genomes data. -Log10 (P-values) indicate the significance of association for each SNP. Arrows on the horizontal blue lines show the direction of transcription, and rectangles represent exons.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

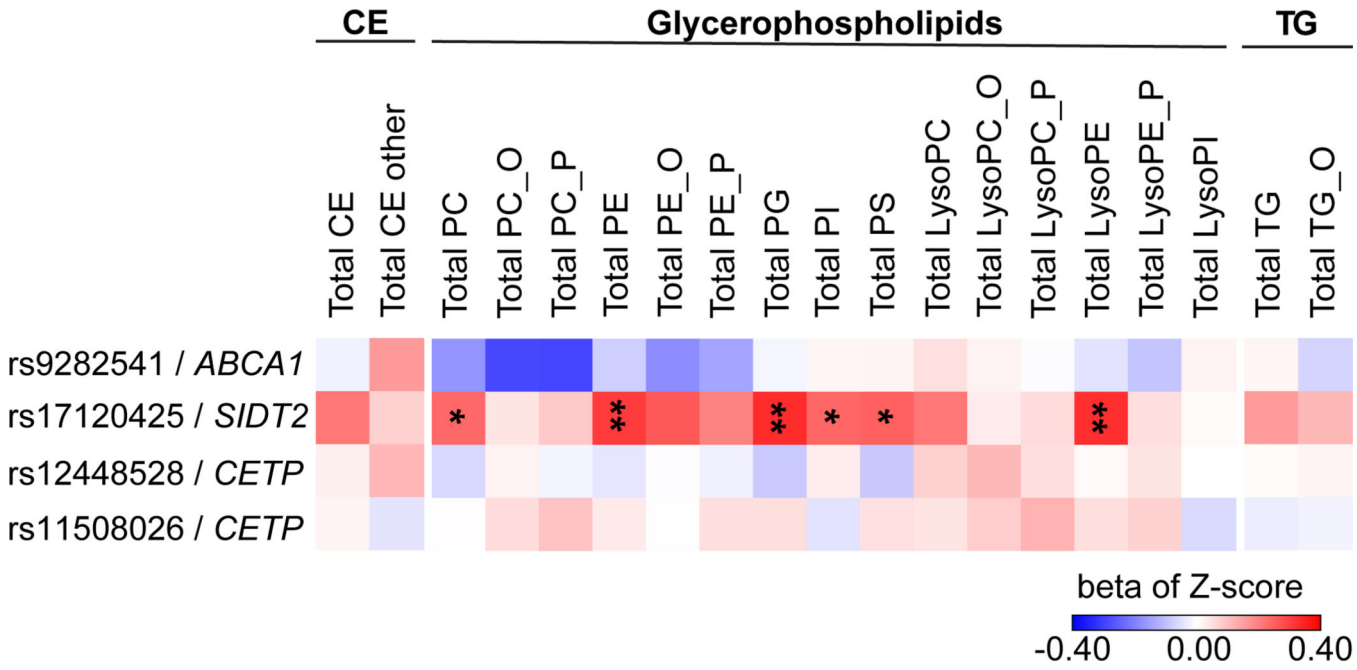

Figure 3. Heat map of associations between the *SIDT2*/Val636Ile variant and lipid classes in the MOBES cohort.

Color intensity reflects the Beta value (red for positive, blue for negative) obtained from linear regression between HDL-C associated SNPs and lipid classes in the MOBES cohort (n=375), adjusted by age, sex and BMI. **P*-value <0.05, ** *P*-value ≤0.001.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

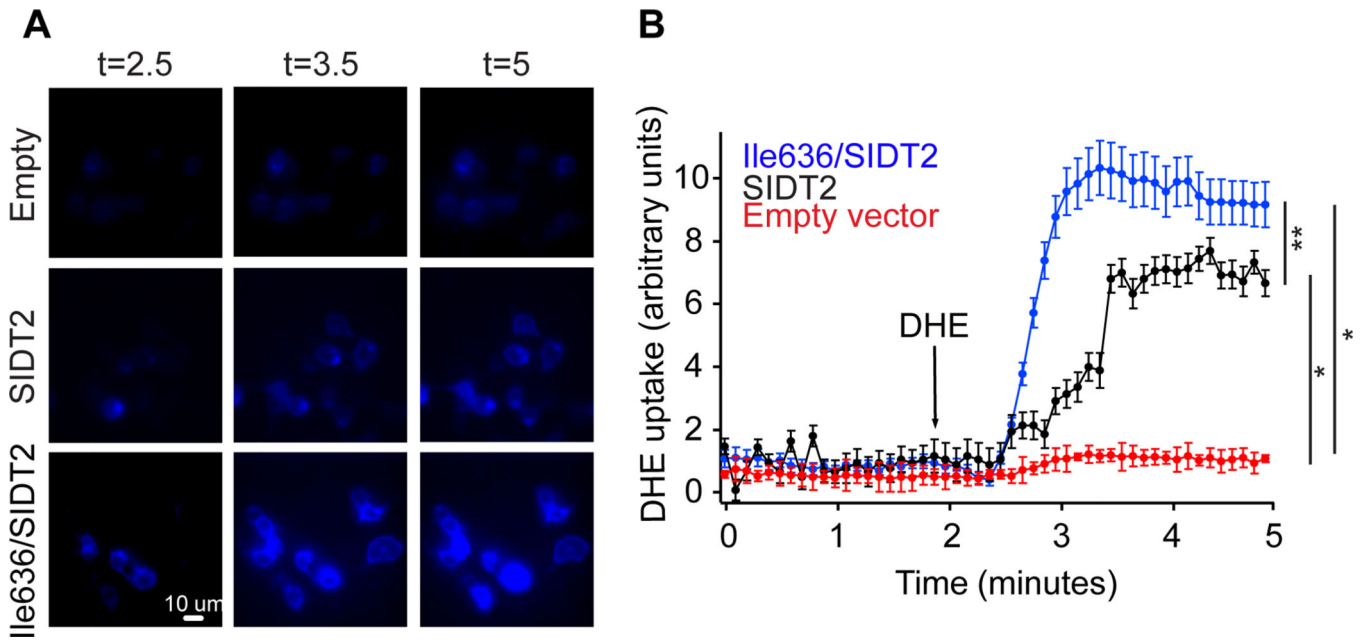

Figure 4. Uptake of the blue fluorescent cholesterol analog dehydroergosterol (DHE) by cells expressing wildtype and Ile636/SIDT2.

(A) Confocal microscopy images of HEK293T cells expressing the empty vector, wildtype and Ile636/SIDT2, at times 2.5, 3.5 and 5 minutes after adding the fluorescent cholesterol analog DHE to the culture. (B) Plot showing mean fluorescence intensity over time after adding DHE to the culture. Values show mean ± standard deviations from at least 4 independent experiments, each experiment shows mean values from all the cells in the focal plane. The red dots represent cells transfected with the empty vector; the black dots represent cells expressing wildtype SIDT2, and blue dots represent cells expressing Ile636/SIDT2. Addition of DHE is indicated with an arrow. **P-value* <0.05; ***P*-value <0.01.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

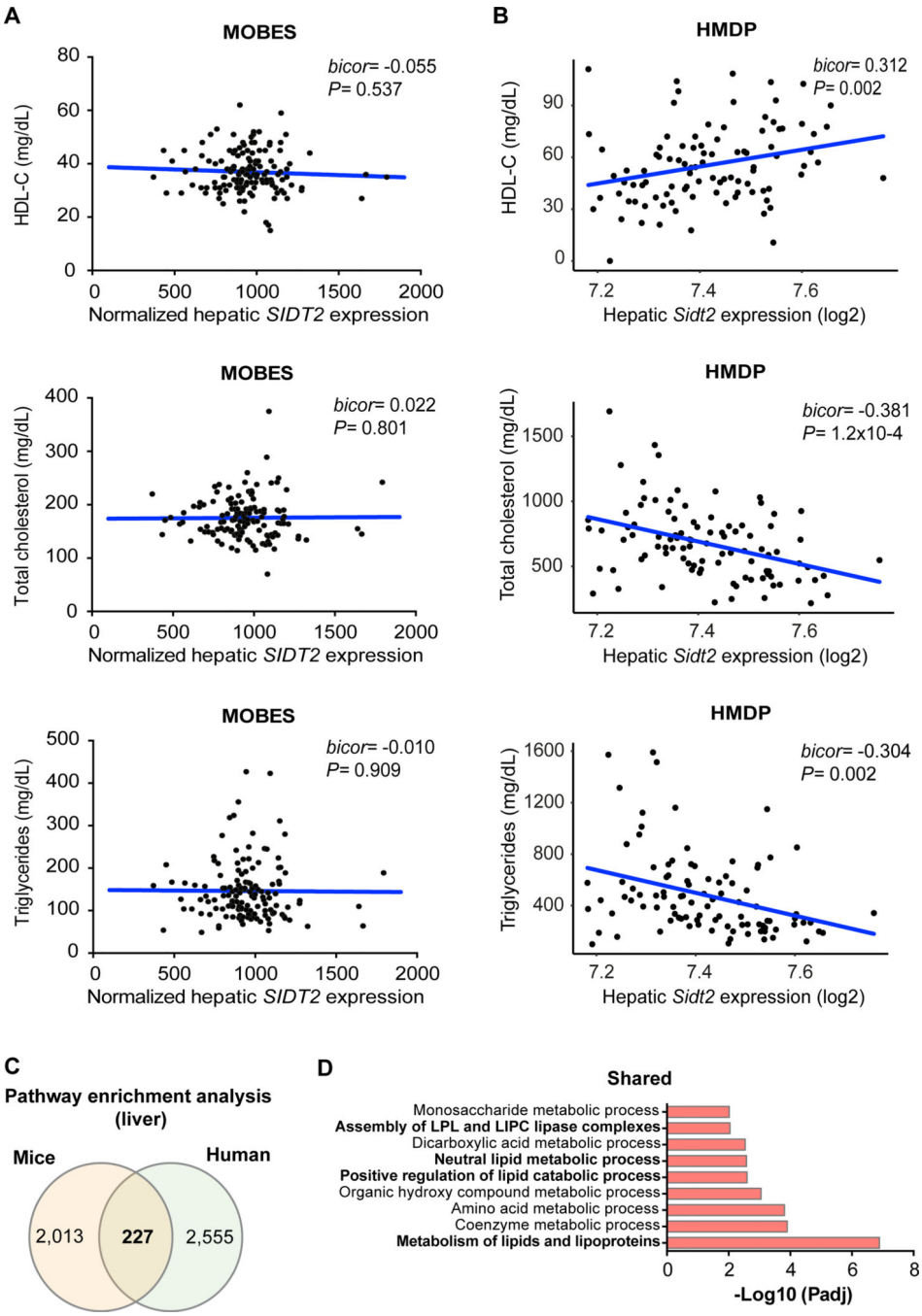

Figure 5. Correlations of liver *SIDT2* expression and lipid levels in mice and humans.

Correlations of normalized liver *Sidt2* expression with HDL-C, total cholesterol, and TG serum levels in MOBES cohort participants (A) and mice from the HMDP (B). (C) Venn diagram depicting the overlap of genes significantly correlated with *SIDT2* liver expression in HMDP mice and MOBES cohort participants ($P<1.0\times10^{-4}$). (D) Significantly enriched pathways in mice and humans.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 1.

Lead SNPs associated with HDL-C levels in the meta-analysis including Mexican children and adults.

Chr: position (GRCh37)	*Locus*	rs ID	Test Allele	Type of variant	Z-Score	*P*-value
3:136235908	*STAG1*	rs711975	A	Intronic	4.636	3.6×10^{-6}
5:130021677	*LOC105379171/LOC107986449*	rs6866550	G	Intergenic	4.489	7.2×10^{-6}
8:9177732	*LOC157273*	rs983309	A	Intronic	−5.056	4.3×10^{-7}
9:107589134	*ABCA1*	rs4149310	A	Intronic	5.734	9.8×10^{-9}
9:107620835	*ABCA1*	rs9282541	A	Missense. Arg230Cys	−8.040	8.9×10^{-16}
11:116529468	*LINC02702/BUD13*	rs948690	A	Intergenic	5.708	1.1×10^{-8}
11:117063003	*SIDT2*	rs17120425	A	Missense. Val636Ile	6.925	4.4×10^{-12}
15:58723479	*LIPC*	rs1077834	A	Upstream	−6.055	1.4×10^{-9}
16:56985555	*CETP*	rs12448528	A	Intergenic	−7.842	4.4×10^{-15}
16:56999328	*CETP*	rs11508026	A	Intronic	9.226	2.8×10^{-20}

HDL-C values were log transformed for the analysis. *P*-values for additive models were adjusted for sex, age, and either BMI or BMI percentile in children (n=2,335). HDL-C, High-density lipoprotein cholesterol; Chr: Chromosome.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 2.

Association of selected SNPs with lipid traits in the GEA cohort.

Chr: position (GRCh37)	*Locus*	rs ID	Test Allele	HDL-C		TC		LDL-C		TG		ApoA1		ApoB	
				B (SE)	*P- value*	B (SE)	*P-value*	B (SE)	*P-value*	B (SE)	*P-value*	B (SE)	*P- value*	B (SE)	*P-value*
3:136235908	*STAG1*	rs711975	A	1.53 (0.55)	0.006	3.21 (1.63)	0.032	2.99 (1.42)	0.023	−5.37 (4.49)	0.038	2.32 (1.67)	0.136	0.70 (1.27)	0.500
5:130021677	*LOC105379171/ LOC107986449*	rs6866550	G	1.80 (0.72)	0.013	4.38 (2.15)	0.044	2.84 (1.86)	0.197	−1.08 (5.91)	0.949	2.23 (2.20)	0.286	1.74 (1.66)	0.403
8:9177732	*LOC157273*	rs983309	A	−2.32 (0.50)	2.0×10^{-6}	−4.30 (1.60)	0.005	−2.69 (1.35)	0.016	4.03 (6.46)	0.076	−4.80 (1.43)	0.001	0.19 (1.19)	0.982
9:107589134	*ABCA1*	rs4149310	A	1.25 (0.49)	0.007	3.68 (1.55)	0.017	0.83 (1.32)	0.650	14.86 (6.22)	0.011	4.22 (1.38)	0.001	0.94 (1.15)	0.371
9:107620835	*ABCA1*	rs9282541	A	−3.81 (0.71)	5.5×10^{-9}	−6.49 (2.21)	0.004	−2.54 (1.88)	0.312	−1.53 (8.39)	0.639	−8.81 (2.06)	2.0×10^{-6}	−1.36 (1.68)	0.452
11:116529468	*LINC02702/ BUD13*	rs948690	A	1.10 (0.52)	0.017	−1.11 (1.56)	0.374	−0.57 (1.35)	0.705	−12.19 (4.27)	0.001	1.94 (1.60)	0.126	−2.93 (1.21)	0.011
11:117063003	*SIDT2*	rs17120425	A	3.35 (0.75)	4.2×10^{-5}	−3.66 (2.31)	0.157	−6.22 (1.96)	0.004	−4.75 (8.75)	0.477	10.29 (2.14)	3.0×10^{-6}	−3.92 (1.74)	0.038
15:58723479	*LIPC*	rs1077834	A	−0.59 (0.46)	0.204	−0.15 (1.46)	0.930	1.66 (1.23)	0.090	−5.44 (5.87)	0.025	−3.77 (1.30)	0.009	−0.90 (1.08)	0.281
16:56985555	*CETP*	rs12448528	A	−3.77 (0.57)	2.9×10^{-12}	−6.02 (1.85)	0.001	−2.96 (1.57)	0.062	3.91 (7.48)	0.340	−5.26 (1.66)	2.1×10^{-4}	−2.09 (1.39)	0.118
16:56999328	*CETP*	rs11508026	A	2.79 (0.47)	3.8×10^{-10}	2.25 (1.51)	0.130	−0.62 (1.28)	0.830	1.38 (6.06)	0.621	4.08 (1.35)	0.001	−0.49 (1.12)	0.979

P-values were calculated using log transformed HDL-C, TC, LDL-C, TG, ApoA1 and ApoB levels. B values were calculated for the minor alleles and are expressed as mg/dL. *P*-values for additive models were adjusted for sex, age, and BMI (n=1,592), SNPs rs711975, rs6866550 and rs948690 were genotyped in a subset of 1,036 individuals recruited as controls. Chr: Chromosome; HDL-C, High density lipoprotein cholesterol; TC, Total cholesterol; LDL-C, Low density lipoprotein cholesterol; TG, Triglycerides; ApoA1, Apolipoprotein A1; ApoB, Apolipoprotein B.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 3.

Association of selected SNPs with lipid traits in the MOBES cohort.

Chr: position (GRCh37)	*Locus*	rs ID	Test Allele	HDL-C		TC		LDL-C		TG	
				B (SE)	*P- value*	B (SE)	*P- value*	B (SE)	*P- value*	B (SE)	*P- value*
3:136235908	*STAG1*	rs711975	A	−0.10 (0.63)	0.903	−3.10 (2.71)	0.275	−4.66 (2.22)	0.036	5.77 (5.13)	0.582
5:130021677	*LOC105379171/ LOC107986449*	rs6866550	G	−0.06 (0.76)	0.753	−0.62 (3.31)	0.752	−0.45 (2.73)	0.694	−1.15 (6.26)	0.840
8:9177732	*LOC157273*	rs983309	A	−0.38 (0.65)	0.810	−2.80 (2.81)	0.648	−3.54 (2.30)	0.255	6.25 (5.24)	0.236
9:107589134	*ABCA1*	rs4149310	A	0.66 (0.60)	0.302	3.70 (2.62)	0.305	3.16 (2.15)	0.162	2.98 (4.88)	0.345
9:107620835	*ABCA1*	rs9282541	A	−3.77 (0.96)	4.8×10^{-4}	−8.78 (4.27)	0.077	−6.38 (3.50)	0.073	−4.26 (7.99)	0.844
11:116529468	*LINC02702/BUD13*	rs948690	A	−1.22 (0.57)	0.095	−2.13 (2.47)	0.652	−2.09 (2.03)	0.617	6.57 (4.67)	0.142
11:117063003	*SIDT2*	rs17120425	A	2.92 (0.98)	0.005	7.79 (4.30)	0.072	3.72 (3.54)	0.226	9.15 (8.04)	0.196
15:58723479	*LIPC*	rs1077834	A	−0.86 (0.57)	0.192	3.13 (2.50)	0.347	4.08 (2.05)	0.143	−2.57 (4.67)	0.264
16:56985555	*CETP*	rs12448528	A	−1.77 (0.72)	0.013	−1.71 (3.14)	0.705	0.12 (2.58)	0.668	−0.29 (5.86)	0.999
16:56999328	*CETP*	rs11508026	A	2.16 (0.55)	8.8×10^{-5}	0.11 (2.44)	0.774	−1.14 (1.99)	0.695	−3.99 (4.54)	0.277

P-values were calculated using log transformed HDL-C, TC, LDL-C and TG levels. B values were calculated for the minor alleles and are expressed as mg/dL. *P*-values for additive models were adjusted for sex, age, and BMI (n=555). Chr: Chromosome; HDL-C, High-density lipoprotein cholesterol; TC, Total cholesterol; LDL-C, Low-density lipoprotein cholesterol; TG, Triglycerides.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 4.

Association of HDL-C associated SNPs with premature coronary artery disease in the GEA cohort.

Chr: position (GRCh37)	*Locus*	rs ID	Allele Frequency			OR (95% CI)	*P-value*
			Test Allele	Controls (n=1,592)	Premature CAD (n=1,095)		
3:136235908	*STAG1*	rs711975	A	35.8	30.6	0.77 (0.67–0.90)	0.001
5:130021677	*LOC105379171/ LOC107986449*	rs6866550	G	16.7	15.0	0.86 (0.71–1.05)	0.149
8: 9177732	*LOC157273*	rs983309	A	29.3	30.5	1.06 (0.93–1.21)	0.370
9: 107589134	*ABCA1*	rs4149310	A	34.4	36.6	1.11 (0.98–1.26)	0.116
9: 107620835	*ABCA1*	rs9282541	A	9.9	7.1	0.70 (0.56–0.86)	0.001
11:116529468	*LINC02702/ BUD13*	rs948690	A	51.6	46.7	0.81 (0.70–0.93)	0.003
11: 117063003	*SIDT2*	rs17120425	A	9.5	8.0	0.80 (0.65–0.99)	0.039
15: 58723479	*LIPC*	rs1077834	A	42.6	39.1	0.87 (0.77–0.98)	0.027
16: 56985555	*CETP*	rs12448528	A	20.5	18.1	0.87 (0.75–1.02)	0.087
16: 56999328	*CETP*	rs11508026	A	36.3	39.0	1.11 (0.99–1.26)	0.087

Associations with premature CAD were adjusted for age, sex and BMI. SNPs rs711975, rs6866550 and rs948690 were genotyped in a subset of 1,036 controls and 625 cases. *P*-values were calculated by logistic regression. HDL-C, High-density lipoprotein cholesterol; Chr: Chromosome; CAD, coronary artery disease; OR, Odds ratio; CI, confidence interval.