

1 **A functional variant of the *SIDT2* gene involved in cholesterol transport is**
2 **associated with HDL-C levels and premature coronary artery disease.**

3
4 Paola León-Mimila,^{1,19} Hugo Villamil-Ramírez,^{1,19} Luis R. Macias-Kauffer,^{1,19} Leonor Jacobo-
5 Albavera,^{2,19} Blanca E. López-Contreras,¹ Rosalinda Posadas-Sánchez,³ Carlos Posadas-
6 Romero,³ Sandra Romero-Hidalgo,⁴ Sofía Morán-Ramos,^{1,5} Mayra Domínguez-Pérez,² Marisol
7 Olivares-Arevalo,¹ Priscilla Lopez-Montoya,¹ Roberto Nieto-Guerra,¹ Víctor Acuña-Alonzo,⁶
8 Gastón Macín-Pérez,⁶ Rodrigo Barquera-Lozano,⁶ Blanca E. del Río-Navarro,⁷ Israel González-
9 González,⁸ Francisco Campos-Pérez,⁸ Francisco Gómez-Pérez,⁹ Victor J. Valdés,¹⁰ Alicia
10 Sampieri,¹⁰ Juan G. Reyes-García,¹¹ Miriam del C. Carrasco-Portugal,¹² Francisco J. Flores-
11 Murrieta,^{11,12} Carlos A. Aguilar-Salinas,^{9,13} Gilberto Vargas-Alarcón,¹⁴ Diana Shih,¹⁵ Peter J.
12 Meikle,¹⁶ Anna C. Calkin,^{17,18} Brian G. Drew,^{17,18} Luis Vaca,¹⁰ Aldons J. Lusic,¹⁵ Adriana Huertas-
13 Vazquez,¹⁵ Teresa Villarreal-Molina,^{2*} and Samuel Canizales-Quinteros,^{1*}

14
15 ¹Unidad de Genómica de Poblaciones Aplicada a la Salud, Facultad de Química, UNAM/Instituto Nacional
16 de Medicina Genómica (INMEGEN), Mexico City, 14610, Mexico

17 ²Laboratorio de Enfermedades Cardiovasculares, INMEGEN, Mexico City, 14610, Mexico

18 ³Departamento de Endocrinología, Instituto Nacional de Cardiología Ignacio Chávez, Mexico City, 14080,
19 México

20 ⁴Departamento de Genómica Computacional, INMEGEN, Mexico City, 14610, Mexico

21 ⁵Consejo Nacional de Ciencia y Tecnología (CONACyT), Mexico City, 03940, México

22 ⁶Escuela Nacional de Antropología e Historia, Mexico City, 14030, Mexico

23 ⁷ Hospital Infantil de México Federico Gómez, Mexico City, 06720, Mexico

24 ⁸Hospital General Rubén Leñero, Mexico City, 11340, Mexico

25 ⁹Unidad de Investigación en Enfermedades Metabólicas and Departamento de Endocrinología y
26 Metabolismo, Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, Mexico City, 14000,
27 Mexico

28 ¹⁰Instituto de Fisiología Celular, UNAM, Mexico City, 04510, Mexico

29 ¹¹Sección de Estudios de Posgrado e Investigación, Escuela Superior de Medicina, Instituto Politécnico
30 Nacional, Mexico City, 11340, México

31 ¹²Unidad de Investigación en Farmacología, Instituto Nacional de Enfermedades Respiratorias Ismael
32 Cosío Villegas, Mexico City, 14080, Mexico

33 ¹³Tecnológico de Monterrey, Escuela de Medicina y Ciencias de la Salud. Monterrey, N.L. 64710, Mexico

34 ¹⁴Departamento de Biología Molecular, Instituto Nacional de Cardiología Ignacio Chávez, Mexico City,
35 14080, Mexico

36 ¹⁵ Division of Cardiology, Department of Medicine, David Geffen School of Medicine, University of
37 California, Los Angeles, CA, 90095, USA

38 ¹⁶Head Metabolomics Laboratory, Baker Heart and Diabetes Institute, Melbourne, VIC 3004, Australia

39 ¹⁷Baker Heart and Diabetes Institute, Melbourne, VIC, 3004, Australia

40 ¹⁸Central Clinical School, Monash University, Melbourne, VIC, 3004, Australia.

41 ¹⁹These authors contributed equally to this work.

42

43 *Correspondence: mvallareal@inmegen.gob.mx (T.V.-M.), cani@unam.mx (S.C.-Q.).

44

45

46

47

48

49

50

51

52

53

54 **ABSTRACT**

55

56 Low HDL-C is the most frequent dyslipidemia in Mexicans, but few studies have
57 examined the underlying genetic basis. Moreover, few lipid-associated variants have been
58 tested for coronary artery disease (CAD) in Hispanic populations. Here, we performed a GWAS
59 for HDL-C levels in 2,183 Mexican individuals, identifying 7 *loci*, including three with genome-
60 wide significance and containing the candidate genes *CETP*, *ABCA1* and *SIDT2*. The *SIDT2*
61 missense Val636Ile variant was associated with HDL-C levels for the first time, and this
62 association was replicated in 3 independent cohorts ($P=5.5 \times 10^{-21}$ in the conjoint analysis). The
63 *SIDT2*/Val636Ile variant is more frequent in Native American and derived populations than in
64 other ethnic groups. This variant was also associated with increased ApoA1 and
65 glycerophospholipid serum levels, decreased LDL-C and ApoB levels and a lower risk of
66 premature CAD. Because *SIDT2* was previously identified as a protein involved in sterol
67 transport, we tested whether the *SIDT2*/Ile636 protein affected this function using an *in vitro* site-
68 directed mutagenesis approach. The *SIDT2*/Ile636 protein showed increased uptake of the
69 cholesterol analog dehydroergosterol, suggesting this variant is functional. Finally, liver
70 transcriptome data from humans and the Hybrid Mouse Diversity Panel (HMDP) are consistent
71 with the involvement of *SIDT2* in lipid and lipoprotein metabolism. In conclusion, this is the first
72 study assessing genetic variants contributing to HDL-C levels and coronary artery disease in the
73 Mexican population. Our findings provide new insight into the genetic architecture of HDL-C and
74 highlight *SIDT2* as a new player in cholesterol and lipoprotein metabolism in humans.

75

76

77 INTRODUCTION

78
79 Observational epidemiologic studies have reported that low plasma high density
80 lipoprotein cholesterol (HDL-C) concentrations are an independent risk factor for cardiovascular
81 disease.¹⁻⁴ Heritability of HDL-C serum levels has been estimated as high as 70% in various
82 populations, including Mexican-Americans.⁵⁻⁸ Genome wide association studies (GWAS) have
83 successfully identified more than 150 loci associated with lipid levels mainly in European
84 populations,⁹⁻¹² while relatively few GWAS have been performed in Mexicans.¹³⁻¹⁶ The main
85 HDL-C associated loci identified in Europeans are also associated with HDL-C levels in
86 Mexicans, although novel loci have been reported in the latter group.¹⁴ Notably, a functional
87 variant apparently private to the Americas (*ABCA1* Arg230Cys) was found to be associated with
88 lower HDL-C levels in Mexicans.^{17; 18} Although low HDL-C levels are a well-established
89 cardiovascular risk factor, Mendelian randomization studies have shown that most genetic
90 variants associated with this trait are not associated with cardiovascular risk, suggesting that this
91 relationship is not necessarily causal.¹⁹⁻²¹ In this regard, it has been postulated that pleiotropic
92 effects of the genetic variants or HDL-C particle functionality rather than HDL-C plasma
93 concentrations may affect cardiovascular risk.²²⁻²⁵

94
95 Low HDL-C levels are highly prevalent in the Mexican population.²⁶⁻²⁸ This population
96 group has been underrepresented in GWAS, and few lipid-associated variants have been tested
97 for coronary artery disease (CAD) risk in Mexicans.²⁹ Therefore, we performed a GWAS for
98 HDL-C levels in a cohort of Mexican individuals, using a multiethnic array that includes rare and
99 common genetic variants for Hispanic populations. We then sought to replicate these
100 associations in two independent cohorts: the Genetics of Atherosclerotic Disease (GEA) and the
101 Morbid Obesity Surgery (MOBES) studies, and to test their possible effect on CAD risk. Lastly,

102 we explored the effect of the *SIDT2* gene, a member of a novel family of cholesterol
103 transporters,³⁰ on lipid metabolism using existing liver transcriptome data from the Hybrid Mouse
104 Diversity Panel, and the effect of the *SIDT2* Val636Ile variant on the cholesterol analog
105 dehydroergosterol (DHE) uptake *in vitro* using a site-directed mutagenesis approach.

106

107 **SUBJECTS AND METHODS**

108

109 ***Study populations***

110

111 *Discovery phase cohorts*

112 *Obesity Research Study for Mexican Children (ORSMEC) and Mexican Adult cohorts*

113 The ORSMEC cohort includes 1,080 school-aged children (6-12 years), 553 with normal-
114 weight and 527 with obesity. The Mexican adult cohort includes 1,073 adults aged 18-82 years,
115 486 with normal-weight and 587 with obesity. Recruitment strategies, inclusion criteria,
116 anthropometric and biochemical characteristics of both ORSMEC and the Mexican Adult cohorts
117 have been previously described.^{31; 32} Demographic and biochemical characteristics are
118 described in Tables S1 and S2. Briefly, obesity, height and weight were measured following
119 standard protocols and calibrated instruments as previously described.³¹ BMI was calculated as
120 body weight in kilograms divided by the square of height in meters (kg/m²). In adults, obesity
121 was defined as a BMI ≥ 30 kg/m² and normal weight as BMI < 25 kg/m² and ≥ 18.5 kg/m²
122 according to World Health Organization (WHO) criteria.³³ In children, BMI percentile was
123 calculated using age and sex specific BMI reference data, as recommended by the Centers for
124 Disease Control and Prevention, and obesity was defined as BMI percentile ≥ 95 and normal
125 weight as BMI percentile > 25 and ≤ 75 .³⁴

126

127 *Replication phase cohorts*

128 *Genetics of Atherosclerotic Disease (GEA) cohort*

129 The GEA study was designed to examine the genetic bases of premature CAD in the
130 Mexican population. We included 1,095 adults with premature coronary artery disease and 1,559
131 individuals recruited as controls without CAD or family history of premature CAD, 413 of which
132 had subclinical atherosclerosis defined by the presence of coronary artery calcification on helical
133 computed axial tomography. Recruitment strategy, inclusion criteria, anthropometric and
134 biochemical characteristics have been previously described.²⁹ Demographic and biochemical
135 characteristics are described in Table S3.

136

137 *The Mexican Obesity Surgery (MOBES) cohort*

138 This cohort was designed to study the genetic and metabolomic bases of obesity and
139 metabolic traits in Mexicans and includes 555 individuals with obesity (BMI $\geq 35\text{kg/m}^2$) aged 18
140 to 59 years who underwent bariatric surgery at the Rubén Leñero General Hospital in Mexico
141 City. Inclusion criteria of this cohort were previously described.³⁵ Demographic and biochemical
142 characteristics of this cohort are described in Table S4.

143

144 *Native Americans*

145 This cohort included 302 unrelated Native American individuals (Totonacs and Nahuas
146 from East-central Mexico) aged over 18 years. Inclusion criteria, anthropometric and biochemical
147 characteristics of these individuals have been previously described.¹⁸ Demographic and
148 biochemical characteristics of these individuals are described in Table S5.

149

150 *Ethics Statement*

151 This study was conducted according to the principles expressed in the Declaration of
152 Helsinki and was approved by the Ethics Committees of participant institutions. All adult
153 participants provided written informed consent prior to inclusion in the study. For children,

154 parents or guardians of each child signed the informed consent and children assented to
155 participate. For Totonac and Nahua participants a translator was used as needed.

156

157 *Biochemical measurements*

158 For all cohorts, blood samples were drawn after 8-12 hours of overnight fasting to
159 determine the serum levels of total cholesterol (TC), triglycerides (TG) and high-density
160 lipoprotein cholesterol (HDL-C) by enzymatic assays as previously described.³¹ Low density
161 lipoprotein-cholesterol (LDL-C) levels were calculated with the equation of Friedewald *et al.*³⁶ In
162 the GEA cohort, LDL-C levels were estimated with the equation of Friedewald modified by
163 DeLong *et al.*,³⁷ and serum levels of Apolipoprotein A1 and B (ApoA1 and ApoB) were measured
164 by immunonephelometry in a BN Pro Spec nephelometer (Dade Behring Marburg GmbH). In
165 children, low HDL-C levels were defined as HDL-C <40 mg/dL according to the US National
166 Cholesterol Education Program (NCEP) Expert Panel on Cholesterol Levels in Children and
167 Adolescents. For adults low HDL-C levels were defined as HDL-C ≤40 mg/dL for men and ≤50
168 mg/dL for women according to the US NCEP Adult Treatment Panel III (ATP III).³⁸

169

170 *Lipidomic study in the MOBES cohort*

171 Lipidomic analysis was performed using 375 serum samples from MOBES cohort
172 participants with previously described methods.^{39: 40} Briefly, liquid chromatography–electrospray
173 ionization–tandem mass spectrometry (LC–ESI–MS/MS) was used for lipidomic analysis on an
174 Agilent 1290 liquid chromatography system. Liquid chromatography was performed on a Zorbax
175 Eclipse Plus 1.8 μm C18, 50 × 2.1 mm column (Agilent Technologies). The mass spectrometer
176 was operated in dynamic/scheduled multiple reaction monitoring (dMRM) mode. There were 630
177 unique lipid species belonging to 31 lipid classes identified together with 15 stable isotope or
178 non-physiological lipid standards.

179

180 **GWAS and quality control**

181 Genomic DNA was isolated from peripheral white blood cells using standard methods. A
182 total of 2,153 children and adults included in the discovery phase were genotyped using the
183 Multi-Ethnic Genotyping Array (MEGA, Illumina, San Diego, CA, USA), which included >1600k
184 SNPs. This array includes both common and rare variants in Latin American, African, European
185 and Asian populations. Standard quality control (QC) measures were as previously described.³²
186 Identity-by-descent (IBD) was estimated using Plink v1.07.⁴¹ A total of 624,242 SNPs remained
187 after QC measures. After imputation with Beagle,⁴² 865,896 SNPs were included in the final
188 analysis. The quantile-quantile (QQ) plot for the HDL-C GWAS was well calibrated for the null
189 hypothesis ($\lambda_{GC} = 1.017$), indicating adequate control for confounders (Figure S1).

190

191 **Selection of SNPs for the replication analyses**

192 Ten SNPs at 7 loci found to be significantly associated with HDL-C levels in the
193 discovery phase ($P < 1 \times 10^{-6}$) were selected for replication in two independent cohorts (GEA
194 controls and MOBES). These were the lead SNP of 4 loci (rs9457930 in *LPAL2*, rs983309 in
195 *PPP1R3B*, rs1514661 in *ADAMTS20* and rs1077834 in *LIPC*), and 6 independent SNPs at 3 loci
196 (rs12448528 and rs11508026 in *CETP*, rs9282541 and rs4149310 in *ABCA1*, and rs17120425
197 and rs10488698 near the *APOA5* cluster). SNPs were defined as independent when LD with
198 other variants within the locus was low (r^2 and $D' < 0.2$). Selected SNPs were genotyped using
199 KASP assays (LGC, U.S. <http://www.lgcgroup.com>). The *SIDT2*/rs17120425 variant was also
200 genotyped in 302 Native Mexicans using a TaqMan assay (ABI Prism 7900HT Sequence
201 Detection System, Applied Biosystems). Call rates exceeded 95% and no discordant genotypes
202 were found in 10% of duplicate samples. No SNPs deviated from Hardy–Weinberg equilibrium in
203 any group ($P > 0.05$).

204

205 **Mendelian Randomization**

206 In order to test the causal effect of HDL-C levels on CAD we performed Mendelian
207 Randomization (MR) analyses by using HDL-C associated SNPs as an instrument. In MR
208 analyses, genetic variants act as proxies for HDL-C levels in a manner independent of
209 confounders.⁴³ We used the inverse-variance weighted (IVW) method⁴⁴ which assumes that all
210 genetic variants satisfy the instrumental variable assumptions (including zero pleiotropy). We
211 also performed MR-Egger regression,⁴⁵ which allow each variant to exhibit pleiotropy. Only
212 SNPs found to be associated with HDL-C in the GEA cohort (n=7) were included in the MR
213 analyses. Both methods were performed with the aid of the Mendelian Randomization R
214 package.⁴⁶

215

216 **Ancestry Analysis**

217 Global ancestry was estimated as previously described.³² Briefly, European (CEU) and
218 Yoruba (YRI) individuals from the 1000 genomes project and fifteen Nahua and Totonac trios
219 (Native American or NAT) were used as reference populations for ancestry analyses.
220 Multidimensional scaling components were calculated in Plink v1.07.⁴¹ Ancestral proportions
221 were determined with Admixture.⁴⁷ Local ancestry was determined to identify the origin of
222 chromosomal segments within the 11q23 region using RFMix.⁴⁸

223

224 **Positive selection analysis of the *SIDT2* Val636Ile variant**

225 Extended haplotype homozygosity (EHH) values were estimated to seek whether the
226 *SIDT2* derived “A” allele shows evidence of positive selection in a sample of Native Mexican
227 individuals. Sabeti *et al.*⁴⁹ introduced EHH to exploit the decay of haplotype homozygosity as a
228 function of genetic distance from a focal SNP. For this purpose, we used SNP 6.0 microarray
229 data (Affymetrix) from 233 Native Mexicans (Nahuas and Totonacs), and genotyped
230 *SIDT2*/rs17120425 in these individuals. The merged data were phased using Beagle.⁴² Phased

231 alleles were coded as 0/1 (ancestral “G”/derived “A”) to obtain EHH values using the Selscan
232 program.⁵⁰

233

234 ***HEK293T cell cultures and wildtype and SIDT2/Ile636 transfection***

235 Human embryonic kidney (HEK293T) cells were purchased from the American Type
236 Culture Collection (ATCC, Manassas, VA). Cells were grown on 35 mm culture dishes using
237 Dulbecco’s modified Eagles medium (DMEM) (Invitrogen, Carlsbad, CA) supplemented with
238 10% of heat inactivated fetal bovine serum (Wisent, premium quality, Canada), penicillin-
239 streptomycin and glutamine (Life Technologies) in an incubator with humidity control at 37 °C
240 and 5% CO₂.

241

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

250

251 The plasmids containing *SIDT2-GFP* and *Ile636/SIDT2-GFP* were transfected to
252 HEK293T cells grown on 35 mm Petri dishes with a glass bottom (MatTek, Ashland, MA, USA).
253 For transfection we used a mixture of 1 µg of DNA and 6 µl CaCl₂ reaching 60 µl of final volume
254 with distilled H₂O. The mixture was added by dropping to HeBS buffer (50mM HEPES, 280mM
255 NaCl, 1.5mM Na₂HPO₄, pH 7.05) and the final mixture was incubated for 30 minutes prior to

256 replacing with DMEM. The cells were incubated overnight with the transfection mixture, which
257 was then replaced with fresh medium to perform the assays 24 hours later.

258

259 ***Fluorescent cholesterol analog dehydroergosterol (DHE) uptake experiments***

260 The naturally occurring blue fluorescent cholesterol analog DHE was purchased from
261 Sigma (Saint Louis, MO). HEK293T cells expressing either SIDT2-GFP or SIDT2/Ile636-GFP
262 were incubated with 5mM DHE solution, adding 300 mM methyl- β -cyclodextrin (M β CD). The
263 solution was carefully resuspended and diluted in PBS to obtain a DHE/M β CD ratio of 1:8
264 (mol/mol) and 100 μ L of this solution were added to the HEK293T cells. Cells were monitored
265 using the scanning confocal microscope (FV1000, Olympus Japan). Focal plane was positioned
266 at the middle of the cells. Confocal images were acquired every 15 seconds with no averaging to
267 reduce photobleaching. Excitation was at 300 nm using a solid-state laser and emission was
268 collected at 535 nm.

269

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

271 *HMDP*: The HMDP is a collection of approximately 100 well-characterized inbred strains
272 of mice that can be used to analyze the genetic and environmental factors underlying complex
273 traits such as dyslipidemia, obesity, diabetes, atherosclerosis and fatty liver disease. We
274 analyzed the liver transcriptome of strains from this panel carrying the human cholesteryl ester
275 transfer protein (CETP) and the human ApoE3-Leiden transgenes. At the age of about 8 weeks,
276 these mice were placed on a “Western Style” synthetic high fat diet supplemented with 1%
277 cholesterol.⁵¹ After 16 weeks on this diet, plasma lipid profiles were measured by colorimetric
278 analysis as previously described^{52; 53} and animals were euthanized for the collection of liver
279 tissue. Total RNA was isolated from the left lobe using the Qiagen (Valencia, CA) RNeasy kit
280 (cat# 74104), as described.⁵⁴ Genome wide expression profiles were determined by
281 hybridization to Affymetrix HT-MG_430 PM microarrays. Microarray data were filtered as

282 previously described.⁵⁵ The ComBat method from the SVA Bioconductor package was used to
283 remove known batch effects.⁵⁶ All animal work was conducted according to relevant national and
284 international guidelines and was approved by the UCLA Institutional Animal Care and Use
285 Committee (IACUC).

286
287 *Humans:* Total RNA was extracted from liver biopsies of 144 MOBES participants using
288 Trizol reagent (Invitrogen). Clinical and biochemical characteristics of this subgroup of patients
289 are described in Table S6. RNA sequencing was performed as previously described.⁵⁷ Briefly,
290 RNA quality was assessed using the Bioanalyzer RNA chip analysis to ensure that the RNA
291 integrity number was >7. Complementary DNA libraries were prepared using the TruSeq RNA
292 Stranded Total RNA Library Preparation kit (Illumina) and sequenced using an Illumina
293 HiSeq2500 instrument, generating approximately 50 million reads/sample. After data quality
294 control, sequencing reads were mapped to the human reference genome using TopHat software
295 v2.0.⁵⁸ and quantified using Cufflinks software.⁵⁹

296
297 ***Statistical Methods***

298 For the discovery phase, genome-wide association with HDL-C was tested independently
299 in 4 groups (normal-weight and obese children and normal-weight and obese adults) under an
300 additive linear mixed model with sex, age and BMI percentile (children) or BMI (adults) as fixed
301 effects, and the genetic relatedness matrix as a random effect. An inverse variance method was
302 used to perform a meta-analysis of the 4 groups.⁶⁰ Genetic relationship matrices from genome-
303 wide data were considered for the analysis using GCTA software.⁶¹ A P -value $<1 \times 10^{-8}$ was
304 considered genome-wide significant, suggestive significance was defined as a P -value $<1 \times 10^{-6}$.
305 Group heterogeneity in the meta-analysis was evaluated by I^2 and Cochrane's Q ⁶² using the R
306 package meta. We used publicly available databases such as the GWAS Catalog

307 (<https://www.ebi.ac.uk/gwas/home>) to annotate associated SNPs. SNPs within a 1Mb range of
308 the *SIDT2*/Val636Ile variant were included in a locus zoom plot.⁶³

309
310 For the validation phase, linear regression under additive models was used to test for
311 genetic associations with lipid traits (HDL-C, LDL-C, TC and TG levels) in the GEA control,
312 MOBES and Native American cohorts, and to test for associations with lipid classes in the
313 MOBES cohort. Genetic associations with premature CAD in the GEA cohort were tested using
314 multiple logistic regression under additive models. All tests were adjusted by age, sex and BMI.
315 Associations were tested using the SPSS Statistics package (IBM SPSS Statistics, version 24,
316 Chicago, IL, USA), and statistical significance was considered at $P < 0.05$.

317
318 Correlations of *SIDT2* liver expression with serum lipid levels and the liver transcriptome
319 were performed using biweight midcorrelation (bicor) coefficient with the R package WGCNA, a
320 robust alternative to Pearson's correlation coefficient not sensitive to outliers.⁶⁴ Genes
321 significantly correlated with *SIDT2* expression in liver ($P \leq 1.0 \times 10^{-4}$) were tested for pathway
322 enrichment analysis using Metascape and ToppGene Suite software.^{65; 66} Enrichment P values
323 < 0.05 after FDR correction were considered significant.⁶⁷

324

325

326 **RESULTS**

327

328 Low HDL-C levels were highly prevalent in Mexican children and adults (38 and 27%,
329 respectively). This trait was significantly more frequent in obese as compared to lean individuals
330 (43.6% vs 12.7% respectively in children, and 81.6% vs 35.4% respectively in adults; $P < 0.001$)
331 (Tables S1 and S2). To identify loci associated with HDL-C levels, we carried out a GWAS in 2
332 independent cohorts of Mexican children and adults using a multi-ethnic array. Of note, the

333 same HDL-C associated loci were found in children and adults, regardless of obesity status, and
334 effect sizes were similar and showed the same directionality. There was no significant evidence
335 of heterogeneity (Table S7), and therefore a fixed-effects meta-analysis was conducted.

336
337 In total, we identified 64 variants distributed across 7 *loci* associated with HDL-C levels
338 with genome wide or suggestive significance ($P < 1.0 \times 10^{-6}$) after adjusting for age, sex, BMI and
339 ancestry (Figure 1, Table S7). Most of the 64 variants were also associated with other lipid
340 parameters (Table S7). Four SNPs showed genome-wide significance ($P < 1.0 \times 10^{-8}$),
341 rs11508026 and rs12448528 within the *CETP* locus on chromosome 16 ($B = 3.02$ mg/dL;
342 $P = 4.46 \times 10^{-18}$ and $B = -2.79$ mg/dL; $P = 5.92 \times 10^{-15}$, respectively); rs9282541 within the *ABCA1*
343 gene on chromosome 9 (Arg230Cys, $B = -3.44$ mg/dL; $P = 3.99 \times 10^{-13}$); and rs17120425 within the
344 *SIDT2* gene (Val636Ile, $B = 3.31$ mg/dL; $P = 1.52 \times 10^{-11}$), the latter associated with HDL-C levels for
345 the first time (Table 1). The gender stratified meta-analysis showed that the effect size of
346 *SIDT2*/Val636Ile on HDL-C levels was similar in females ($B = 3.48$, $P = 1.9 \times 10^{-7}$) and males
347 ($B = 3.27$, $P = 1.4 \times 10^{-5}$) (Table S8). Figure 2 shows the locus zoom plot for the rs17120425
348 (*SIDT2*) region, including 658 SNPs spanning 1 Mb in chromosome 11q23. No nearby SNPs in
349 LD ($r^2 > 0.2$) with rs17120425 (*SIDT2*) were found.

350
351 Notably, minor allele frequencies of *ABCA1*/Arg230Cys and *SIDT2*/Val636Ile variants
352 are highest in populations from the Americas (Table S9). We then genotyped the
353 *SIDT2*/Val636Ile variant in two independent Native American populations from central Mexico
354 (Totonacs and Nahuas), which was also associated with higher HDL-C levels in these
355 indigenous groups ($B = 2.81$ mg/dL; $P = 0.027$) (Table S10). Moreover, the rs17120425 “A” allele
356 was significantly more frequent in Native Mexicans (15%) than in Mexican Mestizos (10.3%,
357 $P = 0.001$) (Table S9), and local ancestry analysis revealed that this allele was found in a block of
358 Native American origin in 98% of individuals. However, according to EHH analysis, LD

359 extension break down was similar in the ancestral and derived rs17120425 alleles, with no
360 evidence of positive selection (Figure S2).

361

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

364 We then sought to replicate associations with HDL-C levels in 1,559 controls without
365 CAD from the GEA cohort. Seven of the 10 variants associated with HDL-C levels in the
366 discovery phase replicated in GEA controls ($P < 0.010$, Table 2): rs12448528 and rs11508026
367 within the *CETP* gene ($P = 7.2 \times 10^{-11}$ and $P = 2.9 \times 10^{-9}$, respectively), rs983309 (*PPP1R3B*,
368 $P = 4.0 \times 10^{-6}$), rs17120425/Val636Ile (*SIDT2*, $P = 7.0 \times 10^{-6}$), rs9282541 and rs4149310 (*ABCA1*,
369 $P = 2.0 \times 10^{-5}$ and $P = 0.01$, respectively) and rs10488698 (*BUD13-ZPR1-APOC3-APOA4-APOA5-*
370 *APOA1* cluster, $P = 6.5 \times 10^{-5}$). Altogether, these seven variants explained ~25% of HDL-C level
371 variation ($P_{\text{Genetic Risk Score}} = 7.0 \times 10^{-15}$). Table 2 shows associations of these SNPs with other lipid
372 traits in the GEA cohort. *SIDT2*/Val636Ile and rs10488698 (*BUD13-ZPR1-APOC3-APOA4-*
373 *APOA5-APOA1* cluster) were both associated with higher HDL-C and ApoA1 levels and lower
374 LDL-C and ApoB levels. SNP rs10488698, but not *SIDT2*/Val636Ile was also significantly
375 associated with lower TG levels. These two SNPs were the only variants associated both with
376 higher ApoA1 and lower ApoB levels. In the MOBES cohort which includes 555 individuals with
377 obesity, rs11508026 (*CETP*, $P = 9.6 \times 10^{-5}$), rs9282541 (*ABCA1*, $P = 1.1 \times 10^{-4}$) and *SIDT2*/Val636Ile
378 ($P = 0.003$) were significantly associated with HDL-C levels (Table 3).

379

380 The association of *SIDT2*/Val636Ile with HDL-C levels was highly significant in the
381 conjoint analysis including the discovery phase and replication cohorts ($B = 3.21$, $P = 5.5 \times 10^{-21}$).

382

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

385 We then sought associations of the 4 HDL-C associated SNPs in the MOBES cohort with
386 lipid classes known to be the main components of HDL-C lipoprotein particles (19 classes of
387 cholesterol esters, phospholipids and TG).^{68; 69} *SIDT2/Val636Ile* was significantly associated with
388 higher glycerophospholipid serum levels ($P<0.05$) (Figure 3), particularly with total
389 phosphatidylcholine (PC), phosphatidylethanolamine (PE), phosphatidylglycerol (PG),
390 phosphatidylinositol (PI) and phosphatidylserine (PS) as adjusted by age, sex, BMI and lipid
391 lowering treatment. In contrast, *ABCA1* (rs9282541) and *CETP* variants (rs12448528 and
392 rs11508026) were not significantly associated with any of these lipid classes.

393

394 **Association with premature CAD and Mendelian randomization**

395 We then tested whether HDL-C variants were also associated with premature CAD in the
396 GEA cohort. Four of the ten variants were significantly associated with CAD risk
397 (rs9282541/*ABCA1*, rs10488698/*APOA1* cluster, rs1077834/*LIPC* and rs17120425/*SIDT2*).
398 Rs10488698/*APOA1*-cluster and *SIDT2/Val636Ile* variants were associated with higher HDL-C
399 levels and lower CAD risk, while rs9282541 (*ABCA1*) and rs1077834 (*LIPC*) were associated
400 with both lower HDL-C levels and lower CAD risk (Table 4). Mendelian randomization analyses
401 were performed including only the 7 variants significantly associated with HDL-C levels in the
402 GEA control cohort. There was no evidence of a causal effect of HDL-C levels on CAD (IVW
403 $P=0.253$ and MR-Egger $P= 0.509$) (Table S11).

404

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

406 It was previously suggested that the transmembrane conserved cholesterol binding
407 (CRAC) domain of murine *Sidt2* associates with the cholesterol analogue dehydroergosterol in
408 HEK293 cells.³⁰ Because the *SIDT2/Val636Ile* variant is near the transmembrane CRAC domain
409 in the human *SIDT2* protein, we evaluated the effect of this variant on DHE uptake in HEK293T
410 cells. Interestingly, DHE uptake was enhanced in cells expressing the Ile636/*SIDT2* protein as

411 compared to cells expressing wildtype *SIDT2*, reaching the highest difference approximately 1.5
412 minutes after adding DHE to the culture (6.85 ± 0.42 AU vs 9.40 ± 0.73 AU, $P < 0.01$) (Figure 4,
413 Video S1).

414
415 ***SIDT2* liver expression correlates with the expression of genes involved in lipid and**
416 ***lipoprotein metabolism***

417 In the MOBES cohort, hepatic *SIDT2* expression did not differ according to the presence
418 of the Val636Ile variant ($P = 0.486$). Moreover, human *SIDT2* liver expression showed no
419 significant correlation with lipid traits including HDL-C, TC, or TG levels (Figure 5A). In contrast,
420 in HMDP mice fed with an atherogenic diet, hepatic *SidT2* expression correlated positively with
421 HDL-C levels ($r = 0.312$, $P = 0.002$), and negatively with total cholesterol and TG levels ($r = -0.381$,
422 $P = 1.2 \times 10^{-4}$; $r = -0.304$, $P = 0.002$, respectively) (Figure 5B).

423
424 To analyze correlations between *SIDT2* expression and the liver transcriptome in mice,
425 we used data from the genetically diverse HMDP where environmental factors are controlled. A
426 total of 2,240 genes showed significant correlations with *SidT2* expression, and the most
427 significantly enriched pathway was lipid and lipoprotein metabolism ($P_{FDR} = 1.5 \times 10^{-7}$). In human
428 liver tissue *SIDT2* expression correlated with the expression of 2,782 genes. Consistent with
429 observations in mice, the lipid and lipoprotein metabolism pathways were among the most
430 significantly enriched. *SIDT2* expression correlated with the expression of 227 genes that were
431 shared by both mice and humans, and as expected metabolism of lipids and lipoproteins was
432 the most significantly enriched pathway (Figure 5C-D).

433

434

435 **DISCUSSION**

436

437 Low HDL-C levels are the most common dyslipidemia in Mexicans, both in adults and
438 children.^{26; 28} Consistently, low HDL-C levels were highly prevalent in children and adults of the
439 present study. Moreover, the prevalence of low HDL-C levels was significantly higher in obese
440 than in lean individuals, in line with the high prevalence of metabolic syndrome observed in the
441 Mexican population.⁷⁰

442
443 Here, using a GWAS for HDL-C levels in different cohorts of the Mexican population, we
444 identified three loci associated with HDL-C levels at chromosomes 9, 11 and 16. Of note, the
445 effect and direction of the associations were consistent in the 4 study groups (normal weight and
446 obese children and adults). The most significant signal corresponds to the *CETP* locus, which is
447 known to be one of the main drivers of HDL-C levels across populations.^{10; 15; 71} The signal in
448 chromosome 9 is the *ABCA1*/Arg230Cys variant (rs9282541) previously reported as associated
449 with lower HDL-C levels by our group, apparently private to Native American and derived
450 populations.^{17; 18} The third genome-wide significant signal corresponds to a missense variant
451 within the *SIDT2* gene (Val636Ile, rs17120425), associated with higher HDL-C levels. This
452 association was replicated in 3 independent cohorts including Native Americans from Mexico.
453 This is relevant because Native Americans live in rural areas, while the GEA and MOBES
454 cohorts are urban populations, known to have different dietary habits, which may affect HDL-C
455 levels.⁷²

456
457 To our knowledge, this is the first time that the *SIDT2*/Val636Ile variant has been
458 associated with HDL-C levels. However, intronic *SIDT2* variants have been previously
459 associated with lipid traits and the metabolic syndrome in multi-ethnic cohorts, mainly in
460 Koreans.^{73; 74} Nevertheless, these variants (rs2269399, rs7107152, rs1242229 and rs1784042)
461 are not in LD with *SIDT2*/Val636Ile ($r^2 < 0.2$). Although the *SIDT2*/Val636Ile variant is not private
462 to the Americas, it is more frequent in Native Americans (15%) and Hispanics (6%), than other

463 ethnic groups (internationalgenome.org). Local ancestry analyses revealed that the derived “A”
464 allele was of Native American origin in most individuals. However, the extended haplotype
465 homozygosity (EHH) analysis of *SIDT2/Val636Ile* did not show evidence of recent positive
466 selection.

467
468 It is likely that previous GWAS in Mexicans failed to identify this SNP as associated with
469 HDL-C because this variant was not included in the microarray platforms used in these studies
470 and the paucity of Native American references for imputation.¹³⁻¹⁶ This 11q23 region contains
471 several lipid-associated signals, and it is thus necessary to define whether these signals are
472 independent of *SIDT2/Val636Ile*. The *BUD13-ZPR1-APOC3-APOA4-APOA5-APOA1* cluster
473 associated with TG levels is 400 Kb upstream *SIDT2*. Our regional LD analysis demonstrated
474 the lead TG-associated SNP rs964184 (*APOA5*)^{9; 13} and all other SNPs analyzed within this
475 cluster were in low LD with *SIDT2/Val636Ile* ($r^2 < 0.2$). This indicates that the *SIDT2* signal
476 associated with HDL-C and SNPs within the *BUD13-ZPR1-APOC3-APOA4-APOA5-APOA1*
477 cluster are in fact independent. Moreover, a previous GWAS in the Mexican population
478 identified two intronic *SIK3* gene variants within the 11q23 region, rs139961185 associated with
479 TG levels and rs11216230 with higher HDL-C levels.¹⁴ The latter association was replicated in
480 an independent Hispanic population.¹⁶ Of note, rs11216230 is more frequent in Mexicans (11%)
481 than in Europeans (1%) and is in high LD with *SIDT2/Val636Ile* in Mexican Americans from the
482 1000 Genomes project ($r^2 = 0.75$). This suggests that the association observed by Ko *et al.*¹⁴
483 could be driven by the *SIDT2/Val636Ile* variant.

484
485 In Mendelian randomization analyses, genetic variants act as proxies for HDL-C levels in
486 a manner independent of confounders to analyze the causality of HDL-C levels on coronary
487 artery disease. Our MR analysis is consistent with previous studies suggesting that higher HDL-
488 C levels are not causally protective against coronary heart disease.¹⁹⁻²¹ This suggests that the

489 effect of individual variants on CAD risk may be mediated by pleiotropic effects on other
490 cardiovascular risk factors, or on HDL-C composition and functionality.^{22-25; 75} The
491 *SIDT2/Val636Ile* variant was associated with higher HDL-C levels and with lower risk of
492 premature CAD. We thus explored whether this variant affects other cardiovascular risk
493 parameters in addition to HDL-C levels. Notably, *SIDT2/Val636Ile* was also associated with
494 higher ApoA1 levels, and lower LDL-C and ApoB serum levels in the GEA cohort. APOA1 is the
495 major protein component of HDL-C particles, and a Mendelian randomization analysis in Finnish
496 individuals reported that ApoA1 was not associated with risk of CAD.⁷⁶ A multivariable
497 Mendelian randomization study examining serum lipid and apolipoprotein levels reported that
498 only ApoB retained a robust relationship with the risk of CAD,⁷⁷ and recent Mendelian
499 randomization studies suggest that ApoB is the primary lipid determinant of cardiovascular
500 disease risk.^{78; 79} Thus, the association of *SIDT2/Val636Ile* with decreased cardiovascular risk
501 could be mediated by its effect on ApoB levels.

502

503 HDL lipidome composition has been associated with HDL-C functional properties,⁸⁰⁻⁸²
504 Notably, of the 4 main variants associated with HDL-C levels in the present study, only
505 *SIDT2/Val636Ile* was associated with lipid species, specifically with higher serum concentrations
506 of several glycerophospholipid classes including PE, PG, PC, PI and PS. It has been reported
507 that HDL-C particles enriched in phospholipids can increase HDL-C stability,⁸³ while decreased
508 levels of phospholipids in HDL-C were found to impair cholesterol efflux and decrease the
509 cardiovascular protective effects of HDLs.^{69; 82; 84} Particularly, recent studies indicate that
510 phosphatidylserine, a minor component of the monolayer surface of HDL-C, is enriched in small,
511 dense HDL-C particles, which display potent anti-atherosclerotic activities.^{83; 85} Although we
512 measured phospholipids in serum and not directly in HDL-C particles, a limitation of the study,
513 the association of *SIDT2/Val636Ile* with higher phospholipid levels is consistent with the lower
514 cardiovascular risk conferred by this variant.

515
516 The SIDT2 protein is found mainly in lysosome membranes,⁸⁶ is a lysosomal nucleic acid
517 transporter, and is expressed in several tissues, including the liver.⁸⁷⁻⁹¹ The mammalian SIDT2
518 protein has high homology to the *C. elegans* cholesterol uptake protein-1 (CUP-1).⁹² SIDT2 has
519 been identified as a sterol-interacting protein⁹³ and more recently as a cholesterol-binding
520 protein.³⁰ Moreover, SIDT2 is predicted to contain two CRAC domains (Cholesterol
521 Recognition/interaction Amino Acid Consensus), found in a broad range of proteins involved in
522 cholesterol transport, metabolism and regulation.^{30; 94} Specifically, the transmembrane CRAC
523 domain from human SIDT1 and mouse SIDT2 appears to bind cholesterol.³⁰ The Val636Ile
524 variant and the CRAC domain are located within the same transmembrane segment, and this
525 variant is 19 amino acids upstream the tyrosine CRAC domain residue, suggested to interact
526 with the cholesterol OH-polar group.³⁰ It is unknown if the Val636Ile variant modifies the
527 interactions of the CRAC domain with cholesterol, thus affecting circulating lipid levels.

528
529 In the present study, HEK293T cells expressing the Ile636/SIDT2 protein showed higher
530 cholesterol analog DHE uptake than those expressing the wildtype protein. This increased
531 uptake observed *in vitro*, may affect circulating levels of cholesterol-rich lipoproteins. Although
532 the mechanism by which this variant increases HDL-C serum levels is unknown, ABCA1-
533 mediated cholesterol efflux and HDL-C formation are primarily dependent on autophagy for
534 cholesterol source.⁹⁵ Because *Sidt2*^{-/-} deficient mice show blocked autophagosome maturation
535 and altered hepatic lipid homeostasis,⁹⁶ it is tempting to speculate that the putative increased
536 function of the Ile636 SIDT2 protein may enhance autophagy-mediated cholesterol flux, and thus
537 ABCA1-mediated HDL-C formation.

538
539 *Sidt2* knockout mice show a wide range of metabolic phenotypes including impaired
540 glucose tolerance likely due to compromised NAADP-involved insulin secretion.⁹⁶⁻¹⁰⁰ Meng *et*

541 *al.*,¹⁰⁰ showed that *Sid2* deficient mice present significantly increased serum total cholesterol,
542 TG and LDL-C levels, and significantly lower HDL-C serum levels as compared to *Sid2*^{+/-} mice.
543 These findings are consistent with the inverse correlation between hepatic *Sid2* expression and
544 total cholesterol and triglycerides levels, and the direct correlation of *Sid2* expression with HDL-
545 C levels observed in the HMDP. Moreover, *Sid2*^{-/-} mice not only have altered lipid serum levels,
546 but also showed impaired liver function and liver steatosis.^{96; 99; 100} Consistently, *Sid2* expression
547 correlated with increased fat liver content in HMDP mice (data not shown). In contrast, in the
548 MOBES cohort *SIDT2* liver expression did not correlate with serum lipid levels or liver steatosis,
549 and the *SIDT2/Val636Ile* variant showed no association with non-alcoholic fatty liver disease in
550 the GEA or MOBES cohorts (data not shown). Because MOBES cohort participants had morbid
551 obesity and do not represent the overall population, the lack of correlation between *SIDT2* liver
552 expression with lipid levels in this cohort must be interpreted with caution. Despite these
553 differences between mice and humans, *SIDT2* expression correlations with the liver
554 transcriptome revealed that the most significantly enriched pathways were lipid and lipoprotein
555 metabolism in both species. Altogether, these findings support a role of *SIDT2* in lipid
556 metabolism. Further studies are required to better understand the mechanisms by which *SIDT2*
557 participates in cholesterol and lipoprotein metabolism at the cellular and systemic levels, and its
558 role in cardiovascular risk.

559
560 In conclusion, this is the first study assessing genetic variants contributing to HDL-C
561 levels and coronary artery disease in the Mexican population. Our GWAS revealed for the first
562 time that the *SIDT2/Val636Ile* variant is associated with increased HDL-C and phospholipid
563 levels and decreased risk of CAD. We also provide evidence that the *SIDT2/Val636Ile* variant is
564 functional, increasing the uptake of a cholesterol analog *in vitro*. Our data support a role of
565 *SIDT2* in cholesterol and lipid metabolism. The mechanisms by which this protein affects lipid
566 and metabolic parameters in humans require further investigation.

567

568

569 **Supplemental Data**

570 Supplemental Data include two figures, eleven tables and one video.

571

572 **Declaration of Interests**

573 The authors declare no competing interests.

574

575 **Acknowledgements**

576 We thank Luz E. Guillén for her technical support. We also thank Productos Medix S.A. de C.V.

577 for their support to perform this study. This work was supported by grants FOSISS-289699 and

578 PEI-230129 from Mexican National Council for Science and Technology (CONACyT).

579

580 **Web Resources**

581 R, <https://www.r-project.org>

582 PLINK, <http://zzz.bwh.harvard.edu/plink>

583 gnomAD, <https://gnomad.broadinstitute.org>

584 1000 Genomes, <https://www.internationalgenome.org>

585 GWAS Catalog, <https://www.ebi.ac.uk/gwas/home>

586 WCGNA, <https://horvath.genetics.ucla.edu/html/CoexpressionNetwork/Rpackages/WGCNA>

587 ToppGene, <https://toppgene.cchmc.org>

588 Metascape, <https://metascape.org>

589 OMIM, <http://www.omim.org>

590 HGNC, <http://www.genenames.org>

591

592 **Data Availability**

593 The datasets supporting the current study have not been deposited in a public repository
594 because they are part of other studies in progress, but are available from the corresponding
595 authors on reasonable request.

596 **REFERENCES**

- 597
- 598 1. Castelli, W.P. (1988). Cholesterol and lipids in the risk of coronary artery disease--the
599 Framingham Heart Study. *Can J Cardiol* 4 Suppl A, 5A-10A.
- 600 2. Boden, W.E. (2000). High-density lipoprotein cholesterol as an independent risk factor in
601 cardiovascular disease: assessing the data from Framingham to the Veterans Affairs
602 High--Density Lipoprotein Intervention Trial. *Am J Cardiol* 86, 19L-22L.
- 603 3. Sharrett, A.R., Ballantyne, C.M., Coady, S.A., Heiss, G., Sorlie, P.D., Catellier, D., Patsch, W.,
604 and Atherosclerosis Risk in Communities Study, G. (2001). Coronary heart disease
605 prediction from lipoprotein cholesterol levels, triglycerides, lipoprotein(a), apolipoproteins
606 A-I and B, and HDL density subfractions: The Atherosclerosis Risk in Communities
607 (ARIC) Study. *Circulation* 104, 1108-1113.
- 608 4. Barter, P.J., Puranik, R., and Rye, K.A. (2007). New insights into the role of HDL as an anti-
609 inflammatory agent in the prevention of cardiovascular disease. *Curr Cardiol Rep* 9, 493-
610 498.
- 611 5. Mitchell, B.D., Kammerer, C.M., Blangero, J., Mahaney, M.C., Rainwater, D.L., Dyke, B.,
612 Hixson, J.E., Henkel, R.D., Sharp, R.M., Comuzzie, A.G., et al. (1996). Genetic and
613 environmental contributions to cardiovascular risk factors in Mexican Americans. The
614 San Antonio Family Heart Study. *Circulation* 94, 2159-2170.
- 615 6. Qasim, A., and Rader, D.J. (2006). Human genetics of variation in high-density lipoprotein
616 cholesterol. *Curr Atheroscler Rep* 8, 198-205.
- 617 7. Gao, C., Tabb, K.L., Dimitrov, L.M., Taylor, K.D., Wang, N., Guo, X., Long, J., Rotter, J.I.,
618 Watanabe, R.M., Curran, J.E., et al. (2018). Exome Sequencing Identifies Genetic
619 Variants Associated with Circulating Lipid Levels in Mexican Americans: The Insulin
620 Resistance Atherosclerosis Family Study (IRASFS). *Sci Rep* 8, 5603.
- 621 8. Williams, P.T. (2020). Quantile-specific heritability of high-density lipoproteins with
622 implications for precision medicine. *J Clin Lipidol* 14, 448-458 e440.
- 623 9. Teslovich, T.M., Musunuru, K., Smith, A.V., Edmondson, A.C., Stylianou, I.M., Koseki, M.,
624 Pirruccello, J.P., Ripatti, S., Chasman, D.I., Willer, C.J., et al. (2010). Biological, clinical
625 and population relevance of 95 loci for blood lipids. *Nature* 466, 707-713.
- 626 10. Willer, C.J., Schmidt, E.M., Sengupta, S., Peloso, G.M., Gustafsson, S., Kanoni, S., Ganna,
627 A., Chen, J., Buchkovich, M.L., Mora, S., et al. (2013). Discovery and refinement of loci
628 associated with lipid levels. *Nat Genet* 45, 1274-1283.
- 629 11. Wu, Y., Waite, L.L., Jackson, A.U., Sheu, W.H., Buyske, S., Absher, D., Arnett, D.K.,
630 Boerwinkle, E., Bonnycastle, L.L., Carty, C.L., et al. (2013). Trans-ethnic fine-mapping of
631 lipid loci identifies population-specific signals and allelic heterogeneity that increases the
632 trait variance explained. *PLoS Genet* 9, e1003379.
- 633 12. Klarin, D., Damrauer, S.M., Cho, K., Sun, Y.V., Teslovich, T.M., Honerlaw, J., Gagnon, D.R.,
634 DuVall, S.L., Li, J., Peloso, G.M., et al. (2018). Genetics of blood lipids among ~300,000
635 multi-ethnic participants of the Million Veteran Program. *Nat Genet* 50, 1514-1523.
- 636 13. Weissglas-Volkov, D., Aguilar-Salinas, C.A., Nikkola, E., Deere, K.A., Cruz-Bautista, I.,
637 Arellano-Campos, O., Munoz-Hernandez, L.L., Gomez-Munguia, L., Ordonez-Sanchez,
638 M.L., Reddy, P.M., et al. (2013). Genomic study in Mexicans identifies a new locus for
639 triglycerides and refines European lipid loci. *J Med Genet* 50, 298-308.
- 640 14. Ko, A., Cantor, R.M., Weissglas-Volkov, D., Nikkola, E., Reddy, P.M., Sinsheimer, J.S.,
641 Pasaniuc, B., Brown, R., Alvarez, M., Rodriguez, A., et al. (2014). Amerindian-specific
642 regions under positive selection harbour new lipid variants in Latinos. *Nat Commun* 5,
643 3983.
- 644 15. Below, J.E., Parra, E.J., Gamazon, E.R., Torres, J., Krithika, S., Candille, S., Lu, Y.,
645 Manichakul, A., Peralta-Romero, J., Duan, Q., et al. (2016). Meta-analysis of lipid-traits in

- 646 Hispanics identifies novel loci, population-specific effects, and tissue-specific enrichment
647 of eQTLs. *Sci Rep* 6, 19429.
- 648 16. Graff, M., Emery, L.S., Justice, A.E., Parra, E., Below, J.E., Palmer, N.D., Gao, C., Duan, Q.,
649 Valladares-Salgado, A., Cruz, M., et al. (2017). Genetic architecture of lipid traits in the
650 Hispanic community health study/study of Latinos. *Lipids Health Dis* 16, 200.
- 651 17. Villarreal-Molina, M.T., Aguilar-Salinas, C.A., Rodriguez-Cruz, M., Riano, D., Villalobos-
652 Comparan, M., Coral-Vazquez, R., Menjivar, M., Yescas-Gomez, P., Konigsoerg-
653 Fainstein, M., Romero-Hidalgo, S., et al. (2007). The ATP-binding cassette transporter
654 A1 R230C variant affects HDL cholesterol levels and BMI in the Mexican population:
655 association with obesity and obesity-related comorbidities. *Diabetes* 56, 1881-1887.
- 656 18. Acuna-Alonzo, V., Flores-Dorantes, T., Kruit, J.K., Villarreal-Molina, T., Arellano-Campos,
657 O., Hunemeier, T., Moreno-Estrada, A., Ortiz-Lopez, M.G., Villamil-Ramirez, H., Leon-
658 Mimila, P., et al. (2010). A functional ABCA1 gene variant is associated with low HDL-
659 cholesterol levels and shows evidence of positive selection in Native Americans. *Hum*
660 *Mol Genet* 19, 2877-2885.
- 661 19. Voight, B.F., Peloso, G.M., Orho-Melander, M., Frikke-Schmidt, R., Barbalic, M., Jensen,
662 M.K., Hindy, G., Holm, H., Ding, E.L., Johnson, T., et al. (2012). Plasma HDL cholesterol
663 and risk of myocardial infarction: a mendelian randomisation study. *Lancet* 380, 572-580.
- 664 20. Holmes, M.V., Asselbergs, F.W., Palmer, T.M., Drenos, F., Lanktree, M.B., Nelson, C.P.,
665 Dale, C.E., Padmanabhan, S., Finan, C., Swerdlow, D.I., et al. (2015). Mendelian
666 randomization of blood lipids for coronary heart disease. *Eur Heart J* 36, 539-550.
- 667 21. Kawashiri, M.A., Tada, H., Nomura, A., and Yamagishi, M. (2018). Mendelian randomization:
668 Its impact on cardiovascular disease. *J Cardiol* 72, 307-313.
- 669 22. Khera, A.V., Cuchel, M., de la Llera-Moya, M., Rodrigues, A., Burke, M.F., Jafri, K., French,
670 B.C., Phillips, J.A., Mucksavage, M.L., Wilensky, R.L., et al. (2011). Cholesterol efflux
671 capacity, high-density lipoprotein function, and atherosclerosis. *N Engl J Med* 364, 127-
672 135.
- 673 23. Solovieff, N., Cotsapas, C., Lee, P.H., Purcell, S.M., and Smoller, J.W. (2013). Pleiotropy in
674 complex traits: challenges and strategies. *Nat Rev Genet* 14, 483-495.
- 675 24. Kosmas, C.E., Christodoulidis, G., Cheng, J.W., Vittorio, T.J., and Lerakis, S. (2014). High-
676 density lipoprotein functionality in coronary artery disease. *Am J Med Sci* 347, 504-508.
- 677 25. Burgess, S., Foley, C.N., Allara, E., Staley, J.R., and Howson, J.M.M. (2020). A robust and
678 efficient method for Mendelian randomization with hundreds of genetic variants. *Nat*
679 *Commun* 11, 376.
- 680 26. Aguilar-Salinas, C.A., Olaiz, G., Valles, V., Torres, J.M., Gomez Perez, F.J., Rull, J.A.,
681 Rojas, R., Franco, A., and Sepulveda, J. (2001). High prevalence of low HDL cholesterol
682 concentrations and mixed hyperlipidemia in a Mexican nationwide survey. *J Lipid Res* 42,
683 1298-1307.
- 684 27. Aguilar-Salinas, C.A., Canizales-Quinteros, S., Rojas-Martinez, R., Mehta, R., Villarreal-
685 Molina, M.T., Arellano-Campos, O., Riba, L., Gomez-Perez, F.J., and Tusie-Luna, M.T.
686 (2009). Hypoalphalipoproteinemia in populations of Native American ancestry: an
687 opportunity to assess the interaction of genes and the environment. *Curr Opin Lipidol* 20,
688 92-97.
- 689 28. Hernandez-Alcaraz, C., Aguilar-Salinas, C.A., Mendoza-Herrera, K., Pedroza-Tobias, A.,
690 Villalpando, S., Shamah-Levy, T., Rivera-Dommarco, J., Hernandez-Avila, M., and
691 Barquera, S. (2020). Dyslipidemia prevalence, awareness, treatment and control in
692 Mexico: results of the Ensanut 2012. *Salud Publica Mex* 62, 137-146.
- 693 29. Villarreal-Molina, T., Posadas-Romero, C., Romero-Hidalgo, S., Antunez-Arguelles, E.,
694 Bautista-Grande, A., Vargas-Alarcon, G., Kimura-Hayama, E., Canizales-Quinteros, S.,
695 Juarez-Rojas, J.G., Posadas-Sanchez, R., et al. (2012). The ABCA1 gene R230C variant

- 696 is associated with decreased risk of premature coronary artery disease: the genetics of
697 atherosclerotic disease (GEA) study. *PLoS One* 7, e49285.
- 698 30. Mendez-Acevedo, K.M., Valdes, V.J., Asanov, A., and Vaca, L. (2017). A novel family of
699 mammalian transmembrane proteins involved in cholesterol transport. *Sci Rep* 7, 7450.
- 700 31. Leon-Mimila, P., Villamil-Ramirez, H., Villalobos-Comparan, M., Villarreal-Molina, T.,
701 Romero-Hidalgo, S., Lopez-Contreras, B., Gutierrez-Vidal, R., Vega-Badillo, J., Jacobo-
702 Albavera, L., Posadas-Romeros, C., et al. (2013). Contribution of common genetic
703 variants to obesity and obesity-related traits in mexican children and adults. *PLoS One* 8,
704 e70640.
- 705 32. Macias-Kauffer, L.R., Villamil-Ramirez, H., Leon-Mimila, P., Jacobo-Albavera, L., Posadas-
706 Romero, C., Posadas-Sanchez, R., Lopez-Contreras, B.E., Moran-Ramos, S., Romero-
707 Hidalgo, S., Acuna-Alonzo, V., et al. (2019). Genetic contributors to serum uric acid
708 levels in Mexicans and their effect on premature coronary artery disease. *Int J Cardiol*
709 279, 168-173.
- 710 33. (2000). Obesity: preventing and managing the global epidemic. Report of a WHO
711 consultation. *World Health Organ Tech Rep Ser* 894, i-xii, 1-253.
- 712 34. Kuczmarski, R.J., Ogden, C.L., Grummer-Strawn, L.M., Flegal, K.M., Guo, S.S., Wei, R.,
713 Mei, Z., Curtin, L.R., Roche, A.F., and Johnson, C.L. (2000). CDC growth charts: United
714 States. *Adv Data*, 1-27.
- 715 35. Leon-Mimila, P., Vega-Badillo, J., Gutierrez-Vidal, R., Villamil-Ramirez, H., Villareal-Molina,
716 T., Larrieta-Carrasco, E., Lopez-Contreras, B.E., Kauffer, L.R., Maldonado-Pintado, D.G.,
717 Mendez-Sanchez, N., et al. (2015). A genetic risk score is associated with hepatic
718 triglyceride content and non-alcoholic steatohepatitis in Mexicans with morbid obesity.
719 *Exp Mol Pathol* 98, 178-183.
- 720 36. Friedewald, W.T., Levy, R.I., and Fredrickson, D.S. (1972). Estimation of the concentration
721 of low-density lipoprotein cholesterol in plasma, without use of the preparative
722 ultracentrifuge. *Clin Chem* 18, 499-502.
- 723 37. DeLong, D.M., DeLong, E.R., Wood, P.D., Lippel, K., and Rifkind, B.M. (1986). A
724 comparison of methods for the estimation of plasma low- and very low-density lipoprotein
725 cholesterol. *The Lipid Research Clinics Prevalence Study. JAMA* 256, 2372-2377.
- 726 38. National Cholesterol Education Program Expert Panel on Detection, E., and Treatment of
727 High Blood Cholesterol in, A. (2002). Third Report of the National Cholesterol Education
728 Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood
729 Cholesterol in Adults (Adult Treatment Panel III) final report. *Circulation* 106, 3143-3421.
- 730 39. Parker, B.L., Calkin, A.C., Seldin, M.M., Keating, M.F., Tarling, E.J., Yang, P., Moody, S.C.,
731 Liu, Y., Zerenturk, E.J., Needham, E.J., et al. (2019). An integrative systems genetic
732 analysis of mammalian lipid metabolism. *Nature* 567, 187-193.
- 733 40. Huynh, K., Barlow, C.K., Jayawardana, K.S., Weir, J.M., Mellett, N.A., Cinel, M., Magliano,
734 D.J., Shaw, J.E., Drew, B.G., and Meikle, P.J. (2019). High-Throughput Plasma
735 Lipidomics: Detailed Mapping of the Associations with Cardiometabolic Risk Factors. *Cell*
736 *Chem Biol* 26, 71-84 e74.
- 737 41. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J.,
738 Sklar, P., de Bakker, P.I., Daly, M.J., et al. (2007). PLINK: a tool set for whole-genome
739 association and population-based linkage analyses. *Am J Hum Genet* 81, 559-575.
- 740 42. Browning, B.L., and Browning, S.R. (2016). Genotype Imputation with Millions of Reference
741 Samples. *Am J Hum Genet* 98, 116-126.
- 742 43. Corbin, L.J., Richmond, R.C., Wade, K.H., Burgess, S., Bowden, J., Smith, G.D., and
743 Timpson, N.J. (2016). BMI as a Modifiable Risk Factor for Type 2 Diabetes: Refining and
744 Understanding Causal Estimates Using Mendelian Randomization. *Diabetes* 65, 3002-
745 3007.

- 746 44. Burgess, S., Butterworth, A., and Thompson, S.G. (2013). Mendelian randomization analysis
747 with multiple genetic variants using summarized data. *Genet Epidemiol* 37, 658-665.
- 748 45. Bowden, J., Davey Smith, G., and Burgess, S. (2015). Mendelian randomization with invalid
749 instruments: effect estimation and bias detection through Egger regression. *Int J*
750 *Epidemiol* 44, 512-525.
- 751 46. Yavorska, O.O., and Burgess, S. (2017). MendelianRandomization: an R package for
752 performing Mendelian randomization analyses using summarized data. *Int J Epidemiol*
753 46, 1734-1739.
- 754 47. Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of
755 ancestry in unrelated individuals. *Genome Res* 19, 1655-1664.
- 756 48. Maples, B.K., Gravel, S., Kenny, E.E., and Bustamante, C.D. (2013). RFMix: a discriminative
757 modeling approach for rapid and robust local-ancestry inference. *Am J Hum Genet* 93,
758 278-288.
- 759 49. Sabeti, P.C., Reich, D.E., Higgins, J.M., Levine, H.Z., Richter, D.J., Schaffner, S.F., Gabriel,
760 S.B., Platko, J.V., Patterson, N.J., McDonald, G.J., et al. (2002). Detecting recent
761 positive selection in the human genome from haplotype structure. *Nature* 419, 832-837.
- 762 50. Szpiech, Z.A., and Hernandez, R.D. (2014). selscan: an efficient multithreaded program to
763 perform EHH-based scans for positive selection. *Mol Biol Evol* 31, 2824-2827.
- 764 51. Bennett, B.J., Davis, R.C., Civelek, M., Orozco, L., Wu, J., Qi, H., Pan, C., Packard, R.R.,
765 Eskin, E., Yan, M., et al. (2015). Genetic Architecture of Atherosclerosis in Mice: A
766 Systems Genetics Analysis of Common Inbred Strains. *PLoS Genet* 11, e1005711.
- 767 52. Puppione, D.L., and Charugundla, S. (1994). A microprecipitation technique suitable for
768 measuring alpha-lipoprotein cholesterol. *Lipids* 29, 595-597.
- 769 53. Hedrick, C.C., Castellani, L.W., Warden, C.H., Puppione, D.L., and Lusis, A.J. (1993).
770 Influence of mouse apolipoprotein A-II on plasma lipoproteins in transgenic mice. *J Biol*
771 *Chem* 268, 20676-20682.
- 772 54. Ghazalpour, A., Bennett, B., Petyuk, V.A., Orozco, L., Hagopian, R., Mungrue, I.N., Farber,
773 C.R., Sinsheimer, J., Kang, H.M., Furlotte, N., et al. (2011). Comparative analysis of
774 proteome and transcriptome variation in mouse. *PLoS Genet* 7, e1001393.
- 775 55. Bennett, B.J., Farber, C.R., Orozco, L., Kang, H.M., Ghazalpour, A., Siemers, N., Neubauer,
776 M., Neuhaus, I., Yordanova, R., Guan, B., et al. (2010). A high-resolution association
777 mapping panel for the dissection of complex traits in mice. *Genome Res* 20, 281-290.
- 778 56. Leek, J.T., Johnson, W.E., Parker, H.S., Jaffe, A.E., and Storey, J.D. (2012). The sva
779 package for removing batch effects and other unwanted variation in high-throughput
780 experiments. *Bioinformatics* 28, 882-883.
- 781 57. Hui, S.T., Kurt, Z., Tuominen, I., Norheim, F., R, C.D., Pan, C., Dirks, D.L., Magyar, C.E.,
782 French, S.W., Chella Krishnan, K., et al. (2018). The Genetic Architecture of Diet-
783 Induced Hepatic Fibrosis in Mice. *Hepatology* 68, 2182-2196.
- 784 58. Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013).
785 TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions
786 and gene fusions. *Genome Biol* 14, R36.
- 787 59. Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg,
788 S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-
789 Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat*
790 *Biotechnol* 28, 511-515.
- 791 60. Willer, C.J., Li, Y., and Abecasis, G.R. (2010). METAL: fast and efficient meta-analysis of
792 genomewide association scans. *Bioinformatics* 26, 2190-2191.
- 793 61. Yang, J., Lee, S.H., Goddard, M.E., and Visscher, P.M. (2011). GCTA: a tool for genome-
794 wide complex trait analysis. *Am J Hum Genet* 88, 76-82.

- 795 62. Rucker, G., Schwarzer, G., Carpenter, J.R., Binder, H., and Schumacher, M. (2011).
796 Treatment-effect estimates adjusted for small-study effects via a limit meta-analysis.
797 *Biostatistics* 12, 122-142.
- 798 63. Pruim, R.J., Welch, R.P., Sanna, S., Teslovich, T.M., Chines, P.S., Gliedt, T.P., Boehnke,
799 M., Abecasis, G.R., and Willer, C.J. (2010). LocusZoom: regional visualization of
800 genome-wide association scan results. *Bioinformatics* 26, 2336-2337.
- 801 64. Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation
802 network analysis. *BMC Bioinformatics* 9, 559.
- 803 65. Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A.H., Tanaseichuk, O., Benner, C.,
804 and Chanda, S.K. (2019). Metascape provides a biologist-oriented resource for the
805 analysis of systems-level datasets. *Nat Commun* 10, 1523.
- 806 66. Chen, J., Bardes, E.E., Aronow, B.J., and Jegga, A.G. (2009). ToppGene Suite for gene list
807 enrichment analysis and candidate gene prioritization. *Nucleic Acids Res* 37, W305-311.
- 808 67. Reimand, J., Isserlin, R., Voisin, V., Kucera, M., Tannus-Lopes, C., Rostamianfar, A., Wadi,
809 L., Meyer, M., Wong, J., Xu, C., et al. (2019). Pathway enrichment analysis and
810 visualization of omics data using g:Profiler, GSEA, Cytoscape and EnrichmentMap. *Nat*
811 *Protoc* 14, 482-517.
- 812 68. Kontush, A., Lhomme, M., and Chapman, M.J. (2013). Unraveling the complexities of the
813 HDL lipidome. *J Lipid Res* 54, 2950-2963.
- 814 69. Ding, M., and Rexrode, K.M. (2020). A Review of Lipidomics of Cardiovascular Disease
815 Highlights the Importance of Isolating Lipoproteins. *Metabolites* 10.
- 816 70. Rojas, R., Aguilar-Salinas, C.A., Jimenez-Corona, A., Shamah-Levy, T., Rauda, J., Avila-
817 Burgos, L., Villalpando, S., and Ponce, E.L. (2010). Metabolic syndrome in Mexican
818 adults: results from the National Health and Nutrition Survey 2006. *Salud Publica Mex* 52
819 *Suppl* 1, S11-18.
- 820 71. Oliveira, H.C.F., and Raposo, H.F. (2020). Cholesteryl Ester Transfer Protein and Lipid
821 Metabolism and Cardiovascular Diseases. *Adv Exp Med Biol* 1276, 15-25.
- 822 72. Ponce, X., Rodriguez-Ramirez, S., Mundo-Rosas, V., Shamah, T., Barquera, S., and
823 Gonzalez de Cossio, T. (2014). Dietary quality indices vary with sociodemographic
824 variables and anthropometric status among Mexican adults: a cross-sectional study.
825 Results from the 2006 National Health and Nutrition Survey. *Public Health Nutr* 17, 1717-
826 1728.
- 827 73. Gombojav, B., Lee, S.J., Kho, M., Song, Y.M., Lee, K., and Sung, J. (2016). Multiple
828 susceptibility loci at chromosome 11q23.3 are associated with plasma triglyceride in East
829 Asians. *J Lipid Res* 57, 318-324.
- 830 74. Moon, S., Lee, Y., Won, S., and Lee, J. (2018). Multiple genotype-phenotype association
831 study reveals intronic variant pair on *SIRT2* associated with metabolic syndrome in a
832 Korean population. *Hum Genomics* 12, 48.
- 833 75. Prats-Urbe, A., Sayols-Baixeras, S., Fernandez-Sanles, A., Subirana, I., Carreras-Torres,
834 R., Vilahur, G., Civeira, F., Marrugat, J., Fito, M., Hernaez, A., et al. (2020). High-density
835 lipoprotein characteristics and coronary artery disease: a Mendelian randomization study.
836 *Metabolism* 112, 154351.
- 837 76. Karjalainen, M.K., Holmes, M.V., Wang, Q., Anufrieva, O., Kahonen, M., Lehtimaki, T.,
838 Havulinna, A.S., Kristiansson, K., Salomaa, V., Perola, M., et al. (2020). Apolipoprotein
839 A-I concentrations and risk of coronary artery disease: A Mendelian randomization study.
840 *Atherosclerosis* 299, 56-63.
- 841 77. Richardson, T.G., Sanderson, E., Palmer, T.M., Ala-Korpela, M., Ference, B.A., Davey
842 Smith, G., and Holmes, M.V. (2020). Evaluating the relationship between circulating
843 lipoprotein lipids and apolipoproteins with risk of coronary heart disease: A multivariable
844 Mendelian randomisation analysis. *PLoS Med* 17, e1003062.

- 845 78. Ference, B.A., Kastelein, J.J.P., Ray, K.K., Ginsberg, H.N., Chapman, M.J., Packard, C.J.,
846 Laufs, U., Oliver-Williams, C., Wood, A.M., Butterworth, A.S., et al. (2019). Association of
847 Triglyceride-Lowering LPL Variants and LDL-C-Lowering LDLR Variants With Risk of
848 Coronary Heart Disease. *JAMA* 321, 364-373.
- 849 79. Zuber, V., Gill, D., Ala-Korpela, M., Langenberg, C., Butterworth, A., Bottolo, L., and
850 Burgess, S. (2020). High-throughput multivariable Mendelian randomization analysis
851 prioritizes apolipoprotein B as key lipid risk factor for coronary artery disease. *medRxiv*,
852 2020.2002.2010.20021691.
- 853 80. Marmillot, P., Patel, S., and Lakshman, M.R. (2007). Reverse cholesterol transport is
854 regulated by varying fatty acyl chain saturation and sphingomyelin content in
855 reconstituted high-density lipoproteins. *Metabolism* 56, 251-259.
- 856 81. Zerrad-Saadi, A., Therond, P., Chantepie, S., Couturier, M., Rye, K.A., Chapman, M.J., and
857 Kontush, A. (2009). HDL3-mediated inactivation of LDL-associated phospholipid
858 hydroperoxides is determined by the redox status of apolipoprotein A-I and HDL particle
859 surface lipid rigidity: relevance to inflammation and atherogenesis. *Arterioscler Thromb*
860 *Vasc Biol* 29, 2169-2175.
- 861 82. Hancock-Cerutti, W., Lhomme, M., Dauteuille, C., Lecocq, S., Chapman, M.J., Rader, D.J.,
862 Kontush, A., and Cuchel, M. (2017). Paradoxical coronary artery disease in humans with
863 hyperalphalipoproteinemia is associated with distinct differences in the high-density
864 lipoprotein phosphosphingolipidome. *J Clin Lipidol* 11, 1192-1200 e1193.
- 865 83. Camont, L., Lhomme, M., Rached, F., Le Goff, W., Negre-Salvayre, A., Salvayre, R.,
866 Calzada, C., Lagarde, M., Chapman, M.J., and Kontush, A. (2013). Small, dense high-
867 density lipoprotein-3 particles are enriched in negatively charged phospholipids:
868 relevance to cellular cholesterol efflux, antioxidative, antithrombotic, anti-inflammatory,
869 and antiapoptotic functionalities. *Arterioscler Thromb Vasc Biol* 33, 2715-2723.
- 870 84. Holzer, M., Birner-Gruenberger, R., Stojakovic, T., El-Gamal, D., Binder, V., Wadsack, C.,
871 Heinemann, A., and Marsche, G. (2011). Uremia alters HDL composition and function. *J*
872 *Am Soc Nephrol* 22, 1631-1641.
- 873 85. Darabi, M., and Kontush, A. (2016). Phosphatidylserine in atherosclerosis. *Curr Opin Lipidol*
874 27, 414-420.
- 875 86. Jialin, G., Xuefan, G., and Huiwen, Z. (2010). SID1 transmembrane family, member 2
876 (Sidt2): a novel lysosomal membrane protein. *Biochem Biophys Res Commun* 402, 588-
877 594.
- 878 87. Ren, R., Xu, X., Lin, T., Weng, S., Liang, H., Huang, M., Dong, C., Luo, Y., and He, J.
879 (2011). Cloning, characterization, and biological function analysis of the SidT2 gene from
880 *Siniperca chuatsi*. *Dev Comp Immunol* 35, 692-701.
- 881 88. Li, W., Koutmou, K.S., Leahy, D.J., and Li, M. (2015). Systemic RNA Interference
882 Deficiency-1 (SID-1) Extracellular Domain Selectively Binds Long Double-stranded RNA
883 and Is Required for RNA Transport by SID-1. *J Biol Chem* 290, 18904-18913.
- 884 89. Aizawa, S., Contu, V.R., Fujiwara, Y., Hase, K., Kikuchi, H., Kabuta, C., Wada, K., and
885 Kabuta, T. (2017). Lysosomal membrane protein SIDT2 mediates the direct uptake of
886 DNA by lysosomes. *Autophagy* 13, 218-222.
- 887 90. Nguyen, T.A., Smith, B.R.C., Elgass, K.D., Creed, S.J., Cheung, S., Tate, M.D., Belz, G.T.,
888 Wicks, I.P., Masters, S.L., and Pang, K.C. (2019). SIDT1 Localizes to Endolysosomes
889 and Mediates Double-Stranded RNA Transport into the Cytoplasm. *J Immunol* 202,
890 3483-3492.
- 891 91. Hase, K., Contu, V.R., Kabuta, C., Sakai, R., Takahashi, M., Kataoka, N., Hakuno, F.,
892 Takahashi, S.I., Fujiwara, Y., Wada, K., et al. (2020). Cytosolic domain of SIDT2 carries
893 an arginine-rich motif that binds to RNA/DNA and is important for the direct transport of
894 nucleic acids into lysosomes. *Autophagy*, 1-15.

- 895 92. Valdes, V.J., Athie, A., Salinas, L.S., Navarro, R.E., and Vaca, L. (2012). CUP-1 is a novel
896 protein involved in dietary cholesterol uptake in *Caenorhabditis elegans*. *PLoS One* 7,
897 e33962.
- 898 93. Hulce, J.J., Coggnetta, A.B., Niphakis, M.J., Tully, S.E., and Cravatt, B.F. (2013). Proteome-
899 wide mapping of cholesterol-interacting proteins in mammalian cells. *Nat Methods* 10,
900 259-264.
- 901 94. Sharpe, L.J., Rao, G., Jones, P.M., Glancey, E., Aleidi, S.M., George, A.M., Brown, A.J., and
902 Gelissen, I.C. (2015). Cholesterol sensing by the ABCG1 lipid transporter: Requirement
903 of a CRAC motif in the final transmembrane domain. *Biochim Biophys Acta* 1851, 956-
904 964.
- 905 95. Ouimet, M., Franklin, V., Mak, E., Liao, X., Tabas, I., and Marcel, Y.L. (2011). Autophagy
906 regulates cholesterol efflux from macrophage foam cells via lysosomal acid lipase. *Cell*
907 *Metab* 13, 655-667.
- 908 96. Chen, X., Gu, X., and Zhang, H. (2018). Sidt2 regulates hepatocellular lipid metabolism
909 through autophagy. *J Lipid Res* 59, 404-415.
- 910 97. Gao, J., Gu, X., Mahuran, D.J., Wang, Z., and Zhang, H. (2013). Impaired glucose tolerance
911 in a mouse model of sidt2 deficiency. *PLoS One* 8, e66139.
- 912 98. Chang, G., Yang, R., Cao, Y., Nie, A., Gu, X., and Zhang, H. (2016). SIDT2 is involved in the
913 NAADP-mediated release of calcium from insulin secretory granules. *J Mol Endocrinol*
914 56, 249-259.
- 915 99. Gao, J., Zhang, Y., Yu, C., Tan, F., and Wang, L. (2016). Spontaneous nonalcoholic fatty
916 liver disease and ER stress in Sidt2 deficiency mice. *Biochem Biophys Res Commun*
917 476, 326-332.
- 918 100. Meng, Y., Wang, L., and Ling, L. (2018). Changes of lysosomal membrane permeabilization
919 and lipid metabolism in sidt2 deficient mice. *Exp Ther Med* 16, 246-252.
- 920
921
922

923 **FIGURE TITLES AND LEGENDS**

924
925 **Figure 1. Manhattan plot for HDL-C levels in the discovery phase.** Plot showing the $-\log_{10}$
926 transformed P -value of SNPs for 2153 Mexican children and adults. The red line indicates the
927 genome-wide significance level ($P=5 \times 10^{-8}$). Genes closest to the SNP with the lowest P -value at
928 each locus are indicated.

929
930 **Figure 2. Locus zoom view of variants within the 11q23 region (*SIDT2* locus) associated**
931 **with HDL-C levels.** SNPs are colored based on their correlation (r^2) with the *SIDT2*/Val636Ile
932 variant (purple diamond), which showed the strongest association with HDL-C levels ($P=1.5 \times 10^{-$
933 11). Arrows on the horizontal blue lines show the direction of transcription, and rectangles
934 represent exons. P -values indicate significance of associations found in the discovery phase.

935
936 **Figure 3. Heat map of associations between the *SIDT2*/Val636Ile variant and lipid classes**
937 **in the MOBES cohort.** Color intensity reflects the Beta value (red for positive, blue for negative)
938 obtained from linear regression between *SIDT2*/Val636Ile and lipid classes in the MOBES study
939 ($n=375$) adjusted by age, sex and BMI. * P -value ≤ 0.05 , ** P -value ≤ 0.001 .

940
941 **Figure 4. Uptake of the blue fluorescent cholesterol analog dehydroergosterol (DHE) by**
942 **cells expressing wildtype and Ile636/*SIDT2*.** (A) Confocal microscopy images of HEK293T
943 cells expressing the empty vector, wildtype and Ile636/*SIDT2*, at times 2.5, 3.5 and 5 minutes
944 after adding the fluorescent cholesterol analog DHE to the culture. (B) Plot showing mean
945 fluorescence intensity over time after adding DHE to the culture. Values show mean \pm standard
946 deviations from at least 4 independent experiments, each experiment shows mean values from
947 all the cells in the focal plane. The red dots represent cells transfected with the empty vector; the
948 black dots represent cells expressing wildtype *SIDT2*, and blue dots represent cells expressing
949 Ile636/*SIDT2*. Addition of DHE is indicated with an arrow. * $P < 0.05$; ** $P < 0.01$.

950
951 **Figure 5. Liver *SIDT2* expression and lipids in mice and humans.** Correlations of liver *SidT2*
952 expression with HDL-C, total cholesterol, and TG serum levels in MOBES cohort participants (A)
953 and mice from the HMDP (B). (C) Venn diagram depicting the overlap of genes significantly
954 correlated with *SIDT2* liver expression in HMDP mice and MOBES cohort participants
955 ($P < 1.0 \times 10^{-8}$). (D) Significantly enriched pathways in mice and humans.

956

957 **Video S1. Uptake of the fluorescent cholesterol analog dehydroergosterol (DHE) in**
958 **HEK293T cells expressing wildtype or Ile636/SIDT2.** Representative experiments illustrating
959 cells in the focal plane. Fluorescence was monitored in time using confocal microscopy. DHE
960 uptake in cells expressing the empty vector (upper panel), wildtype SIDT2 and the Ile636/SIDT2
961 proteins (lower panels).

Table 1. Lead SNPs associated with HDL-C levels in the meta-analysis including Mexican children and adults.

CHR: position (hg19)	Locus	rs ID	Type of variant	β HDL-C (SE)	P-value
6: 160921566	<i>LPAL2</i>	rs9457930	Intron variant	1.19 (0.33)	6.93 x 10 ⁻⁶
8: 9177732	<i>PRPF31, PPP1R3B, TNKS</i>	rs983309	Intron variant	-1.83 (0.34)	4.48 x 10 ⁻⁶
9: 107589134	<i>ABCA1</i>	rs4149310	Intron variant	1.44 (0.32)	7.25 x 10 ⁻⁷
9: 107620835	<i>ABCA1</i>	rs9282541	Missense variant	-3.44 (0.48)	3.99 x 10 ⁻¹³
11: 116633947	<i>BUD13-ZPR1-APOC3-APOA4-APOA5-APOA1</i>	rs10488698	Missense variant	2.38 (0.44)	8.17 x 10 ⁻⁸
11: 117063003	<i>SIDT2</i>	rs17120425	Missense variant	3.31 (0.51)	1.52 x 10 ⁻¹¹
12: 43679673	<i>ADAMTS20</i>	rs1514661	Regulatory region variant	1.12 (0.34)	4.23 x 10 ⁻⁶
15: 58723479	<i>LIPC, ALDH1A2</i>	rs1077834	Intron variant	-1.55 (0.33)	8.57 x 10 ⁻⁷
16: 56985555	<i>CETP</i>	rs12448528	Regulatory region variant	-2.79 (0.39)	5.92 x 10 ⁻¹⁵
16: 56999328	<i>CETP</i>	rs11508026	Intron variant	3.02 (0.32)	4.46 x 10 ⁻¹⁸

HDL-C values were log transformed for the analysis. Effect sizes were calculated for minor alleles. P-values for additive models were adjusted for sex, age, and either BMI or BMI percentile in children (n=2,153). HDL-C, High-density lipoprotein cholesterol; CHR: Chromosome; SE, Standard error.

Table 2. Association of selected SNPs with lipid traits in the GEA cohort.

CHR: position (hg19)	Locus	rs ID	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
6: 160921566	<i>LPAL2</i>	rs9457930	0.59 (0.49)	0.230	2.40 (1.55)	0.123	1.49 (1.31)	0.322	5.32 (6.24)	0.837	-0.02 (1.39)	0.920	1.16 (1.15)	0.286
8: 9177732	<i>PRPF31, PPP1R3B, TNKS</i>	rs983309	-2.32 (0.50)	4.0x10 ⁻⁶	-4.30 (1.60)	0.007	-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	<i>ABCA1</i>	rs4149310	1.25 (0.49)	0.010	3.68 (1.55)	0.018	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	<i>ABCA1</i>	rs9282541	-4.22 (0.98)	2.0x10 ⁻⁵	-6.29 (2.85)	0.027	-2.23 (2.53)	0.622	0.68 (8.33)	0.640	-11.02 (3.10)	5.3x10 ⁻⁵	-1.46 (2.28)	0.607
11: 116633947	<i>BUD13-ZPR1- APOC3-APOA4- APOA5-APOA1</i>	rs10488698	2.75 (0.69)	6.5x10 ⁻⁵	-4.17 (2.20)	0.058	-4.58 (1.85)	0.027	-16.94 (8.83)	0.048	4.65 (1.96)	0.018	-4.10 (1.63)	0.017
11: 117063003	<i>SIDT2</i>	rs17120425	3.35 (0.75)	7.0x10 ⁻⁶	-3.66 (2.31)	0.114	-6.22 (1.96)	0.004	-4.75 (8.75)	0.477	10.29 (2.14)	2.6x10 ⁻⁶	-3.92 (1.74)	0.038
12: 43679673	<i>ADAMTS20</i>	rs1514661	-0.12 (0.50)	0.805	-0.25 (1.57)	0.874	-0.03 (1.32)	0.773	0.43 (6.35)	0.492	-0.80 (1.41)	0.677	0.84 (1.16)	0.381
15: 58723479	<i>LIPC, ALDH1A2</i>	rs1077834	-0.59 (0.46)	0.200	-0.15 (1.46)	0.919	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	<i>CETP</i>	rs12448528	-3.77 (0.57)	7.2x10 ⁻¹¹	-6.02 (1.85)	0.001	-2.96 (1.57)	0.062	3.91 (7.48)	0.340	-5.26 (1.66)	2.1x10 ⁻⁴	-2.09 (1.39)	0.118
16: 56999328	<i>CETP</i>	rs11508026	2.79 (0.47)	2.9x10 ⁻⁹	2.25 (1.51)	0.136	-0.62 (1.28)	0.830	1.38 (6.06)	0.621	4.08 (1.35)	0.001	-0.49 (1.12)	0.979

LDL-C, TG, ApoA1 and ApoB levels were log transformed for the analysis. Effect sizes were calculated for the minor alleles. P-values for additive models were adjusted for sex, age, and BMI (n=1,559). CHR: Chromosome; MAF, minor allele frequency; HDL-C, High density lipoprotein cholesterol; TC, Total cholesterol; LDL-C, Low density lipoprotein cholesterol; TG, Triglycerides; ApoA1, Apolipoprotein A1; ApoB, Apolipoprotein B.

Table 3. Association of selected SNPs with lipid traits in the MOBES cohort.

CHR: position (hg19)	Locus	rs ID	HDL-C		TC		LDL-C		TG	
			B (SE)	P-value	B (SE)	P-value	B (SE)	P-value	B (SE)	P-value
6: 160921566	<i>LPAL2</i>	rs9457930	0.23 (0.64)	0.724	0.72 (2.89)	0.802	0.22 (2.36)	0.581	-2.49 (5.46)	0.758
8: 9177732	<i>PRPF31, PPP1R3B, TNKS</i>	rs983309	-0.38 (0.65)	0.559	-2.80 (2.81)	0.319	-3.54 (2.30)	0.255	6.25 (5.24)	0.236
9: 107589134	<i>ABCA1</i>	rs4149310	0.66 (0.60)	0.275	3.70 (2.62)	0.158	3.16 (2.15)	0.162	2.98 (4.88)	0.345
9: 107620835	<i>ABCA1</i>	rs9282541	-3.77 (0.96)	1.1x10 ⁻⁴	-8.78 (4.27)	0.040	-6.38 (3.50)	0.073	-4.26 (7.99)	0.844
11: 116633947	<i>BUD13-ZPR1- APOC3-APOA4- APOA5-APOA1</i>	rs10488698	0.81 (0.87)	0.350	2.02 (3.78)	0.594	2.21 (3.10)	0.175	-1.46 (7.05)	0.906
11: 117063003	<i>SIDT2</i>	rs17120425	2.92 (0.98)	0.003	7.79 (4.30)	0.071	3.72 (3.54)	0.226	9.15 (8.04)	0.196
12: 43679673	<i>ADAMTS20</i>	rs1514661	0.46 (0.62)	0.454	-0.87 (2.68)	0.746	-0.60 (2.20)	0.490	-0.81 (5.00)	0.930
15: 58723479	<i>LIPC, ALDH1A2</i>	rs1077834	-0.86 (0.57)	0.134	3.13 (2.50)	0.211	4.08 (2.05)	0.143	-2.57 (4.67)	0.264
16: 56985555	<i>CETP</i>	rs12448528	-1.77 (0.72)	0.014	-1.71 (3.14)	0.586	0.12 (2.58)	0.668	-0.29 (5.86)	0.999
16: 56999328	<i>CETP</i>	rs11508026	2.16 (0.55)	9.6x10 ⁻⁵	0.11 (2.44)	0.965	-1.14 (1.99)	0.695	-3.99 (4.54)	0.277

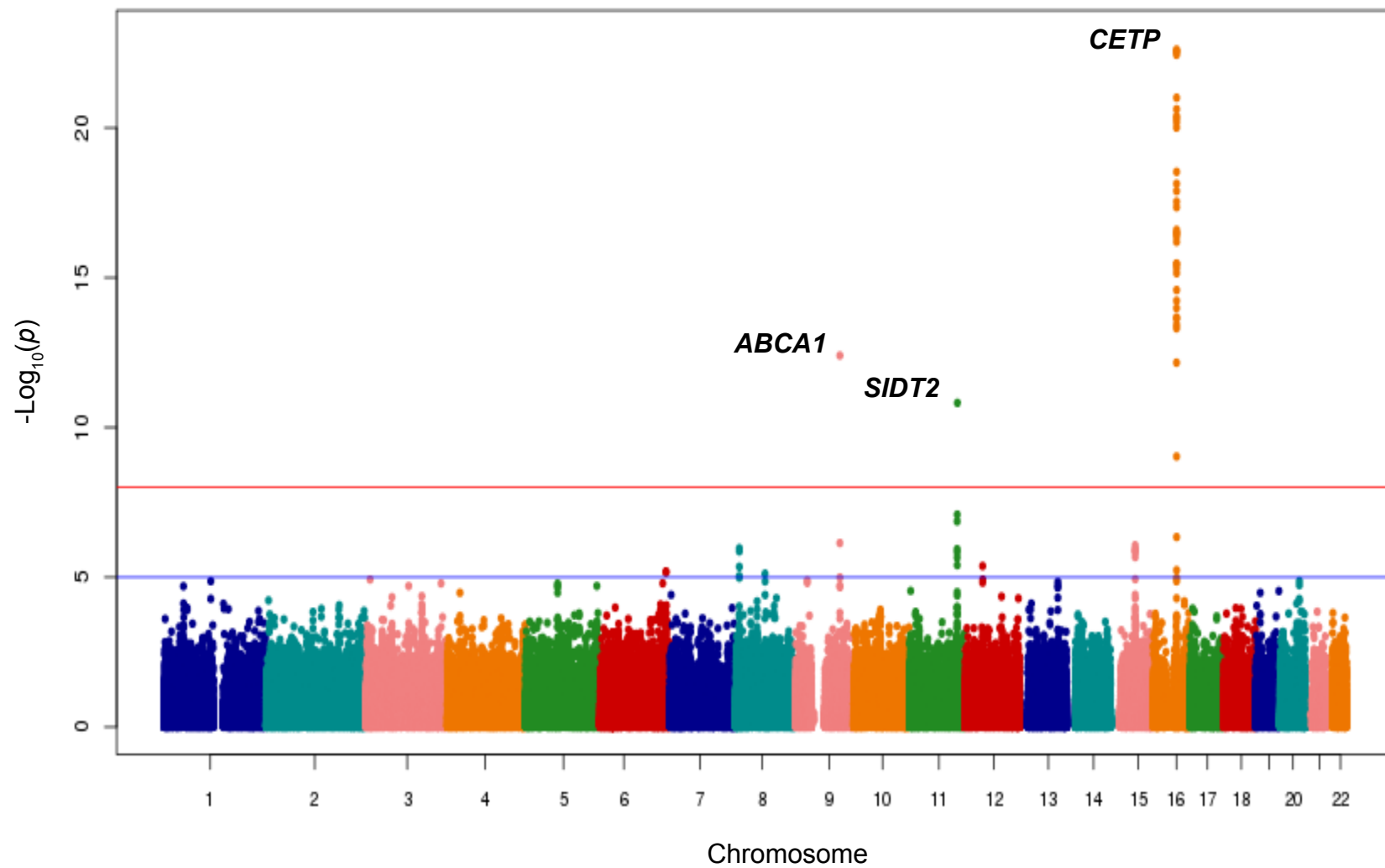
LDL-C and TG levels were log transformed for the analysis. Effect size was calculated for the minor alleles. P-values for additive models were adjusted for sex, age, and BMI (n=555). CHR: Chromosome; MAF, minor allele frequency; HDL-C, High-density lipoprotein cholesterol; TC, Total cholesterol; LDL-C, Low-density lipoprotein cholesterol; TG, Triglycerides.

Table 4. Association of HDL-C associated SNPs with premature coronary artery disease in the GEA cohort.

CHR: position (hg19)	Locus	rs ID	MAF		OR (95% CI)	P-value
			Controls (n=1,559)	Premature CAD (n=1,095)		
6: 160921566	<i>LPAL2</i>	rs9457930	37.9	37.0	0.98 (0.86-1.11)	0.704
8: 9177732	<i>PRPF31, PPP1R3B, TNKS</i>	rs983309	29.3	30.5	1.06 (0.93-1.21)	0.370
9: 107589134	<i>ABCA1</i>	rs4149310	34.4	36.6	1.11 (0.98-1.26)	0.116
9: 107620835	<i>ABCA1</i>	rs9282541	10.4	4.8	0.44 (0.30-0.65)	3.9x10 ⁻⁵
11: 116633947	<i>BUD13-ZPR1-APOC3- APOA4-APOA5-APOA1</i>	rs10488698	12.9	10.7	0.80 (0.67-0.97)	0.023
11: 117063003	<i>SIDT2</i>	rs17120425	9.5	8.0	0.80 (0.65-0.99)	0.039
12: 43679673	<i>ADAMTS20</i>	rs1514661	31.1	30.7	0.98 (0.87-1.12)	0.803
15: 58723479	<i>LIPC, ALDH1A2</i>	rs1077834	42.6	39.1	0.87 (0.77-0.98)	0.027
16: 56985555	<i>CETP</i>	rs12448528	20.5	18.1	0.87 (0.75-1.02)	0.087
16: 56999328	<i>CETP</i>	rs11508026	36.3	39.0	1.11 (0.99-1.26)	0.087

Associations with premature CAD were adjusted for age, sex and BMI. P-values were calculated by logistic regression. HDL-C, High-density lipoprotein cholesterol; CHR: Chromosome; MAF, minor allele frequency; CAD, coronary artery disease; OR, Odds ratio; CI, confidence interval.

Figure 1



Plotted SNPs

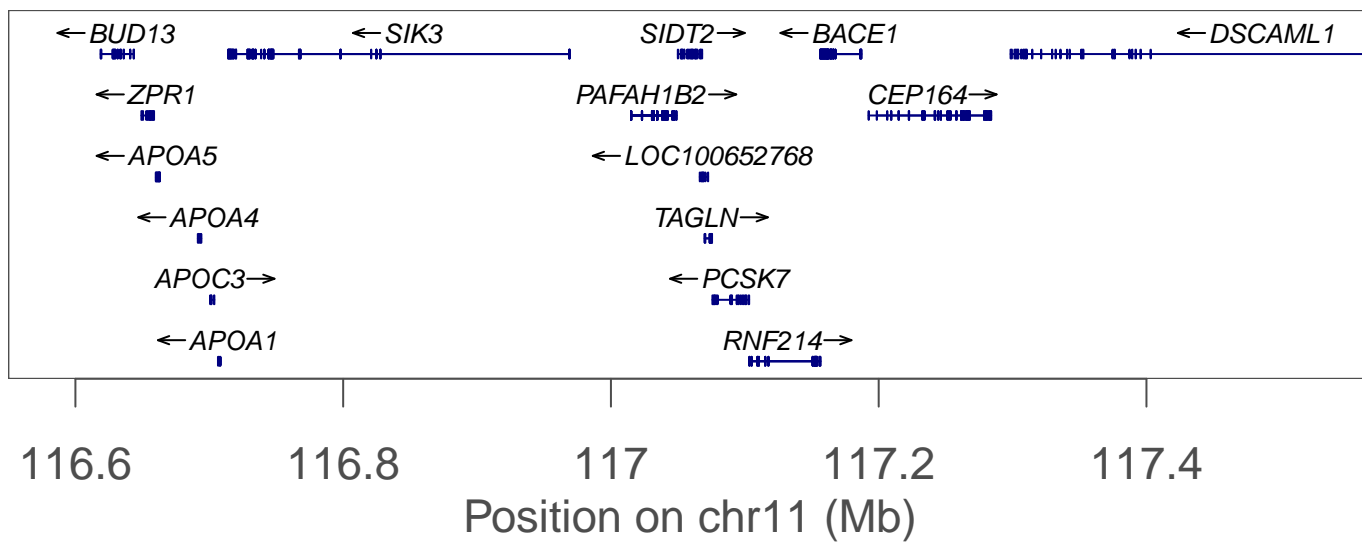
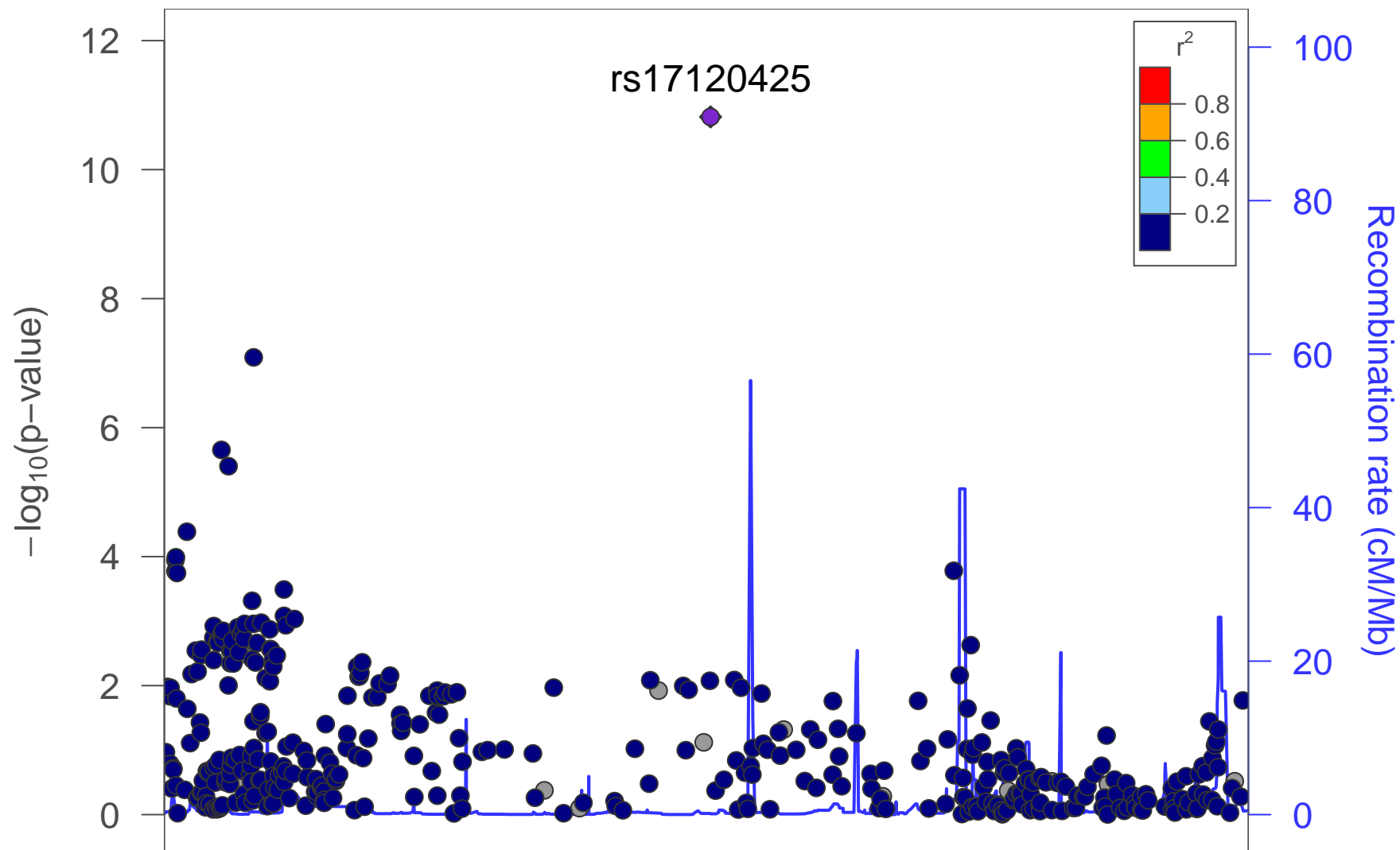


Figure 3

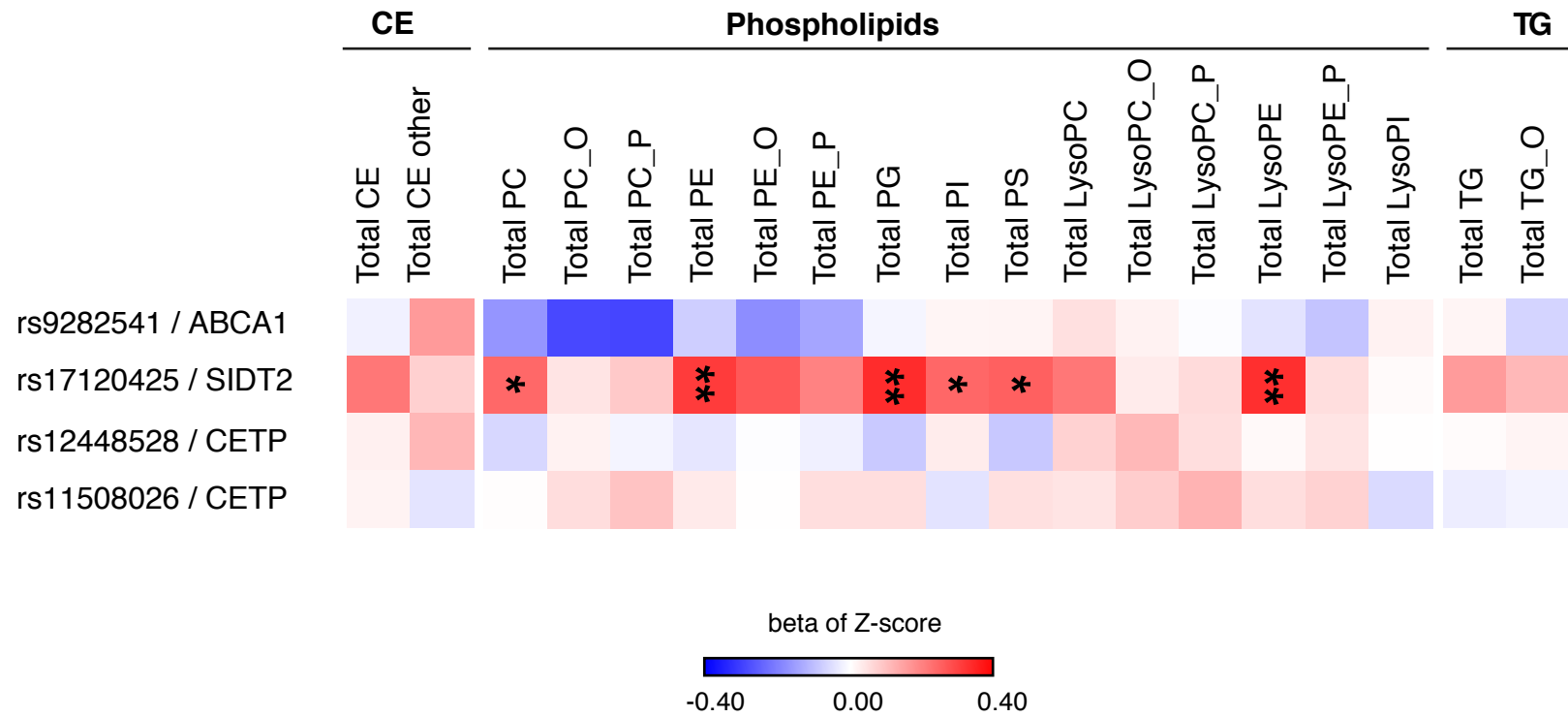
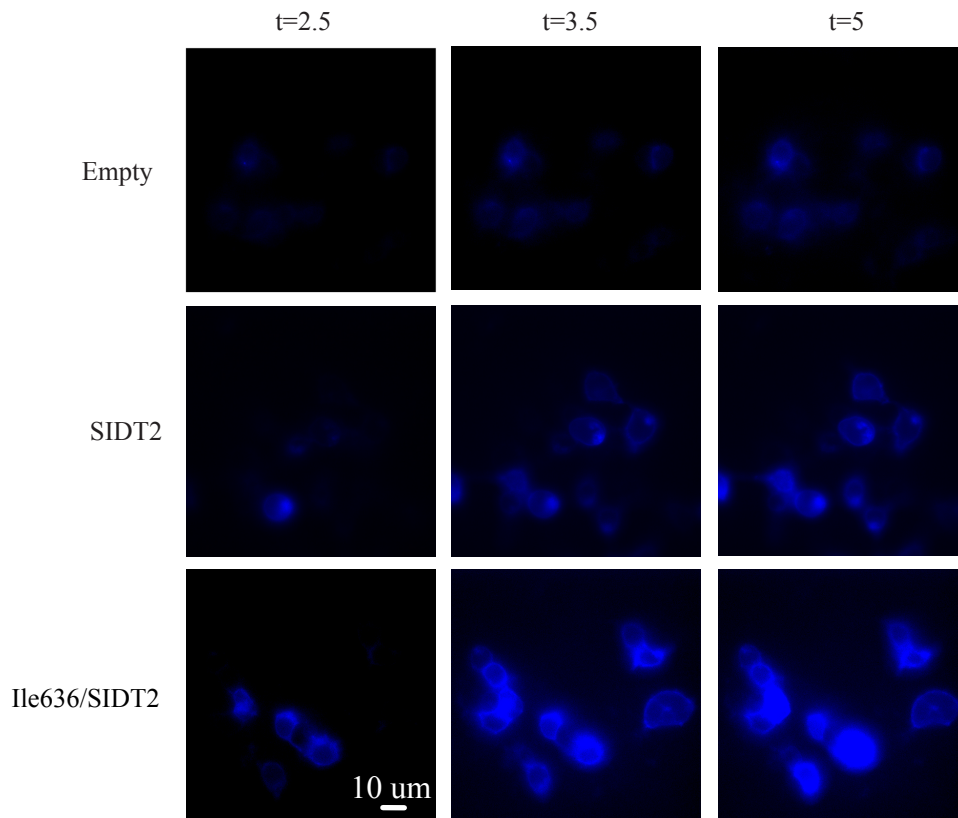


Figure 4

medRxiv preprint doi: <https://doi.org/10.1101/2020.09.19.20197673>; this version posted September 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

A.



B.

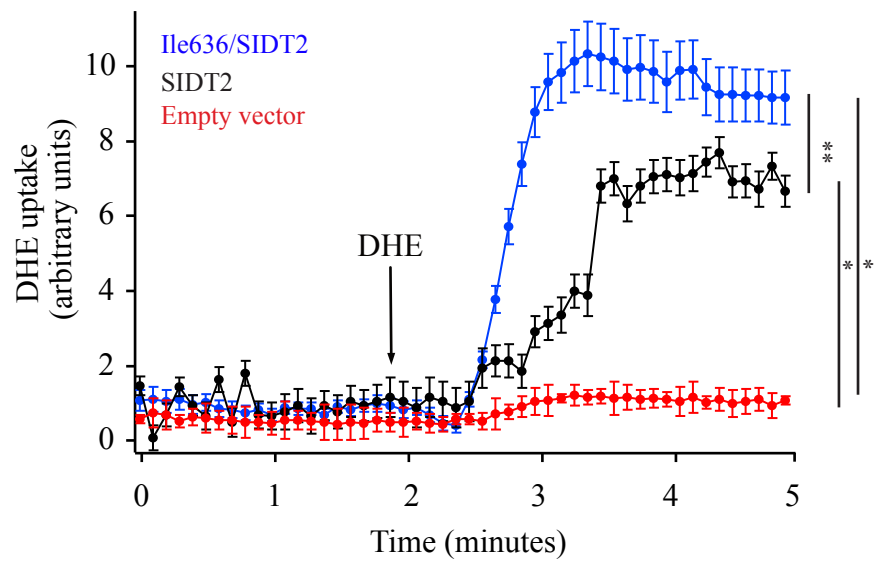
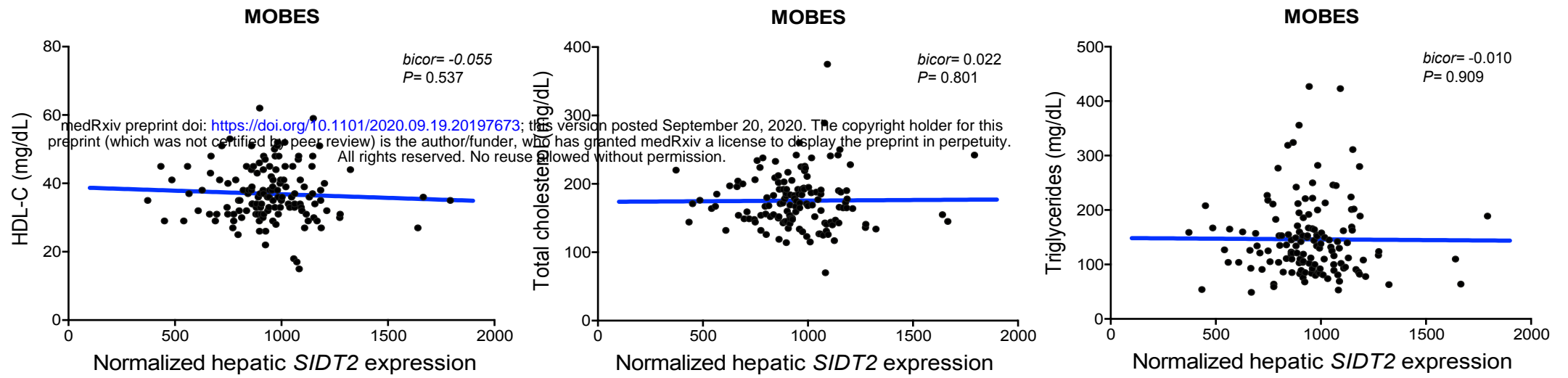
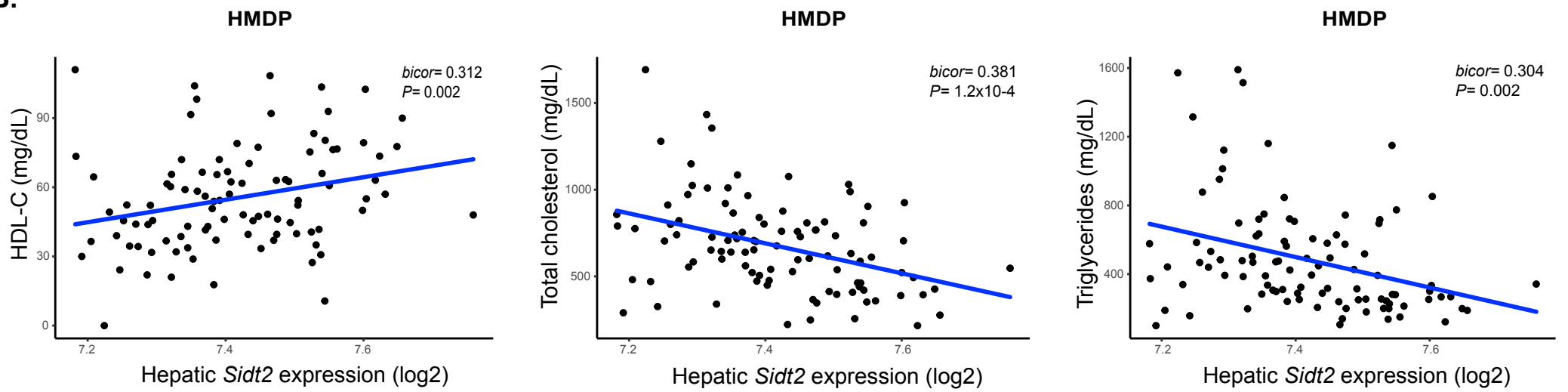


Figure 5

A.

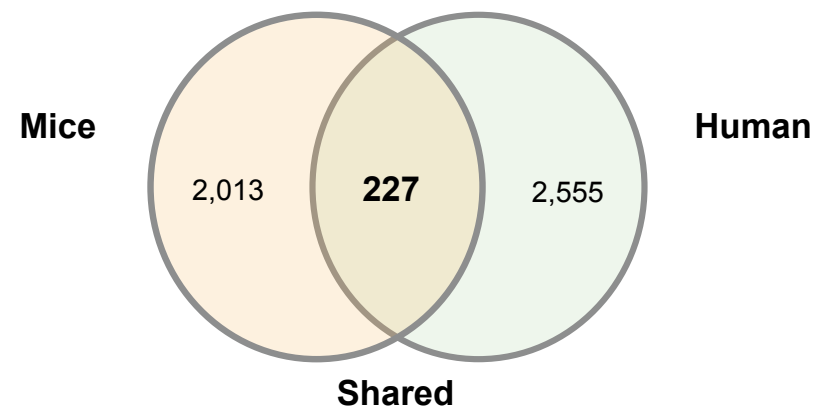


B.



C.

Pathway enrichment analysis (liver)



D.

