



Missense variants in *SORT1* are associated with LDL-C in an Amish population

Kelly A. Mitok¹, Kathryn L. Schueler¹, Sarah M. King², Joseph Orr², Kathleen A. Ryan³, Mark P. Keller¹, Ronald M. Krauss², Braxton D. Mitchell³, Alan R. Shuldiner^{3,4}, and Alan D. Attie^{1,*}

¹Department of Biochemistry, University of Wisconsin-Madison, Madison, WI, USA; ²Department of Pediatrics, University of California-San Francisco, San Francisco, CA, USA; ³Department of Medicine, University of Maryland School of Medicine, Baltimore, MD, USA; and ⁴Regeneron Genetics Center, Tarrytown, NY, USA

Abstract Common noncoding variants at the human *l*p13.3 locus associated with *SORT1* expression are among those most strongly associated with low-density lipoprotein cholesterol (LDL-C) in human genome-wide association studies. However, validation studies in mice and cell lines have produced variable results regarding the directionality of the effect of *SORT1* on LDL-C. This, together with the fact that the *l*p13.3 variants are associated with expression of several genes, has raised the question of whether *SORT1* is the causal gene at this locus. Using whole exome sequencing in members of an Amish population, we identified coding variants in *SORT1* that are associated with increased (rs141749679, K302E) and decreased (rs149456022, Q225H) LDL-C. Further, analysis of plasma lipoprotein particle subclasses by ion mobility in a subset of rs141749679 (K302E) carriers revealed higher levels of large LDL particles compared to noncarriers. In contrast to the effect of these variants in the Amish, the sortilin K302E mutation introduced into a C57BL/6J mouse via CRISPR/Cas9 resulted in decreased non-high-density lipoprotein cholesterol, and the sortilin Q225H mutation did not alter cholesterol levels in mice. This is indicative of different effects of these mutations on cholesterol metabolism in the two species. **Key words:** To our knowledge, this is the first evidence that naturally occurring coding variants in *SORT1* are associated with LDL-C, thus supporting *SORT1* as the gene responsible for the association of the *l*p13.3 locus with LDL-C.

Supplementary key words cholesterol/metabolism • dyslipidemias • LDL/metabolism • lipoproteins/metabolism • lipoproteins/receptors • genetic association • genetic linkage • sortilin • *SORT1* • missense variant

The human *l*p13.3 locus, which includes the *SORT1* gene, has one of the strongest genetic associations with low-density lipoprotein cholesterol (LDL-C) ($P = 5 \times 10^{-324}$) (Common Metabolic Diseases Knowledge Portal (CMDKP) - <https://hugeamp.org/phenotype.html>)

phenotype=LDL). Several noncoding variants in linkage disequilibrium (LD) at this locus are associated with an ~5–8 mg/dl decrease in LDL-C and decreased incidence of coronary artery disease (1). Additionally, this locus is most strongly associated with levels of very small LDL (vsLDL) (2), which are more strongly related to atherosclerosis than large LDL (lgLDL) (3). The variants are associated with hepatic expression of three nearby genes: *CELSR2*, *PSRC1*, and *SORT1* (2, 4, 5). Mediation analysis found that regulation of hepatic *SORT1* expression is primarily responsible for the association of this locus with LDL-C, with *SORT1* levels being negatively correlated with LDL-C (2, 4). These studies suggested that the regulatory mechanism underlying the association of the noncoding variants at the *l*p13.3 locus with LDL-C is sortilin-mediated and liver-specific. However, validation studies in mouse and cell models have led to contradictory results, as we summarized in our recent review (1). This raised the possibility that a different gene regulated by the noncoding variants at this locus might be responsible for the association with LDL-C.

The vast majority of genome-wide association study (GWAS) loci are in noncoding regions of the genome. Within these loci are often several highly correlated variants, making it challenging to identify which may be causal (6–8). In addition to this complexity, there are often multiple genes located at these loci that could plausibly be responsible for the physiological phenotype. A common approach to prioritize candidate genes is to assess the correlations of their expression with variants at a locus. However, a gene with expression most strongly correlated with the variant genotype is not necessarily the responsible gene. The identification of coding variants can provide a direct means for determining causal genes and hypothesizing functional mechanisms. Coding variants often have larger effects on an individual's traits than noncoding variants, but

*For correspondence: Alan D. Attie, adattie@wisc.edu.

their rarity makes their contribution to overall heritability small, such that they are often not identified in GWASs (9–12).

Here, using exome sequencing (ES) in the Old Order Amish founder population of Lancaster County, PA, we identify coding mutations in *SORT1* that are associated with higher (rs141749679, K302E) and lower (rs149456022, Q225H) levels of LDL-C. Carriers of the sortilin K302E mutation have significantly lower levels of IgLDL particles. To study these mutations further, we introduced them into C57BL/6J mice using CRISPR/Cas9 technology. Unexpectedly, the sortilin K302E variant resulted in reduced non-high-density lipoprotein cholesterol (non-HDL-C) and the Q225H variant did not alter cholesterol in mice. We discuss several differences in lipoprotein metabolism between mice and humans that may account for these species-specific effects.

MATERIALS AND METHODS

Amish study population

We performed a community-wide survey in 5,987 Old Order Amish individuals aged 18 years and older from Lancaster County, PA, that included a basic physical examination and fasting blood draw during the period 2010–2018 by the University of Maryland School of Medicine's Amish Research Program (13, 14) (<http://www.medschool.umaryland.edu/endocrinology/Amish-Research-Program>). The mean age of the individuals used in this study was 41.7 ± 15.4 years, and 44% were male. All study participants provided informed consent. Study protocols were approved by the University of Maryland Institutional Review Board and abide by the Declaration of Helsinki principles.

Identification of *SORT1* variants

Rsl41749679 (sortilin K302E) and rs149456022 (sortilin Q225H) were identified from ES conducted by the Regeneron Genetics Center (Tarrytown, NY) as part of an ongoing collaboration. Exome capture was performed using a slightly modified version of the xGen capture reagent available from Integrated DNA Technologies with some modifications, and the captured libraries were sequenced on the Illumina HiSeq 2500 and NovaSeq 6000 platforms. Captured fragments were sequenced to achieve a minimum of 86% of the target bases covered at 20× or greater coverage variants with call rate <90%. Further technical details of the ES methods have been previously published (15). The rs12740374 (noncoding) variant was genotyped at the Regeneron Genetics Center from the Infinium Global Screening array (Illumina, Inc, San Diego, CA). Following sequencing and genotyping, all samples underwent thorough quality control to remove samples exhibiting high levels of mendelian errors, gender discordance, low coverage (for sequencing), and high genotype missingness.

Genetic association analyses

Genetic association analyses were performed using linear mixed regression models with genotype as an independent variable and assuming an additive genetic effect. We accounted for relatedness among study subjects by including

the genetic relationship matrix in the model as a random effect. Association analyses were carried out using the Mixed Model Analysis for Pedigree and Population software program (<http://edn.som.umaryland.edu/mmap/index.php>) (16). All association analyses were adjusted for age, sex, sub-study, and measurement protocol (as needed). HDL-C, LDL-C, and triglyceride (TG) were also adjusted for rs5742904 (APOB R3527Q) and rs76353203 (APOC3 R19*), two variants that are enriched in the Amish population and are highly associated with lipid levels. APOB R3527Q increases LDL-C ~75 mg/dl (13) and APOC3 R19* is associated with decreased TG and increased HDL-C (17) in the Amish. Effect sizes are reported as 1 unit change in trait per allele. To account for the fact that we were testing three variants, we considered genotype differences with p -values < 0.017 (0.05/3) to be statistically significant.

Serum lipid measurements and ion mobility in Amish subjects

Serum was harvested from overnight fasting blood and sent to Quest Diagnostics (Horsham, Pennsylvania) for measurement of TC, HDL-C, TG, and glucose. Non-HDL-C was calculated by subtracting HDL-C from TC. LDL-C was estimated using the Friedewald method (18). Prior to ion mobility (IM) measurement of plasma lipoprotein concentration, plasma was treated with 17% ethanol to remove >97% of fibrinogen, and then all lipoproteins were precipitated with 2 mg/ml dextran sulfate and 0.15 M calcium. Precipitated lipoproteins were harvested on paramagnetic particles, washed to remove free salt and proteins (e.g., IgG, albumin, and transferrin), and then resuspended in 25 mM ammonium acetate, as previously described (19). Following isolation, lipoproteins were fractionated and quantified in a single scan using gas-phase electrophoresis (IM), as previously described (20, 21). In the IM “scans” shown in Fig. 1F, G, lipoproteins were grouped into bins, each spanning ~5 Å diameter. Lipoproteins were then further pooled by summing the total number of particles within specific size ranges that approximately group them into separately defined subclasses that have minimal methodologic and biologic overlap, and as previously characterized (22), to generate the data shown in Fig. 1H–N. [supplemental Table S1](#) shows the lipoprotein subclasses, their size ranges, and nomenclature.

Generation of sortilin K300E and sortilin Q223H mice

Sortilin K300E and sortilin Q223H mice were made via CRISPR-Cas9 genome-editing by the Gene Editing and Animal Models group at the University of Wisconsin-Madison. For the sortilin K300E mice, the highly specific target sequence GAAACGGCCCCAAGACCAA was used to introduce rs141749679 (GRCh37.p13 chr 1 NC_000001.10:g.109888432T>C) into a C57BL/6J mouse obtained from the Jackson Laboratory to mutate lysine (K) 300 to glutamic acid (E) in the sortilin protein (Uniprot Q6PHU5-1). For the sortilin Q223H mice, the highly specific target sequence ACAGGTAATCAGAATTCTGA was used to introduce rs149456022 (GRCh37.p13 chr 1 NC_000001.10:g.109897022C>A) into a C57BL/6J mouse obtained from the Jackson Laboratory to mutate glutamine (Q) 223 to histidine (H) in the sortilin protein (Uniprot Q6PHU5-1). All predicted off-targets varied by at least three nucleotides, and no single predicted off-target had an activity prediction score (CFD) higher than 0.5. An in vitro transcription template was generated by overlap-extension PCR with one oligo carrying a 5' T7

adapter, the target sequence, and a portion of the common gRNA sequence and the other oligo carrying the antisense common gRNA sequence. Following column-purification, the in vitro transcript was transcribed with the MEGAscript kit (ThermoFisher), and the resultant gRNA was cleaned with the MEGAclear kit (ThermoFisher), purified with ammonium acetate, washed with 70% ethanol, and resuspended in injection buffer (10 mM Tris-HCl, 0.1 mM EDTA, pH 7.4). A mixture of gRNA (50 ng/ μ l), ssODN (50 ng/ μ l), and Cas9 protein (40 ng/ μ l) were microinjected into the pronuclei of fertilized C57BL/6J one-cell embryos and then implanted into pseudopregnant female B6D2F1 recipients. Resultant pups were genotyped at weaning by PCR of tail DNA. For the sortilin K300E mice, the targeted region was amplified with forward primer TGCA-GATTCTCTGTGTATGAT and reverse primer TGCCCAA-CACATATATCACA. The PCR product was digested with BstXI. The base pair change to create K300E also created a BstXI site. The digest was run on a 2% agarose gel. Wild-type (WT) mice yielded a 471 bp fragment, sortilin K300E homozygous mice yielded 234 bp and 237 bp fragments, and sortilin K300E heterozygous mice yielded 471 bp, 234 bp, and 237 bp fragments. The 471 bp PCR products were also gel-purified and sequenced to identify a founder mouse that contained rs141749679 (g.109888432T>C) and had no other mutations in this 471 bp sequence. rs141749679 was approximately in the center of the 471 bp sequence. For the sortilin Q223H mice, the targeted region was amplified with forward primer GAAGCCGAGGCG-GAAGAGTG and reverse primer TTCCAGCAGCAGAC ATCCGTTT. The PCR product was digested with EcoRI. The base pair change to create Q223H led to the loss of an EcoRI site. The digest was run on a 2% agarose gel. WT mice yielded 104 bp and 259 bp fragments, sortilin Q223H homozygous mice yielded a 363 bp fragment, and sortilin Q223H heterozygous mice yielded 104 bp, 259 bp, and 363 bp fragments. The 363 bp PCR products were also gel-purified and sequenced to identify a founder mouse that contained rs149456022 (g.109897022C>A) and had no other mutations in this 363 bp sequence. rs149456022 was approximately in the center of the 363 bp sequence. One founder mouse homozygous for either variant was identified and bred to a WT C57BL/6J mouse to generate heterozygotes. Male mice homozygous for either sortilin K300E or sortilin Q223H were compared to age-matched male WT mice produced from the breeding of the respective CRISPR-edited line.

Animal care and housing

Mice were bred, housed, and cared for in the AAALAC-accredited University of Wisconsin-Madison Biochemistry Department vivarium, and all mouse experiments were conducted in accordance with University of Wisconsin-Madison IACUC-approved protocols. Mice were group housed under temperature- and humidity-controlled conditions, a 12-h light/dark cycle (6 a.m. to 6 p.m.) and received ad libitum access to water and food. Starting at weaning (~3 weeks of age), mice were fed a Western-style diet high in fat and sucrose (TD.08811, Envigo Teklad Custom Diet) containing 44.6% kcal from fat, 14.7% kcal from protein, 40.7% kcal from carbohydrate, 34% sucrose by weight, and high saturated fatty acids (>60% of total fatty acids).

Plasma measurements in mice

Mice were fasted for 4-h and bled retro-orbitally to collect plasma whole blood using EDTA as an anticoagulant. Triglyceride was measured using the Triglycerides

Reagent from ThermoFisher (TR22421), and cholesterol was measured using the Total Cholesterol Reagent from ThermoFisher (TR13421). To determine cholesterol that is nonprecipitable by heparin-MnCl₂, plasma was mixed with a solution containing heparin and MnCl₂ with a final concentration of 220 U/ml heparin and 92 mM Mn²⁺, incubated at 4°C for 20 min, centrifuged at 4°C for 20 min at 1,500 g and the cholesterol concentration of the supernatant was measured using the Total Cholesterol Reagent from ThermoFisher (TR13421). Cholesterol that was precipitable by heparin-MnCl₂ was calculated by subtracting the nonprecipitable cholesterol from the total cholesterol (TC).

Statistical methods

Amish genetic association analyses were performed as described above. All other data were analyzed by repeated measures two-way ANOVA with Geisser-Greenhouse correction and Bonferroni's multiple comparisons test with individual variances computed for each comparison. Log-normally distributed data were log-transformed prior to analysis by ANOVA. ANOVAs were performed using GraphPad Prism 9. Data are represented as mean \pm SEM.

RESULTS

rs141749679 and rs149456022, resulting in K302E and Q225H in sortilin, are associated with LDL-C in an Amish population

We performed ES on ~6,000 Old Order Amish residing in Lancaster County, PA, and discovered two rare missense variants in *SORT1* that are associated with LDL-C. The first, rs141749679 (g.109888432T>C, c.A904G, p.K302E), has a minor allele frequency (MAF) of 0.003 in the Amish and in the general European population (Genome Aggregation Database (gnomAD) - https://gnomad.broadinstitute.org/variant/1-109888432-T-C?dataset=gnomad_r2_1). We identified 35 heterozygote (TC) carriers of the K302E variant and no homozygotes. This variant was associated with a substantial 28 mg/dl (13%) increase [95% CI: 12–44 mg/dl (6–21%) increase] in TC ($P = 0.0006$), a 22 mg/dl (17%) increase [95% CI: 8–36 mg/dl (6–27%) increase] in LDL-C ($P = 0.002$), and a 25 mg/dl (17%) increase [95% CI: 10–40 mg/dl (7–27%) increase] in non-HDL-C ($P = 0.001$) and was marginally associated with increased TG ($P = 0.03$) (Table 1). The second missense variant, rs149456022 (g.109897022C>A, c.G675T, p.Q225H), is enriched in the Amish (MAF = 0.026) compared to the general European population (MAF = 0.0002) (gnomAD - https://gnomad.broadinstitute.org/variant/1-109897022-C-A?dataset=gnomad_r2_1). We identified 304 heterozygote (CA) carriers and two homozygotes. This variant was associated with a modest 8 mg/dl (4%) decrease [95% CI: 3–13 mg/dl (1–6%) decrease] in TC ($P = 0.002$), an 8 mg/dl (6%) decrease [95% CI: 4–12 mg/dl (3–9%) decrease] in LDL-C ($P = 0.0004$), and an 8 mg/dl (5%) decrease [95% CI: 3–13 mg/dl (2–9%) decrease] in non-HDL-C ($P = 0.001$) (Table 1). Neither of the two coding variants were associated with HDL-C.

TABLE 1. Association of *SORT1* variants rs141749679 (K302E), rs149456022 (Q225H), and rs12740374 with lipid traits in an Amish population

Variant	Trait	Noncarriers	Het Carriers	Hom Carriers	Effect (95% CI)	P
		Mean ± SD or Median (25%, 75%) (mg/dl)				
rs141749679 (K302E)		TT (N = 5,904)	TC (N = 34)	CC (N = 0)		
	TC	208 ± 50	253 ± 57		28 (12–44)	0.0006
	LDL-C	132 ± 46	168 ± 50		22 (8–36)	0.002
	non-HDL-C	147 ± 49	186 ± 50		25 (10–40)	0.001
	HDL-C	61 ± 17	68 ± 19		3 (–3 to 9)	0.3
	lnTG	58 (43, 84)	73 (54, 109)		0.2 (0.0–0.4)	0.03
rs149456022 (Q225H)		CC (N = 5,633)	CA (N = 303)	AA (N = 2)		
	TC	209 ± 50	199 ± 44	144 ± 18	–8 (–3 to –13)	0.002
	LDL-C	133 ± 46	121 ± 42	72 ± 28	–8 (–4 to –12)	0.0004
	non-HDL-C	148 ± 49	135 ± 45	85 ± 38	–8 (–3 to –13)	0.001
	HDL-C	61 ± 17	64 ± 17	59 ± 20	0.3 (–1.5 to 2.2)	0.7
	lnTG	58 (43, 85)	54 (42, 79)	62 (30, 94)	–0.02 (–0.08 to 0.03)	0.4
rs12740374 (noncoding)		GG (N = 2,619)	GT (N = 2,649)	TT (N = 685)		
	TC	213 ± 53	207 ± 48	198 ± 43	–5 (–3 to –7)	1 × 10^{–10}
	LDL-C	137 ± 49	131 ± 44	122 ± 40	–5 (–4 to –6)	1 × 10^{–10}
	non-HDL-C	148 ± 49	135 ± 45	85 ± 38	–5 (–3 to –7)	1 × 10^{–9}
	HDL-C	61 ± 17	62 ± 17	62 ± 7	–0.06 (–0.7 to 0.6)	0.9
	lnTG	58 (44, 84)	58 (43, 85)	57 (42, 82)	–0.01 (–0.03 to 0.01)	0.3

Abbreviations: Het, heterozygote; Hom, homozygote; LDL-C, low-density lipoprotein cholesterol; non-HDL-C, non-high-density lipoprotein cholesterol; TC, total cholesterol; TG, triglyceride.

Associations were tested using linear mixed regression models. Effect sizes and *P* values are per-allele and adjusted for age, sex, and two coding variants, one in APOB and one in APOC3, previously shown to greatly affect lipid levels in this population (see [Methods](#) for details). Effect sizes are in units of mg/dl and include the 95% confidence interval (CI). Genotype-specific data are represented as mean ± SD for normally distributed traits (TC, LDL-C, non-HDL-C, HDL-C, and lnTG) and median (25%, 75%) for nonnormally distributed traits (TG) and are not adjusted for covariates. All traits are in units of mg/dl. TG was natural log (ln) transformed prior to analysis. *P*-values <0.017 (0.05/3, to account for testing three variants) are statistically significant, indicated in bold.

We repeated the association analyses of rs141749679 and rs149456022 with lipids but removed subjects with the rs149456022 variant minor allele (A) from the analysis of rs141749679 and removed subjects with the rs141749679 variant minor allele (C) from the analysis of rs149456022. The results were highly consistent with those shown in [Table 1](#) (i.e., 24 mg/dl higher LDL-C among those with the rs141749679 C allele (*P* = 0.002) and 7 mg/dl lower LDL-C among those with the rs149456022 A allele (*P* = 0.002)).

We compared the effect size on LDL-C of these *SORT1* coding variants with that of the noncoding variant rs12740374 (g.109817590G>T), the putative causal variant for the association of the human *lpl3.3* locus with LDL-C in GWAS (CMDKP - <https://hugeamp.org/variant.html?variant=1%3A109817590%3AG%3AT>) (2, 4). rs12740374 is a common variant in the general European population (MAF = 0.2) (gnomAD - https://gnomad.broadinstitute.org/variant/1-109817590-G-T?dataset=gnomad_r2_1) but is even more prevalent in the Amish (MAF = 0.3), in whom it was associated with a 5 mg/dl (2%) decrease [95% CI: 3–7 mg/dl (1–6%) decrease] in TC (*P* = 1 × 10^{–10}), a 5 mg/dl (4%) decrease [95% CI: 4–6 mg/dl (3–9%) decrease] in LDL-C (*P* = 1 × 10^{–10}), and a 5 mg/dl (3%) decrease [95% CI: 3–7 mg/dl (2–9%) decrease] in non-HDL-C (*P* = 1 × 10^{–9}) ([Table 1](#)). This is in line with the ~5–8 mg/dl decrease previously reported for rs12740374 and other variants in LD at the *lpl3.3* locus (4, 23–29). We

performed LD analysis for rs141749679 (K302E), rs149456022 (Q225H), and rs12740374 and found that rs149456022 (Q225H) is in LD with the common noncoding variant rs12740374, while rs141749679 (K302E) is not ([supplemental Table S1](#)). When a multivariate model was used to test for the association of each of the three variants with LDL-C, where all three variants were analyzed together, the effect size and strength of the association of rs141749679 (K302E) with LDL-C remained essentially unchanged, while those of rs149456022 (Q225H) decreased ([supplemental Table S2](#)), reflecting the results of the LD analysis.

We looked for any association between the rs141749679 (K302E) or rs149456022 (Q225H) variants and lipids in the most recent lipid GWAS, from the Global Lipids Genetics Consortium (30). Phenome-wide association analysis of the aggregated GWAS results across all ancestries revealed that the rs141749679 (K302E) variant is associated with increased LDL-C (beta = 0.04, *P* = 0.002) and that the rs149456022 (Q225H) variant is associated with decreased LDL-C (beta = –0.10, *P* = 0.02) ([supplemental Table S3](#)), replicating the association and directionality of the association we observe in the Amish population for these variants. When the data are stratified by ancestry, the K302E variant is significantly associated with LDL-C in European ancestry individuals. K302E carriers of Hispanic or African ancestry have increased LDL-C, and

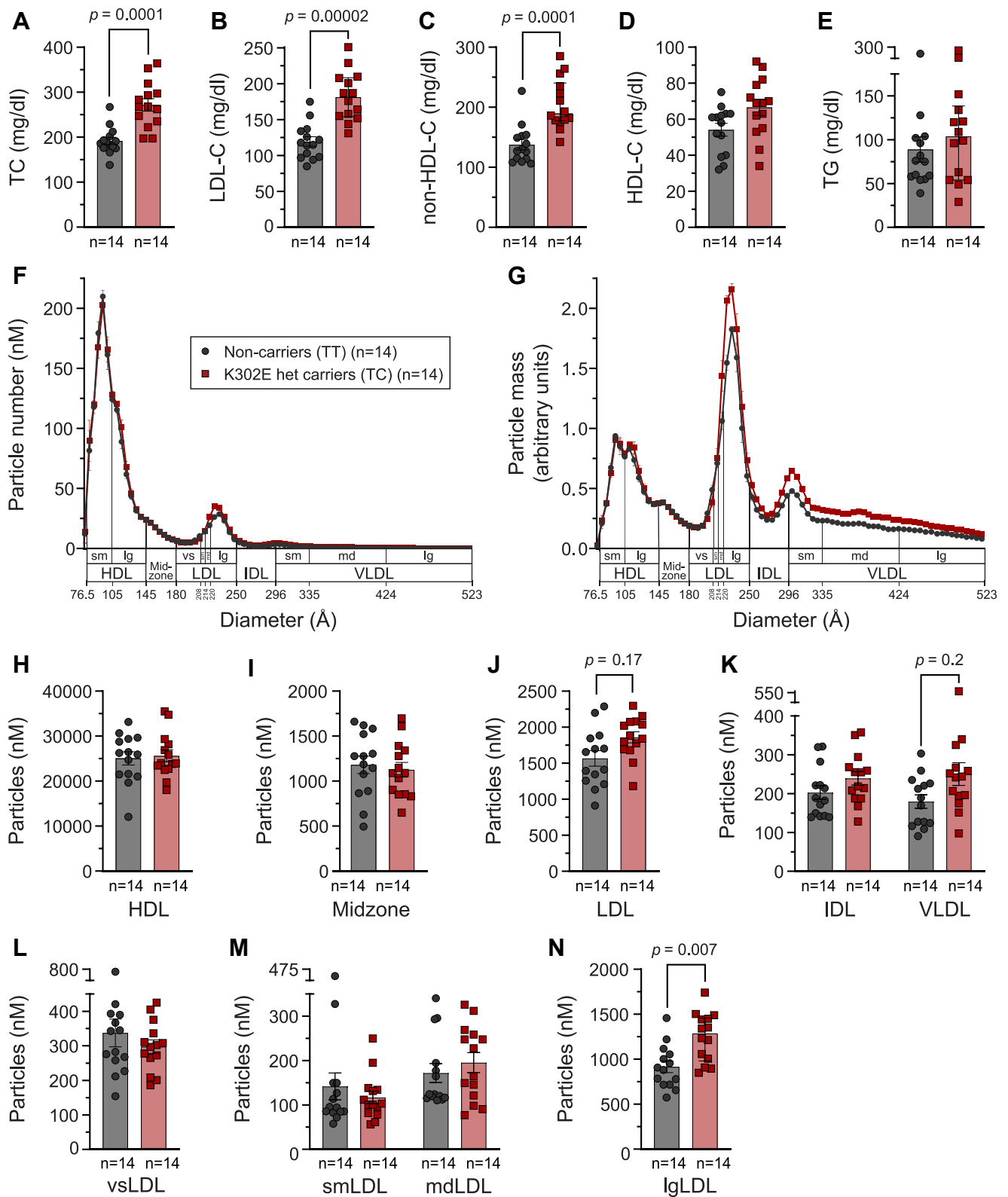


Fig. 1. Heterozygote carriers of rs141749679 (sortilin K302E) in the Amish have significantly increased levels of large LDL particles. A: Total cholesterol (TC), (B) HDL cholesterol (HDL-C), (C) non-HDL-cholesterol (non-HDL-C), (D) LDL cholesterol (LDL-C), (E) triglycerides (TG), (F) lipoprotein particle concentration and (G) mass, as analyzed by ion mobility, and concentration of particles in (H) HDL, (I) midzone, (J) LDL, and (K) IDL and VLDL classes and (L) very small LDL (vsLDL), (M) small LDL (smLDL) and medium LDL (mdLDL) and (N) large LDL (lgLDL) subclasses (as defined in [supplemental Table S3](#)) in 14 rs141749679 (K302E) heterozygote carriers and 14 age- and sex-matched noncarriers. Serum was collected after an overnight fast. Data in panels A–E, H–K, and L–N were analyzed by three separate repeated measures two-way ANOVAs with Geisser-Greenhouse correction followed by Bonferroni’s multiple comparisons test with individual variances computed for each comparison. All data were log transformed prior to analysis. Data are represented as mean ± SEM.

Q223H carriers of European or Hispanic ancestry have decreased LDL-C, but these effects did not reach statistical significance (supplemental Table S3).

Carriers of the sortilin K302E variant have increased levels of IgLDL particles

To investigate if the K302E variant affects specific subclasses of LDL, we performed IM, a method that directly measures the diameter and concentration of lipoprotein particles (20, 21), in a subset of 14 K302E heterozygote carriers and 14 age- and sex-matched noncarriers. Importantly, the K302E heterozygote carriers in this subset of subjects showed increased TC, LDL-C, and non-HDL-C, and no difference in HDL-C or TG (Fig. 1A–E), in concordance with the association analyses in the full set of subjects. Figure 1F, G show the lipoprotein particle concentration and particle mass, respectively, plotted against particle diameter. Lipoproteins were then pooled by summing the total number of particles within diameter ranges that group the lipoproteins into major classes: HDL, midzone (particles with diameters between those of HDL and LDL), LDL, IDL, and VLDL (Fig. 1H–K), as defined in supplemental Table S4 and as previously described (22). Heterozygote carriers of the K302E variant had a trend for increased LDL and VLDL particles ($P = 0.17$ and 0.2 , respectively) (Fig. 1J, K). Division of LDL particles into subclasses that have minimal methodologic and biologic overlap (vsLDL, small LDL, medium LDL, and IgLDL) (Fig. 1L–N) revealed a significant increase in the number of IgLDL particles in K302E carriers compared to noncarriers ($P = 0.007$) (Fig. 1N), which correlated with LDL-C in the 28 subjects (supplemental Fig. S1).

K302 and Q225 residues are highly conserved and located in the ligand-binding domain of sortilin

K302 and Q225 are highly conserved residues that lie within the 10-bladed β -propeller ligand-binding domain of sortilin (Fig. 2A). K302 lies along the outside edge of the β -propeller, near the transmembrane domain, and is oriented such that it would point toward membrane when sortilin is membrane-anchored inside the cell (Fig. 2B). The side chain nearest to that of K302 is that of D320 and is located 4.6 Å away (Fig. 2B). Q225 is located along the inside edge of the β -propeller, near the narrower “back” opening of the tunnel that passes through the middle of the β -propeller (Fig. 2B). The side chain nearest to that of Q225 is that of N265 and is located 5.0 Å away (Fig. 2B).

Mice harboring K300E or Q223H do not recapitulate lipid effects seen in the Amish despite a high structural homology of sortilin between the two species

Using CRISPR/Cas9, we generated C57BL/6J mouse models expressing either the sortilin K302E variant (K300E in mice) or the Q225H variant (Q223H in mice). WT and CRISPR-edited mice harboring each mutation

were fed a high-fat, high-sucrose diet (45% kcal from fat, 34% sucrose by weight) for 15 weeks. We performed heparin-MnCl₂ precipitation on fasting plasma to determine precipitable and nonprecipitable cholesterol, estimates of non-HDL-C and HDL-C in mice, respectively, a method like that used in the Amish. Surprisingly, sortilin K300E mice had drastically decreased TC ($P = 0.0002$) and decreased precipitable cholesterol ($P = 0.09$) (Fig. 3A, B), as well as decreased nonprecipitable cholesterol ($P = 0.0009$) (Fig. 3C). This is opposite to our prediction based on the genetic association in the Amish. There was no difference in plasma TG between WT and K300E mice (Fig. 3D). Sortilin Q223H mice had TC, precipitable cholesterol, nonprecipitable cholesterol, and TG levels comparable to that of WT controls (Fig. 3E–H).

The opposing effects of the mutations on cholesterol in mice versus humans is despite a high level of structural homology between mouse and human sortilin. The linear sequence of the full-length mouse and human sortilin proteins are highly conserved, with 91% identity and 95% similarity (Fig. 4A). Further, the structure of the soluble domain of mouse sortilin is strikingly similar to that of human sortilin (Fig. 4B). Alignment of the two structures using the Vector Alignment Search Tool results in a very low root mean square deviation of 0.85 Å, indicating high structural similarity (Fig. 4C). Each of the 588 amino acids modeled in both the mouse and human structures were determined by Vector Alignment Search Tool to be aligned in 3D space.

DISCUSSION

By conducting ES in the Old Order Amish population of Lancaster County, PA, we identified two coding mutations in *SORT1* associated with LDL-C. To our knowledge, this is the first report that naturally occurring coding mutations in *SORT1* are associated with a cardiovascular disease-related trait. The rs141749679 variant (sortilin K302E) was recently reported to be significantly associated with Alzheimer’s disease (39).

The effect size of the common noncoding regulatory variant rs12740374 on LDL-C is modest compared to that of the rare *SORT1* coding variant rs141749679 (K302E). For several years, the prevailing hypothesis behind the genetic risk of common diseases, such as cardiovascular disease, has been that disease risk alleles with high frequencies cumulatively cause common diseases—the common variant, common disease hypothesis (40). However, more recent work has demonstrated that low-frequency genetic variants with large phenotypic effects can also contribute significantly to complex diseases—the rare variant, common disease hypothesis (12).

In individual GWASs, noncoding variants at the 1p13.3 locus are primarily associated with decreased LDL-C and are not associated with TG (1), similar to

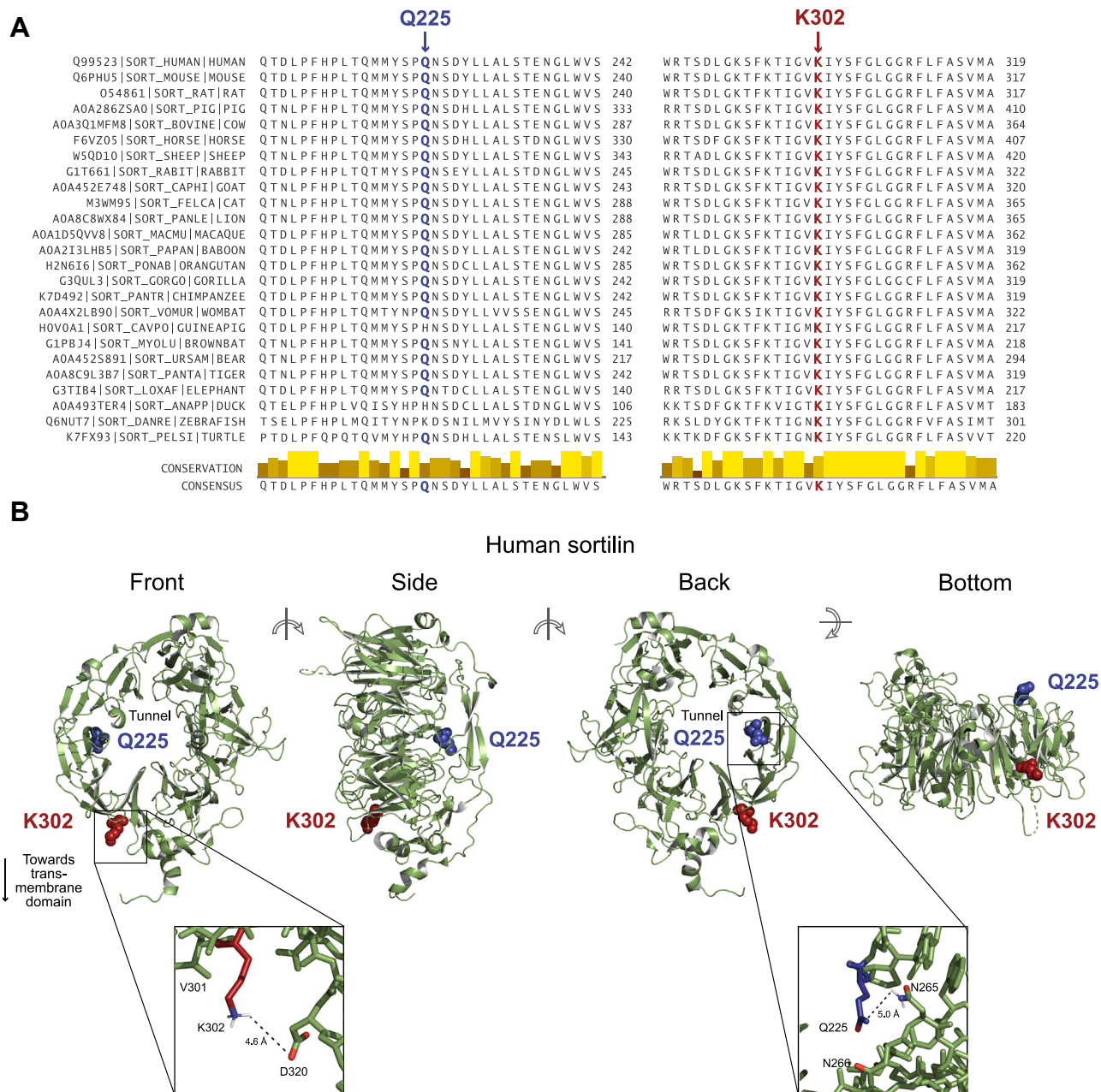


Fig. 2. K302 and Q225 are highly conserved residues and are located in the β -propeller domain of sortilin. A: Multiple species alignment of the regions of sortilin containing lysine 302 (K302) and glutamine 225 (Q225) using Clustal Omega (31, 32) and visualized using Jalview (33). B: Structure of the luminal domain of human sortilin as determined by Quistgaard *et al.* (34) (Protein Data Bank 3F6K) and visualized using PyMol (35). K302 and Q225 are highlighted in red and blue, respectively.

what we see in the present study with one of these variants (rs12740374) (Table 1). However, meta-analysis of GWAS datasets reveals a significant association of this locus with decreased TG ($P = 1 \times 10^{-14}$) (CMDKP - <https://hugeamp.org/region.html?chr=1&end=109990540&phenotype=LDL&start=109802190>), an effect with the same directionality as that of LDL-C. Previous studies have suggested that the manipulation of *SORT1* expression can affect LDL levels in both VLDL (TG)-dependent and -independent manners (1). Even though the K302E variant was only marginally

associated with TG in this study, its effect on LDL-C and TG are also in the same (positive) direction. The opposite directionality of the effect on LDL-C and TG of the K302E variant compared to that of the rs12740374 variant, which is associated with increased hepatic *SORT1* expression, suggests that K302E is a reduced-function mutation. Further studies are required to determine whether the effect of the K302E variant on LDL-C is through an effect on VLDL.

The use of IM allowed us to directly quantify the concentration of lipoprotein particles in different

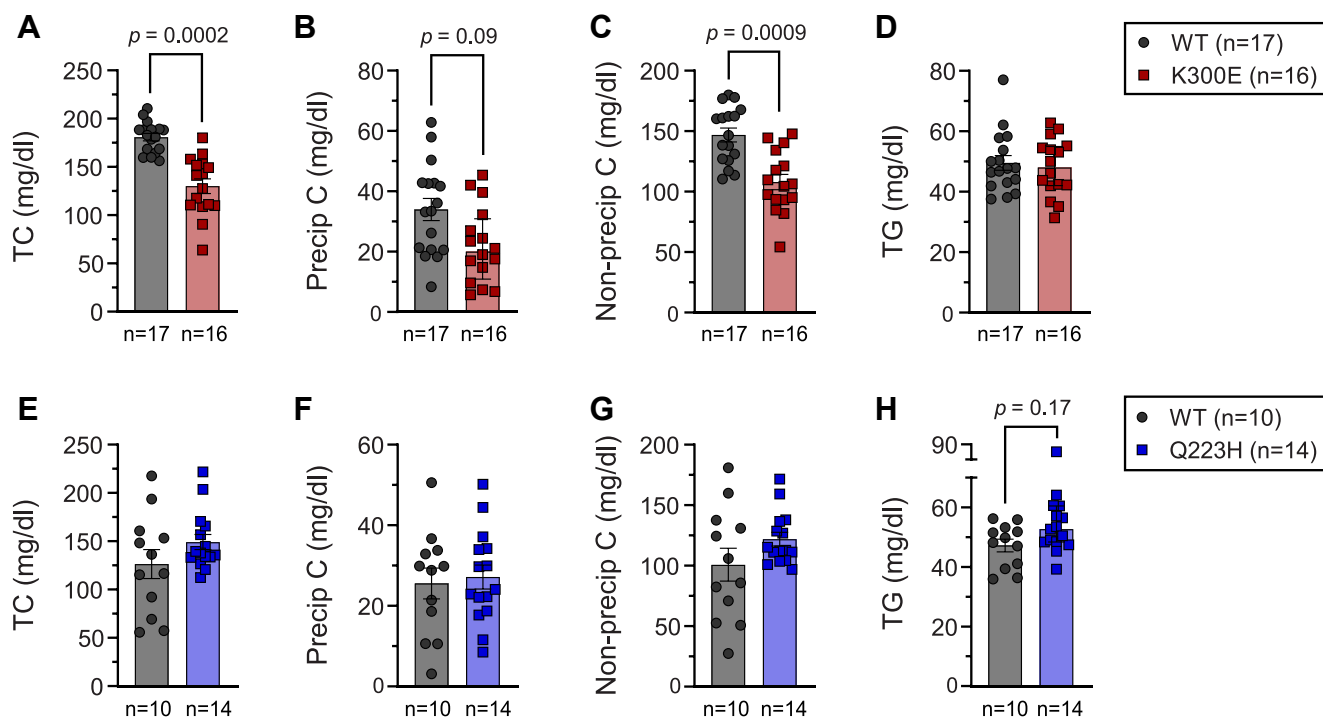


Fig. 3. Mice expressing the sortilin K300E variant have decreased TC and heparin-MnCl₂-precipitable and non-precipitable cholesterol. (A) Total cholesterol (TC), (B) heparin-MnCl₂-precipitable cholesterol (Precip C), (C) non-heparin-MnCl₂-precipitable cholesterol (Non-precip C), and (D) triglycerides (TG) in sortilin K300E mice (n = 17 for WT and n = 16 for K300E) and (E) TC, (F) Precip C, (G) Non-precip C, (H) TG in sortilin Q223H mice (n = 10 for WT and n = 14 for Q223H). All mice were male, on a C57BL/6J background, 18 weeks of age, and fed a high-fat, high-sucrose (HF/HS) diet for 15 weeks. Plasma was collected after a 4-h fast. Data in panels A–D and E and F were analyzed by two separate repeated measures two-way ANOVAs with Geisser-Greenhouse correction followed by Bonferroni's multiple comparisons test with individual variances computed for each comparison. All data were log transformed prior to analysis. Data are represented as mean ± SEM.

subclasses in Amish individuals. We found that heterozygote carriers of rs141749679 (K302E) had an increased concentration of IgLDL particles and that the effect of the variant was specific to this subclass of LDL. Musunuru *et al.* previously used IM to measure subclasses of lipoprotein particles in carriers of the noncoding variant rs646776 at the *lpl3.3* locus, which is in LD with the putative causal variant rs12740374. They reported that homozygote carriers of the rs646776 minor allele had decreased concentrations of all LDL subclasses compared to noncarriers, with the greatest decrease being in vsLDL (20% decrease, $P = 1.1 \times 10^{-11}$) with progressively smaller decreases for larger LDL subclasses (2). The *lpl3.3* noncoding variants are associated with a wider range of LDL particle sizes than the K302E variant, suggesting that the effect of the K302E mutation may be more complex than just a simple reduced function. If sortilin has different binding affinities for different sizes of lipoprotein particles, as has been demonstrated for the LDL receptor (LDLR) (41, 42), a change in these binding affinities induced by the K302E mutation may explain the differential effect of the K302E mutation and the noncoding variant on different sizes of LDL particles.

A major finding of our study is that the sortilin K302E and Q223H mutations have differing effects on cholesterol in mice versus humans. The mouse and

human sortilin proteins are highly homologous, both in linear amino acid sequence and structurally. It is therefore highly unlikely that the differences we observe upon introducing these *SORT1* variants into mice are due to the mutations having different effects on the sortilin protein in the two species. Rather, there are important differences in lipoprotein metabolism between humans and mice that may provide an explanation for this. One difference is that mice lack cholesterol ester transfer protein, which in humans mediates transfer of cholesterol from HDL to apoB-containing lipoproteins (43, 44). A second is that the apolipoprotein B mRNA editing enzyme (*APOBEC1*), which is responsible for post-transcriptional production of apolipoprotein B-48 (apoB48) from the apolipoprotein B-100 (*APOB100*) transcript, is expressed in liver and intestine in mice but is only expressed in intestine in humans (45). Thus, unlike humans, mice produce a mixture of liver-derived apoB48- and apoB100-containing particles. While both the LDLR and sortilin bind to apoB100 (46, 47), neither binds to apoB48 (46, 48). Finally, in contrast to humans, mice carry significant amounts of apolipoprotein E (apoE) on LDL particles (49), with much of the hepatic clearance of apoB48-containing LDL occurring through interaction of apoE with LDLR-related protein 1 (50). Sortilin has

A

Q99523 SORT_HUMAN	MERPWGAADGLSRWPHGLGLLLLLLQLLPPSTLSQDRLDAPPPAAPLPRWSGPIGVSWGLRAAAAGGAFPRGGRWRRSAPGEDEEC	86
Q6PHU5 SORT_MOUSE	MERPRGAADGLLRWPL--LGLLLLLLQLLPPAAVGDQDRLDAPPPAPELLRWAGPVGVSWGLRAAAPGGPVPRAGRWRGAPAEQDC	84
CONSERVATION		
Q99523 SORT_HUMAN	GRVRDFVAKLANNTHQHVFDLGRSVLSWVGDSTGIVLVLTFFHVPLVIMTFGQSKLYRSEYDGKFNKIDITDLINNTFIRTEFGM	172
Q6PHU5 SORT_MOUSE	GRLPDFIAKLNTNTHQHVFDLGSVLSWVGDSTGIVLVLTFFQVPLVIVSFGQSKLYRSEYDGKFNKIDITLNLINNTFIRTEFGM	170
CONSERVATION		
Q99523 SORT_HUMAN	A I G P E N S G K V L T A E V S G G R G R I F R S S D F A K N F V Q T D L P F H P L T Q M M Y S P Q N S D Y L L A L S T E N G L W V S K N F G G K W E E I H K A V C L	258
Q6PHU5 SORT_MOUSE	A I G P E N S G K V I L T A E V S G G R G R V F R S S D F A K N F V Q T D L P F H P L T Q M M Y S P Q N S D Y L L A L S T E N G L W V S K N F G E K W E E I H K A V C L	256
CONSERVATION		
Q99523 SORT_HUMAN	A K W G S D N T I F F T T Y A N G S C K A D L G A L E L W R T S D L G K S F K T I G V K I Y S F G L G G R F L F A S V M A D K D T T R R I H V S T D Q G D T W S M A Q L P S	344
Q6PHU5 SORT_MOUSE	A K W G P N N I I F F T T H V N G S C K A D L G A L E L W R T S D L G K T F K T I G V K I Y S F G L G G R F L F A S V M A D K D T T R R I H V S T D Q G D T W S M A Q L P S	342
CONSERVATION		
Q99523 SORT_HUMAN	V G Q E Q F Y S I L A A N D M V F M H V D E P G D T G F G T I F T S D D R G I V Y S K S L D R H L Y T T T G G E T D F T N V T S L R G V Y I T S V L S E D N S I Q T M I T	430
Q6PHU5 SORT_MOUSE	V G Q E Q F Y S I L A A N E D M V F M H V D E P G D T G F G T I F T S D D R G I V Y S K S L D R H L Y T T T G G E T D F T N V T S L R G V Y I T S L S E D N S I Q S M I T	428
CONSERVATION		
Q99523 SORT_HUMAN	F D Q G G R W T H L R K P E N S E C D A T A K N K N E C S L H I H A S Y S I S Q K L N V P M A P L S E P N A V G I V I A H G S V G D A I S V M V P D V Y I S D D G G Y S W T	516
Q6PHU5 SORT_MOUSE	F D Q G G R W E H L R K P E N S K C D A T A K N K N E C S L H I H A S Y S I S Q K L N V P M A P L S E P N A V G I V I A H G S V G D A I S V M V P D V Y I S D D G G Y S W A	514
CONSERVATION		
Q99523 SORT_HUMAN	K M L E G P H Y Y I L D S G G I I V A I E H S R P I N V I K F S T D E G C Q W Q T Y T F T R D P I Y F T G L A S E P G A R S M N I S I W G F T E S F L T S Q W V S Y T I	602
Q6PHU5 SORT_MOUSE	K M L E G P H Y Y I L D S G G I I V A I E H S N R P I N V I K F S T D E G C Q W S Y V F T Q E P I Y F T G L A S E P G A R S M N I S I W G F T E S F I T R Q W S Y T V	600
CONSERVATION		
Q99523 SORT_HUMAN	D F K D I L E R N C E E K D Y T I W L A H S T D P E D Y E D G C I L G Y K E Q F L R L R K S S V C Q N G R D Y V V T K Q P S I C L C S L E D F L C D F G Y Y R P E N D S K C	688
Q6PHU5 SORT_MOUSE	D F K D I L E R N C E E D D Y T T W L A H S T D P G D Y K D G C I L G Y K E Q F L R L R K S S V C Q N G R D Y V V A K Q P S V C P C S L E D F L C D F G Y F R P E N A S E C	686
CONSERVATION		
Q99523 SORT_HUMAN	V E Q P E L K G H D L E F C L Y G R E E H L T T N G Y R K I P G D K C Q G G V N P V R E V K D L K K K C T S N F L S P E K Q N S K S N S V P I I L A I V G L M L V T V V A G	774
Q6PHU5 SORT_MOUSE	V E Q P E L K G H E L F C L Y G K E E H L T T N G Y R K I P G D K C Q G G M N P A R E V K D L K K K C T S N F L N P T K Q N S K S N S V P I I L A I V G L M L V T V V A G	772
CONSERVATION		
Q99523 SORT_HUMAN	V L I V K Y V C G G R F L V H R Y S V L Q Q H A E A N G V D G V D A L D T A S H T N K S G Y H D D S D E D L L E	831
Q6PHU5 SORT_MOUSE	V L I V K Y V C G G R F L V H R Y S V L Q Q H A E A --D G V E A L D S T S H A -K S G Y H D D S D E D L L E	825
CONSERVATION		

Identity: 91%
Similarity: 95%

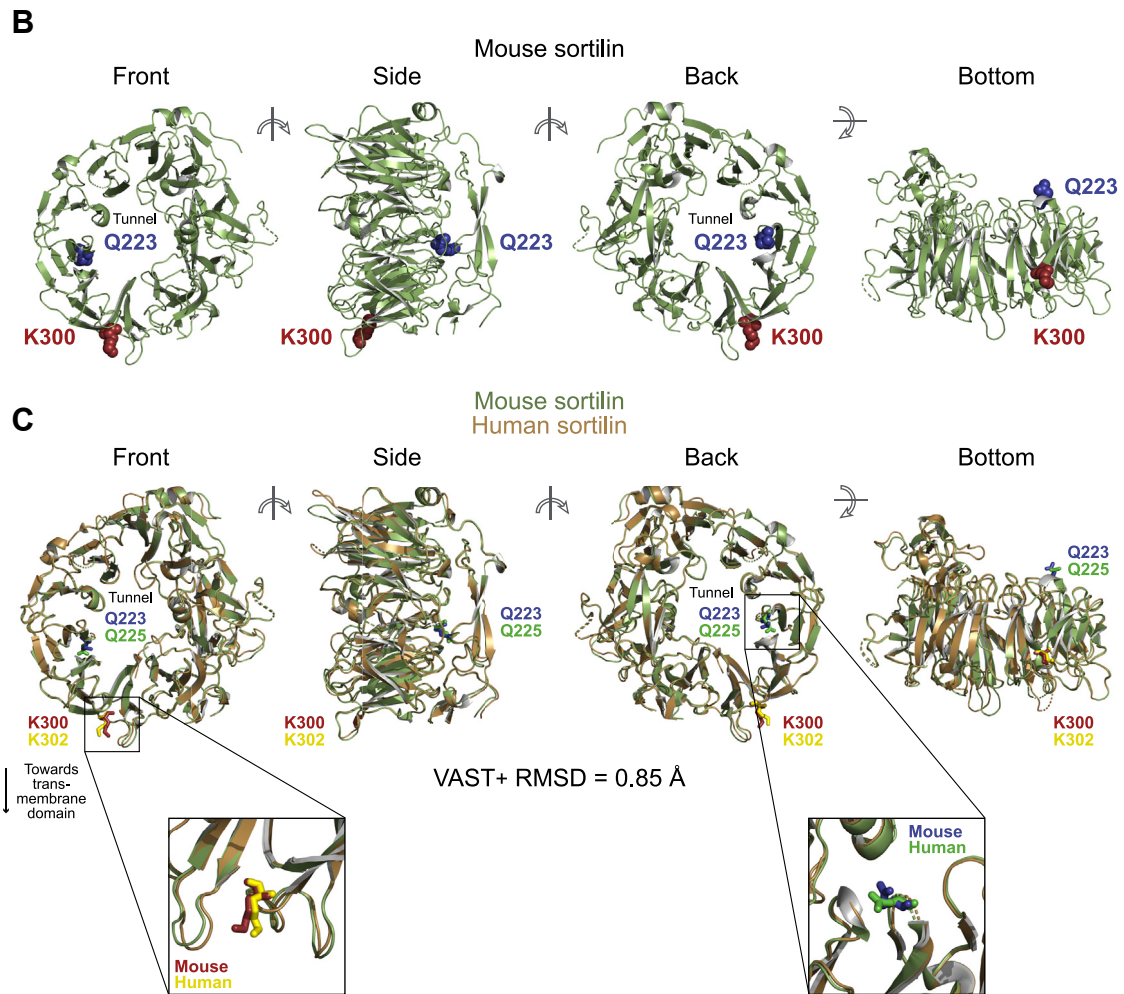


Fig. 4. Mouse and human sortilin have high structural homology. **A:** Alignment of the amino acid sequences of full-length mouse and human sortilin using EMBOSS Needle (36) and visualized using Jalview (33). **B:** Structure of the luminal domain of mouse sortilin as determined by Loupel *et al.* (37) (Protein Data Bank 5NMR) and visualized using PyMol (35). K300 and Q223 are highlighted in red and blue, respectively. **C:** Alignment of the structures of the luminal domains of mouse and human sortilin (Protein Data Bank ID 5NMR and 3F6K, respectively) using the Vector Alignment Search Tool VAST+ (38) and visualized using PyMol (35).

been shown to bind apoE in the brain (51), making it plausible that it also acts as an apoE receptor in the liver. Given these differences, the fate of LDL in mice is likely to be affected by more proteins than in humans. This, combined with the possibility that K300E or Q223H affects the binding of apoE to sortilin, may contribute to the differential effects of these mutations in mice versus humans.

Both K302 and Q225 are highly conserved residues in sortilin and are located in its ligand-binding domain. Studies that have analyzed the structure of sortilin and/or carried out competitive binding experiments with several of its known ligands have indicated the presence of at least two distinct binding sites within the tunnel of its β -propeller (34, 37, 52, 53). There is evidence for allosteric regulation between the binding sites and for some ligands to span both binding sites (52, 53). Therefore, mutations within in the β -propeller, such as K302E and Q225H, may affect the binding of some ligands to sortilin but not others, allowing for partial loss- or gain-of-function.


The binding of ligands within the tunnel of sortilin's β -propeller is regulated through dimerization of sortilin at low pH (37). Dimerization occurs along the front of the β -propeller and causes a conformational change that results in collapse of the tunnel (37). Determination of sortilin's dimerized form predicts a 2-fold axis with the dimer oriented perpendicular to cellular membranes. Based on this proposed orientation, 10 lysine residues reside at the dimer/membrane interface, possibly functioning to stabilize the dimer form by interacting with negatively charged glycolipids in the membrane. K302 is one of the 10 lysine residues at this interface (37). Replacing a positively charged lysine (K) residue with a negatively charged glutamic acid (E) residue may disrupt dimer stability and alter ligand binding, further suggesting a reduced-function effect of this mutation.

A limitation of our study is that it does not address the tissue site of action of sortilin. Our analysis favors the liver as the site of action affecting LDL. Studies following up on the original discovery that the human *Ipl3.3* locus is associated with LDL-C have found a significant association of the locus with hepatic *SORT1* expression levels, but not with *SORT1* expression in other tissues and cells involved in lipoprotein metabolism, namely white adipose tissue (2), blood vessels (54, 55), monocytes (56), and whole blood (57). However, we remain puzzled by the fact that *SORT1* is expressed at a very low level in the liver and at a far higher level in adipose tissue. Adipose tissue is a major site for VLDL lipolysis by lipoprotein lipase, and one of the earliest ligands identified for sortilin was lipoprotein lipase (58). Thus, it is possible that the functions of sortilin and/or sortilin K302E in adipose tissue may contribute to their effects on lipoprotein metabolism.

In conclusion, this study shows a direct causal association between mutations in the sortilin protein and

LDL levels. It provides further granularity to the early studies of Musunuru *et al.* and others, suggesting that the common noncoding locus at human *Ipl3.3* exerts its effect on LDL through its regulation of *SORT1* expression (2, 46, 59). The identification of coding variants in *SORT1* that are associated with LDL-C in humans presented here provides renewed confidence that *SORT1* is the gene responsible for the strong association between the *Ipl3.3* locus and LDL-C in human GWAS.

Data availability

All data described in the manuscript are contained within the manuscript. 

Supplemental data

This article contains [supplemental data](#) (20–22, 30).

Acknowledgments

The authors would like to acknowledge the Regeneron Genetics Center (RGC) (Tarrytown, NY) for conducting the exome sequencing in this study as part of an ongoing collaboration. They would like to thank Dr Ira Goldberg (New York University Langone School of Medicine) for providing his expertise and feedback on the manuscript.

Author contributions

K. A. M., S. M. K., J. O., K. A. R., and B. D. M. validation; K. A. M., S. M. K., J. O., K. A. R., and B. D. M. formal analysis; K. A. M., K. L. S., S. M. K., J. O., K. A. R., and B. D. M. investigation; K. A. M. and A. D. A. writing original draft; K. A. M., M. P. K., R. M. K., B. D. M., A. R. S., and A. D. A. writing review and editing; K. A. M. visualization; K. A. M., R. M. K., and B. D. M. project administration; K. L. S. and R. M. K. methodology; M. P. K., R. M. K., A. R. S., and A. D. A. supervision; R. M. K., B. D. M., A. R. S., and A. D. A. resources; R. M. K., B. D. M., A. R. S., and A. D. A. funding acquisition; B. D. M. data curation; A. R. S. and A. D. A. conceptualization.

Author ORCIDs

Kelly A. Mitok  <https://orcid.org/0000-0002-0167-3990>
Sarah M. King  <https://orcid.org/0000-0002-3907-8804>
Kathleen A. Ryan  <https://orcid.org/0000-0003-1158-8455>
Braxton D. Mitchell  <https://orcid.org/0000-0003-4920-4744>
Alan R. Shuldiner  <https://orcid.org/0000-0001-9921-4305>
Alan D. Attie  <https://orcid.org/0000-0002-0568-2261>

Funding and additional information

This work was supported by the National Institutes of Health grants R01 DK101573 and R01 DK102948 (A. D. A.), the National Institutes of Health grants U01 HL072515, U01 GM074518, R01 AG18728, P30 DK072488, and P30 AG028747 (B. D. M. and A. R. S.), the American Heart Association Predoctoral Fellowship Award 15PRE21280006 (K. A. M.), by Dolores Jordan Endowed Chair (R. M. K.), UCSF Benioff Children's Hospital, Oakland, and by the Regeneron Genetics Center. The content is solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health.

Conflict of interest

The authors declare that they have no conflicts of interest with the contents of this article.

Abbreviations

apoB100, apolipoprotein B-100; apoB48, apolipoprotein B-48; *APOBEC1*, apolipoprotein B mRNA editing enzyme catalytic subunit 1; APOC3, apolipoprotein C3; apoE, apolipoprotein E; *CELSR2*, cadherin EGF LAG seven-pass G-type receptor 2; CMDKP, Common Metabolic Diseases Knowledge Portal; CRISPR/Cas9, clustered regularly interspaced short palindromic repeats/CRISPR-associated protein 9; ES, exome sequencing; gnomAD, Genome Aggregation Database; GWAS, genome-wide association study; Het, heterozygote; Hom, homozygote; IM, ion mobility; LDLR, low-density lipoprotein receptor; lgLDL, large LDL; MAF, minor allele frequency; *SORT1*, sortilin; TC, total cholesterol; TG, triglyceride; vsLDL, very small LDL.

Manuscript received August 9, 2023, and in revised form October 23, 2023. Published, JLR Papers in Press, October 31, 2023, <https://doi.org/10.1016/j.jlr.2023.100468>

REFERENCES

1. Mitok, K. A., Keller, M. P., and Attie, A. D. (2022) Sorting through the extensive and confusing roles of sortilin in metabolic disease. *J. Lipid Res.* **63**, 100243
2. Musunuru, K., Strong, A., Frank-Kamenetsky, M., Lee, N. E., Ahfeldt, T., Sachs, K. V., et al. (2010) From noncoding variant to phenotype via *SORT1* at the 1p13 cholesterol locus. *Nature* **466**, 714–719
3. Liou, L., and Kaptoge, S. (2020) Association of small, dense LDL-cholesterol concentration and lipoprotein particle characteristics with coronary heart disease: a systematic review and meta-analysis. *PLoS One* **15**, e0241993
4. Kathiresan, S., Melander, O., Guiducci, C., Surti, A., Burtt, N. P., Rieder, M. J., et al. (2008) Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. *Nat. Genet.* **40**, 189–197
5. Schadt, E. E., Molony, C., Chudin, E., Hao, K., Yang, X., Lum, P. Y., et al. (2008) Mapping the genetic architecture of gene expression in human liver. *PLoS Biol.* **6**, e107
6. Cano-Gamez, E., and Trynka, G. (2020) From GWAS to function: using functional genomics to identify the mechanisms underlying complex diseases. *Front. Genet.* **11**, 424
7. Keller, B., Martini, S., Sedor, J., and Kretzler, M. (2010) Linking variants from genome-wide association analysis to function via transcriptional network analysis. *Semin. Nephrol.* **30**, 177–184
8. Novikova, G., Andrews, S. J., Renton, A. E., and Marcora, E. (2021) Beyond association: successes and challenges in linking non-coding genetic variation to functional consequences that modulate Alzheimer's disease risk. *Mol. neurodegener.* **16**, 27
9. Backman, J. D., Li, A. H., Marcketta, A., Sun, D., Mbatchou, J., Kessler, M. D., et al. (2021) Exome sequencing and analysis of 454,787 UK Biobank participants. *Nature* **599**, 628–634
10. Karczewski, K. J., Solomonson, M., Chao, K. R., Goodrich, J. K., Tiao, G., Lu, W., et al. (2022) Systematic single-variant and gene-based association testing of thousands of phenotypes in 394,841 UK Biobank exomes. *Cell Genom.* **2**, 100168
11. Sun, B. B., Kurki, M. I., Foley, C. N., Mechakra, A., Chen, C. Y., Marshall, E., et al. (2022) Genetic associations of protein-coding variants in human disease. *Nature* **603**, 95–102
12. Wang, Q., Dhindsa, R. S., Carss, K., Harper, A. R., Nag, A., Tachmazidou, I., et al. (2021) Rare variant contribution to human disease in 281,104 UK Biobank exomes. *Nature* **597**, 527–532
13. Xu, H., Ryan, K. A., Jaworek, T. J., Southam, L., Reid, J. G., Overton, J. D., et al. (2017) Familial Hypercholesterolemia and Type 2 diabetes in the Old Order Amish. *Diabetes* **66**, 2054–2058
14. He, S., Ryan, K. A., Streeten, E. A., McArdle, P. F., Daue, M., Trubiano, D., et al. (2020) Prevalence, control, and treatment of diabetes, hypertension, and high cholesterol in the Amish. *BMJ Open Diabetes Res. Care* **8**, e000912
15. Montasser, M. E., Van Hout, C. V., Milosic, L., Howard, A. D., Rosenberg, A., Callaway, M., et al. (2021) Genetic and functional evidence links a missense variant in *B4GALT1* to lower LDL and fibrinogen. *Science* **374**, 1221–1227
16. Sun, C., VanRaden, P. M., Cole, J. B., and O'Connell, J. R. (2014) Improvement of prediction ability for genomic selection of dairy cattle by including dominance effects. *PLoS One* **9**, e103934
17. Pollin, T. I., Damcott, C. M., Shen, H., Ott, S. H., Shelton, J., Hornestein, R. B., et al. (2008) A null mutation in human *APOC3* confers a favorable plasma lipid profile and apparent cardioprotection. *Science* **322**, 1702–1705
18. Friedewald, W. T., Levy, R. I., and Fredrickson, D. S. (1972) Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. *Clin. Chem.* **18**, 499–502
19. Mora, S., Caulfield, M. P., Wohlgenuth, J., Chen, Z., Superko, H. R., Rowland, C. M., et al. (2015) Atherogenic lipoprotein subfractions determined by ion mobility and first cardiovascular events after random allocation to high-intensity statin or placebo: the justification for the use of statins in prevention: an intervention trial evaluating rosuvastatin (JUPITER) trial. *Circulation* **132**, 2220–2229
20. Caulfield, M. P., Li, S., Lee, G., Blanche, P. J., Salameh, W. A., Benner, W. H., et al. (2008) Direct determination of lipoprotein particle sizes and concentrations by ion mobility analysis. *Clin. Chem.* **54**, 1307–1316
21. Caulfield, M. P., Li, S., Lee, G., Blanche, P. A., Salameh, W. A., Benner, W. H., et al. (2008) In reply. *Clin. Chem.* **54**, 2088–2089
22. Choi, Y. J., Roberts, B. K., Wang, X., Geaney, J. C., Naim, S., Wojnoonski, K., et al. (2012) Effects of the PPAR- δ agonist MBX-8025 on atherogenic dyslipidemia. *Atherosclerosis* **220**, 470–476
23. Aulchenko, Y. S., Ripatti, S., Lindqvist, I., Boomsma, D., Heid, I. M., Pramstaller, P. P., et al. (2009) Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. *Nat. Genet.* **41**, 47–55
24. Sabatti, C., Service, S. K., Hartikainen, A. L., Pouta, A., Ripatti, S., Brodsky, J., et al. (2009) Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. *Nat. Genet.* **41**, 35–46
25. Teslovich, T. M., Musunuru, K., Smith, A. V., Edmondson, A. C., Stylianou, I. M., Koseki, M., et al. (2010) Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**, 707–713
26. Sandhu, M. S., Waterworth, D. M., Debenham, S. L., Wheeler, E., Papadakis, K., Zhao, J. H., et al. (2008) LDL-cholesterol concentrations: a genome-wide association study. *Lancet* **371**, 483–491
27. Meroufel, D. N., Mediene-Benchekor, S., Lardjam-Hetraf, S. A., Ouhaibi-Djellouli, H., Boulouner, H., Hamani-Medjaoui, I., et al. (2015) Associations of common SNPs in the *SORT1*, *GCKR*, *LPL*, *APOA1*, *CETP*, *LDLR*, *APOE* genes with lipid trait levels in an Algerian population sample. *Int. J. Clin. Exp. Pathol.* **8**, 7358–7363
28. Buyske, S., Wu, Y., Carty, C. L., Cheng, I., Assimes, T. L., Dumitrescu, L., et al. (2012) Evaluation of the metabochip genotyping array in African Americans and implications for fine mapping of GWAS-identified loci: the PAGE study. *PLoS One* **7**, e35651
29. Shahid, S. U., Cooper, J. A., Beaney, K. E., Li, K., Rehman, A., Humphries, S. E., et al. (2016) Effect of *SORT1*, *APOB* and *APOE* polymorphisms on LDL-C and coronary heart disease in Pakistani subjects and their comparison with Northwick Park Heart Study II. *Lipids Health Dis.* **15**, 83
30. Graham, S. E., Clarke, S. L., Wu, K. H., Kanoni, S., Zajac, G. J. M., Ramdas, S., et al. (2021) The power of genetic diversity in genome-wide association studies of lipids. *Nature* **600**, 675–679
31. Sievers, F., and Higgins, D. G. (2018) Clustal Omega for making accurate alignments of many protein sequences. *Protein Sci.* **27**, 135–145
32. Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., et al. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539
33. Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M., and Barton, G. J. (2009) Jalview Version 2—a multiple sequence

- alignment editor and analysis workbench. *Bioinformatics*. **25**, 1189–1191
34. Quistgaard, E. M., Madsen, P., Groftehauge, M. K., Nissen, P., Petersen, C. M., and Thirup, S. S. (2009) Ligands bind to sortilin in the tunnel of a ten-bladed beta-propeller domain. *Nat. Struct. Mol. Biol.* **16**, 96–98
 35. Schrodinger, L. (2015) *The PyMOL Molecular Graphics System, Version 2.5.2*. Schrodinger, LLC, New York, NY
 36. Madeira, F., Pearce, M., Tivey, A. R. N., Basutkar, P., Lee, J., Edbali, O., *et al.* (2022) Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Res.* **50**, W276–W279
 37. Leloup, N., Lossl, P., Meijer, D. H., Brennich, M., Heck, A. J. R., Thies-Weesie, D. M. E., *et al.* (2017) Low pH-induced conformational change and dimerization of sortilin triggers endocytosed ligand release. *Nat. Commun.* **8**, 1708
 38. Madej, T., Marchler-Bauer, A., Lanczycki, C., Zhang, D., and Bryant, S. H. (2020) Biological assembly comparison with VAST. *Methods Mol. Biol.* **2112**, 175–186
 39. Bellenguez, C., Kucukali, F., Jansen, I. E., Kleideidam, L., Moreno-Grau, S., Amin, N., *et al.* (2022) New insights into the genetic etiology of Alzheimer's disease and related dementias. *Nat. Genet.* **54**, 412–436
 40. Reich, D. E., and Lander, E. S. (2001) On the allelic spectrum of human disease. *Trends Genet.* **17**, 502–510
 41. Campos, H., Arnold, K. S., Balestra, M. E., Innerarity, T. L., and Krauss, R. M. (1996) Differences in receptor binding of LDL subfractions. *Arterioscler. Thromb. Vasc. Biol.* **16**, 794–801
 42. Nigon, F., Lesnik, P., Rouis, M., and Chapman, M. J. (1991) Discrete subspecies of human low density lipoproteins are heterogeneous in their interaction with the cellular LDL receptor. *J. Lipid Res.* **32**, 1741–1753
 43. Sakai, N., Matsuzawa, Y., Hirano, K., Yamashita, S., Nozaki, S., Ueyama, Y., *et al.* (1991) Detection of two species of low density lipoprotein particles in cholesteryl ester transfer protein deficiency. *Arterioscler. Thromb.* **11**, 71–79
 44. Dinchuk, J., Hart, J., Gonzalez, G., Karmann, G., Schmidt, D., and Wirak, D. O. (1995) Remodelling of lipoproteins in transgenic mice expressing human cholesteryl ester transfer protein. *Biochim. Biophys. Acta.* **1255**, 301–310
 45. Greeve, J., Altkemper, I., Dieterich, J. H., Greten, H., and Windler, E. (1993) Apolipoprotein B mRNA editing in 12 different mammalian species: hepatic expression is reflected in low concentrations of apoB-containing plasma lipoproteins. *J. Lipid Res.* **34**, 1367–1383
 46. Kjolby, M., Andersen, O. M., Breiderhoff, T., Fjorback, A. W., Pedersen, K. M., Madsen, P., *et al.* (2010) Sortl, encoded by the cardiovascular risk locus lp13.3, is a regulator of hepatic lipoprotein export. *Cell Metab.* **12**, 213–223
 47. Brown, M. S., and Goldstein, J. L. (1986) A receptor-mediated pathway for cholesterol homeostasis. *Science*. **232**, 34–47
 48. Marcel, Y. L., Innerarity, T. L., Spilman, C., Mahley, R. W., Protter, A. A., and Milne, R. W. (1987) Mapping of human apolipoprotein B antigenic determinants. *Arteriosclerosis*. **7**, 166–175
 49. Gordon, S. M., Li, H., Zhu, X., Shah, A. S., Lu, L. J., and Davidson, W. S. (2015) A comparison of the mouse and human lipoproteome: suitability of the mouse model for studies of human lipoproteins. *J. Proteome Res.* **14**, 2686–2695
 50. de Silva, H. V., Lauer, S. J., Wang, J., Simonet, W. S., Weisgraber, K. H., Mahley, R. W., *et al.* (1994) Overexpression of human apolipoprotein C-III in transgenic mice results in an accumulation of apolipoprotein B48 remnants that is corrected by excess apolipoprotein E. *J. Biol. Chem.* **269**, 2324–2335
 51. Carlo, A. S., Gustafsen, C., Mastrobuoni, G., Nielsen, M. S., Burgert, T., Hartl, D., *et al.* (2013) The pro-neurotrophin receptor sortilin is a major neuronal apolipoprotein E receptor for catabolism of amyloid-beta peptide in the brain. *J. Neurosci.* **33**, 358–370
 52. Sparks, R. P., Arango, A. S., Jenkins, J. L., Guida, W. C., Tajkhorshid, E., Sparks, C. E., *et al.* (2020) An allosteric binding site on sortilin regulates the trafficking of VLDL, PCSK9, and LDLR in Hepatocytes. *Biochemistry*. **59**, 4321–4335
 53. Quistgaard, E. M., Groftehauge, M. K., Madsen, P., Pallesen, L. T., Christensen, B., Sorensen, E. S., *et al.* (2014) Revisiting the structure of the Vps10 domain of human sortilin and its interaction with neurotensin. *Protein Sci.* **23**, 1291–1300
 54. Jones, G. T., Bown, M. J., Gretarsdottir, S., Romaine, S. P., Helgadottir, A., Yu, G., *et al.* (2013) A sequence variant associated with sortilin-1 (SORT1) on lp13.3 is independently associated with abdominal aortic aneurysm. *Hum. Mol. Genet.* **22**, 2941–2947
 55. Folkersen, L., van't Hooft, F., Chernogubova, E., Agardh, H. E., Hansson, G. K., Hedin, U., *et al.* (2010) Association of genetic risk variants with expression of proximal genes identifies novel susceptibility genes for cardiovascular disease. *Circ. Cardiovasc. Genet.* **3**, 365–373
 56. Zeller, T., Wild, P., Szymczak, S., Rotival, M., Schillert, A., Castagne, R., *et al.* (2010) Genetics and beyond-the transcriptome of human monocytes and disease susceptibility. *PLoS One*. **5**, e10693
 57. Arvind, P., Nair, J., Jambunathan, S., Kakkar, V. V., and Shanker, J. (2014) CELSR2-PSRC1-SORT1 gene expression and association with coronary artery disease and plasma lipid levels in an Asian Indian cohort. *J. Cardiol.* **64**, 339–346
 58. Nielsen, M. S., Jacobsen, C., Olivecrona, G., Gliemann, J., and Petersen, C. M. (1999) Sortilin/neurotensin receptor-3 binds and mediates degradation of lipoprotein lipase. *J. Biol. Chem.* **274**, 8832–8836
 59. Kathiresan, S., Willer, C. J., Peloso, G. M., Demissie, S., Musunuru, K., Schadt, E. E., *et al.* (2009) Common variants at 30 loci contribute to polygenic dyslipidemia. *Nat. Genet.* **41**, 56–65