Functional Characterization of Lipid Regulatory Effects of Three Genes Using Knockout Mouse Models

Chen Yao^{1,2}, Holly Savage³, Tong Hao^{4,5}, Gha Young Lee^{1,2}, Yuka Takemon³, Wenting Bian^{4,5}, David E Hill^{4,5}, Marc Vidal^{4,5}, Ron Korstanje^{3*} and Daniel Levy^{1,2*}

¹The Framingham Heart Study, Framingham, MA;

²The Population Sciences Branch, Division of Intramural Research, National Heart, Lung, and Blood Institute, Bethesda, MD;

³The Jackson Laboratory, Bar Harbor, ME;

⁴Center for Cancer Systems Biology (CCSB) and Dept. of Cancer Biology, Dana-Farber Cancer Institute, Boston MA, USA; ⁵Dept. of Genetics, Blavatnik Institute, Harvard Medical School, Boston MA

***Corresponding author:** Daniel Levy, MD
Email: levyd@nih.gov

Abstract:

Integrative analysis that combines genome-wide association data with expression quantitative trait analysis and network representation may illuminate causal relationships between genes and diseases. To identify causal lipid genes, we utilized genotype, gene expression, protein-protein interaction networks, and phenotype data from 5,257 Framingham Heart Study participants and performed Mendelian randomization to investigate possible mechanistic explanations for observed associations. We selected three putatively causal candidate genes (*ABCA6*, *ALDH2*, and *SIDT2*) for lipid traits (LDL cholesterol, HDL cholesterol and triglycerides) in humans and conducted mouse knockout studies for each gene to confirm its causal effect on the corresponding lipid trait. We conducted the RNA-seq from mouse livers to explore transcriptome-wide alterations after knocking out the target genes. Our work builds upon a lipid-related gene network and expands upon it by including protein-protein interactions. These resources, along with the innovative combination of emerging analytical techniques, provide a groundwork upon which future studies can be designed to more fully understand genetic contributions to cardiovascular diseases.

Introduction

Cardiovascular disease (CVD) remains the leading cause of death worldwide¹. A key goal of CVD research is the identification of specific genes and genetic variants that contribute to the disease². Genome-wide association studies (GWAS) have revealed numerous inherited DNA sequence variants associated with CVD and its risk factors³⁻⁵. For many known CVD-associated variants, however, the biological mechanisms and underlying causal genes remain largely unknown. The overwhelming majority of CVD-related genetic variants lie within intergenic or noncoding regions of the genome³, indicating that these single nucleotide polymorphisms (SNPs) impact CVD risk indirectly, for example, by altering transcription levels of nearby (*cis*) or remote (*trans*) genes. The identification of genetic variants that alter gene expression – expression quantitative trait loci (eQTLs) – has paved the way for functional studies linking candidate genes to disease⁶.

We postulated *a priori* that a network approach could portray biological systems underlying CVD traits and provide insights into disease-related genes and pathways. Network-based approaches have many biological and clinical applications and have yielded promising results in recent studies^{7, 8}. Not only do networks allow for the integration of different types of biological information, but they also allow for unbiased discovery of disease genes that, when integrated with protein-protein interactions (PPI), can provide mechanistic explanations for diseases⁹. In a previous study, we linked 21 CVD traits based on their shared SNP associations, and in doing so, recapitulated the clustering of metabolic risk factors observed in epidemiological studies¹⁰. Here, we expand that network model to identify causal genes and pathways for lipid traits and tested three genes for causality.

By integrating high-quality PPI data¹¹ with SNPs associated in GWAS of lipids, gene expression, and fasting blood lipid levels in 5,257 Framingham Heart Study (FHS) participants (Supplementary table 1), we selected *ABCA6*, *ALDH2*, and *SIDT2* as candidate causal genes for lipid traits (low-density lipoprotein (LDL) cholesterol, high-density lipoprotein (HDL) cholesterol, and triglycerides (TG)). Dysregulation of circulating lipid levels has been implicated in the pathogenesis of CVD¹². To confirm that the candidate genes identified from the network model are causal for the corresponding lipid traits, we created mouse knockout (KO) models to assess, *in vivo*, the consequences of perturbing key genes as well as the networks in which they function. The mouse KO models were therefore phenotyped for the corresponding lipid traits and additionally liver gene expression was measured by RNA-Seq to study perturbations within the network (Figure 1).

We posit that creating a network based on the relationships among genes, traits, and PPIs can further our understanding of the genetic basis of lipid traits and their contributions to lipid dysregulation. By combining human genotype data with mouse KO experiments, we also provide a framework for selecting and validating lipid-related candidate genes.

Results

Identifying lipid trait-related expression quantitative loci

We identified 4670 (84%) lipid trait-associated expression quantitative loci (eQTLs; SNPs association with gene expression) by linking 5568 lipid trait GWAS SNPs (Supplementary Table 2) that were also associated with gene expression in whole blood in 5,257 Framingham participants¹³ (Supplementary Table 3). Seventy-two percent of these eQTLs affected the expression of a nearby (*cis*) transcript (SNP located within 1 Mb of the transcript), suggesting that

they may play a role in regulating gene expression. For these eQTLs, we conducted mediation testing to detect if the association of the GWAS SNP with the corresponding trait was mediated by gene expression. At $P < 0.005$, we identified 464 SNPs with significant mediation effects on lipids traits (Supplementary Table 4).

Expanding the candidate gene network

To expand the candidate gene network, we built a CVD gene network by incorporating all protein coding genes containing CVD GWAS SNPs. Our CVD-centric network contained binary PPIs for 846 CVD-associated proteins (out of ~1,300 tested) extracted from a dataset of ~58,000 binary PPIs among 10,690 human proteins obtained from both systematic binary mapping and literature curations¹⁴. The 846 CVD-associated proteins and their first-degree neighbors constituted a network of ~8,600 interactions among 4,336 proteins (Supplementary Table 5), including 444 intra-CVD PPIs among 349 lipids-associated proteins (Supplementary Table 6). This expanded network is amenable to investigating disease models by analyzing other features such as shared gene ontology (GO) terms, shared expression profiles, and functional similarity.

Ranking candidate genes

Combining GWAS lipid trait SNPs, eQTLs, and mediation results, we selected seven candidate genes to test against three lipid traits in Mendelian randomization (MR) analyses: *DOCK7*, *TAGLN*, *SIDT2*, *ALDH2*, *SLC44A4*, *ABCA6*, and *ATG4C* (See Methods for gene selection). We used genetic variants (focusing on those that are eQTLs variants) as instruments to investigate the causal relations between gene expression and lipid phenotypes. The causal effects of the seven genes were further tested by MR using independent GWAS data from the Global Lipids Genetics Consortium (GLGL)¹⁵ (Table 1) and filtered based on 1) gene novelty (not previously studied in a mouse model for cardiovascular phenotypes), 2) repository availability of knockout sperm or embryos that have been shown to produce viable animals, and 3) interaction with CVD-related proteins in an integrated network based on PPIs. This approach identified three genes as suitable for detailed mouse KO experiments: *ABCA6* for non-HDL, *ALDH2* for HDL, and *SIDT2* for TG (Table 2). Though *ABCA6* was not found to be causally associated with lipids traits by MR. rs740516, a variant in *ABCA6*, was significantly associated with LDL in GWAS ($P = 7 \times 10^{-9}$) and was associated with expression of *ABCA6* ($P = 3.3 \times 10^{-5}$). *SIDT2* has been previously studied in relation to lipid metabolism and homeostasis, our finding is one of the first to utilize an integrative genomics framework to better understand the genetic basis of its effect on lipid traits^{16, 17}. The effects of these genes on secondary lipid traits (*ABCA6* on HDL and TG, *ALDH2* on non-HDL and TG, *SIDT2* on HDL and non-HDL) are also summarized.

ABCA6, *ALDH2*, *SIDT2*, and the corresponding mouse knockout models

ABCA6 Knockout

The protein ABCA6, encoded by *ABCA6*, is a member of the superfamily of ATP-binding cassette (ABC) transporters. ABC proteins transport various substrates, including lipids, peptides, vitamins, and ions¹⁸, across extra- and intracellular membranes. *ABCA6* is expressed exclusively in multicellular eukaryotes and has been suspected of acting as an intracellular transporter in macrophage lipid homeostasis¹⁹.

Comparing total cholesterol and non-HDL cholesterol levels (which are primarily comprised of LDL cholesterol) between C57BL/6N-*Abca6*^{Gtm2a(KOMP)Wtsi}/TcpRkorJ mice (*Abca6* knockout) and WT C57BL/6N mice, we observed a significant increase in total cholesterol in males on the chow

diet at 22 weeks of age and in males on the high-fat diet at both 18 and 22 weeks of age (Figure 2). A similar trend was identified in females.

In addition, knockout of *ABCA6* led to a significant increase in HDL cholesterol in male mice on the chow diet but not on a high-fat diet with a similar but nonsignificant trend observed in females. This finding is supported by previous studies that have shown that *ABCA6* is involved in reverse cholesterol transport, specifically by facilitating phospholipid and cholesterol export from the cell^{19, 20}. There were no significant differences in TG levels in the female mice on both chow and high-fat diet, but the male *Abca6*^{-/-} mice had lower TG levels on the chow diet.

ALDH2 Knockout

ALDH2 encodes aldehyde dehydrogenase, the second enzyme of the major oxidative pathway of alcohol metabolism. This gene encodes a mitochondrial isoform, which has a high affinity for acetaldehydes and is localized in the mitochondrial matrix. Deficiencies in *ALDH2* have been shown to be correlated with differences in lipid levels due to its role in the metabolism of lipid-peroxidation-derived aldehydes²¹.

ALDH2 was causally related to all three lipids in MR (Table 1). In addition, SNPs in *ALDH2* were significantly associated with HDL and LDL in GWAS and gene expression in FHS (Supplementary Table 3). Comparing HDL cholesterol levels between B6DnK;B6N-*Aldh2*^{tm1a(EUCOMM)Wtsi/J}egRkorJ mice (*Aldh2* knockout) and WT C57BL/6N mice, we found significantly increased HDL cholesterol levels in both female and male *Aldh2* KO mice at several time points compared to controls when fed the chow diet (Figure 3). On a high-fat diet, we no longer observed a difference in males, but there was a significant difference in females at 18 weeks of age.

Except for a few specific time points, there were no differences in non-HDL levels for either diet in males or females. There were no differences in TG levels in the female mice, but in the male mice we observed increased TG levels in the KO animals on a high-fat diet. This suggests that the knockout of *Aldh2* in males might lead to a decreased risk of CVD on the chow diet, but an increased risk on the high-fat diet.

SIDT2 Knockout

SIDT2 is a transmembrane protein that predominantly localizes on lysosomes but is also detectable in the plasma membrane of human embryonic kidney cells. Overexpression of *SIDT2* in some cells is accompanied by a significant reduction of detectable lysosomes, indicating that the overexpressed protein leads to lysosomal dysfunction. Lysosomes are thought to be the major intracellular compartment for the degradation of macromolecules and play an important role in regulating lipid degradation pathways^{22, 23}. Previous *SIDT2*-knockout experimentation has shown that *SIDT2* plays a major role in regulating lipid autophagy and metabolism and cholesterol and triglyceride transport in mammalian cells, particularly in the liver^{16, 17}.

SIDT2 was significant in MR test for all three lipids traits. SNPs in *SIDT2* were significantly associated with LDL in GWAS and gene expression in FHS (Supplementary Table 3). We identified *SIDT2* in our human cohort to be associated primarily with triglycerides and therefore compared TG levels between B6;129S5-*Sid2*^{tm1Lex}/MmucdRkorJ mice (*Sid2* knockout) and WT control littermates. We observed a significant increase in TG levels in the female mice on the chow diet at 18 and 22 weeks of age (Figure 4). Male TG levels were only increased at 8 weeks of age.

In addition, HDL cholesterol levels were significantly decreased in male and female KO mice on the high-fat diet. On a chow diet, HDL cholesterol levels appeared to decrease even further in male mice. HDL cholesterol levels, however, increased slightly in females. Female non-HDL cholesterol levels did not change on the chow diet; in the early stages there was a large decrease in non-HDL cholesterol on the high-fat diet that ultimately dissipated. Male non-HDL cholesterol levels did not change on the high-fat diet and significantly increased on a chow diet.

Expression changes after gene knock out

To explore transcriptome-wide alterations resulting from knocking out the target genes, we conducted RNA-Seq for *ALDH2* and *ABCA6* in three male and three female knockout animals and three male and three female control littermates for each of the two diets. We didn't conduct the RNA-Seq for *SIDT2* in light of previously reported studies of this knockout^{16,17}. At FDR<0.05 (T test, P<0.014), we identified an average of 238 differentially-expressed genes (ranging from 69 to 403 for eight comparisons) after knockout (Supplementary Table 7). The differentially-expressed genes were analyzed through the use of Ingenuity Pathway Analysis (IPA, QIAGEN Inc.)²⁴. Function enrichment analyses were conducted on the differentially expressed genes using curated information from the QIAGEN Knowledge Base²⁴. Though canonical pathways of differentially expressed genes from mice of different diet conditions are enriched in different metabolic and cell signaling pathways, the affected diseases and molecular functions are highly consistent with regard to metabolic disease and lipid metabolism (Supplementary figure 1). The top toxicological functions were liver steatosis and cardiac dysfunction. Network analysis based on Ingenuity Knowledge Base²⁴ revealed that *Abca6* affected cholesterol level via cytokine IFNG (interferon gamma) and that *Aldh2* affected cholesterol level through enzyme GNMT (glycine N-methyltransferase) and acetaldehyde (Supplementary Figure 2). Comparing RNA-seq data from *Abca6* and *Aldh2* knockout of the same sex and diet, we found 19%-60% overlap of differentially expressed genes. Using IPA functional enrichment at P<0.005, we found that the overlapping genes are enriched in lipid metabolism (Supplementary Table 8), suggesting a common pathway affected by these two very different genes.

Discussion

To represent the complex molecular dynamics that underlie CVD, we constructed a network that integrates SNPs, gene expression, protein-protein interactions, and lipid phenotypes. From this integrated network approach, we identified three genes – *ABCA6*, *ALDH2*, and *SIDT2* – that were linked to lipid levels and were investigated further through mouse knockout models. The mouse experiments recapitulated network-predicted relations of all three genes to the corresponding lipid traits and provided experimental support for the integrative approach we developed.

Disease processes involve many interacting molecules and physical and biochemical processes, and thus the analysis of single data types is often insufficient to explain the etiology of complex traits. For example, carriers of risk alleles do not always have phenotypic consequences as such genetic variants do not necessarily alter the expression of disease-related genes or proteins. Therefore, in order to draw a more comprehensive view of biological processes, high-throughput data from different elements of multidimensional molecular information must be integrated and analyzed.

Results from our mouse knockout models provide proof of principle for the utility of the integrative network approach that we developed. Results of the mouse experiments largely support our network model, with a few exceptions. Knockout of *Abca6*, which contains GWAS SNPs for

LDL cholesterol and total cholesterol, resulted in higher total cholesterol (which is largely LDL cholesterol) and HDL cholesterol in male mice, suggesting a potential role of this gene in CVD development. *ABCA6* plays a key role in macrophage lipid homeostasis by facilitating phospholipid and cholesterol export from the cell^{19, 20}. *ABCA6* dysfunction may cause a buildup of intracellular phospholipids and cholesterol. Knockout of *Aldh2*, which was chosen for its putatively causal association with HDL cholesterol, resulted in significantly increased HDL cholesterol levels in both female and male mice. Aldehyde dehydrogenase, the protein encoded by *ALDH2*, is an enzyme in the oxidative pathway and plays a major role in the metabolism of lipid peroxidation-derived aldehydes²¹. In our knockout models, the absence of *Aldh2* resulted in higher lipid levels, which supports the idea that *ALDH2* is involved in lipid degradation/clearance. Knockout of *Sid12*, which contains GWAS SNPs for HDL cholesterol, resulted in a significant increase in triglyceride levels in the females on the chow diet at 18 and 22 weeks of age. *SID12* plays a crucial role in the uptake and intracellular transport of triglycerides and cholesterol^{22, 23}. Previous studies¹⁶ have shown that the absence of *Sid12* causes increased serum triglycerides and free fatty acids in mice. After analyzing the knockout results, we compared them with expression data for the corresponding human genes. We discovered that while each gene knockout produced some significantly different results in the female and male mice, FHS human gene expression data did not reveal sex differences in lipid effects for any of the genes ($P > 0.05$, data not shown). This may be due to innate differences in lipid metabolism and regulation between the two species.

In contrast to similar studies that have relied on public databases to identify eQTLs²⁵, our study takes a direct and thorough approach, analyzing extensive phenotypic datasets in relation to gene expression in human peripheral blood. This method, combined with mediation analysis, allowed us not only to identify significant pathway enrichment from GWAS SNPs, but also to investigate mechanisms underlying lipids traits. Our study is also unique in that we combined several emerging techniques to identify and validate genes; these techniques included the use of functional eQTL studies and the integration of PPI data. Functional studies, especially those involving the use of networks to mirror biological systems, have shown great promise in the ability to identify candidate genes for various diseases²⁶. In one study, Atanasovska et. al used a network-based approach to prioritize genes for hundreds of cardiometabolic SNPs to identify disease-predisposing genes²⁷. The integration of PPI with eQTL analyses also has great potential to offer mechanistic explanations for diseases^{11, 14}.

One limitation of our study is the use of whole blood for expression profiling. Although the lipid traits were measured in blood, and as such, whole blood-derived eQTLs may be highly relevant to the phenotypes we studied, in some cases, we observed opposite direction between MR predicted effects and mouse knockout (*ALDH2* on HDL). Another limitation is that our MR analyses could not infer sex-specific causal effects because the underlying GWAS studies were conducted in pooled-sex analyses, even though our mouse experiments revealed notable sex differences. Differential lipid metabolic responses between male and female mice have also been previously identified in other gene knockout studies²⁸.

In summary, our network approach allowed us to identify novel candidate genes that may contribute to CVD via lipid effects. As such, these genes represent attractive targets for the treatment of dyslipidemia and the prevention of CVD. Looking forward, we plan to continue to expand the CVD network and investigate additional causal genes. As we do so, we anticipate that it will increasingly explain genes, pathways, and mechanisms underlying CVD and point toward promising precision drug targets. Our integrative network also provides insights into how

dysfunction of novel CVD-associate genes is manifested at the molecular level²⁹. Approaches similar to those used in this study may not only be expanded upon in relation to CVD but also be used to investigate other diseases impacted by genetic and epigenetic factors. To make future studies more comprehensive and accurate, it will be important to expand eQTL databases to other disease-related tissues and to include additional clinically important traits as well.

Materials and Methods

Samples and Phenotypes

In 1948, the FHS started recruiting participants (original cohort) from Framingham, MA to begin the first round of extensive physical examinations and lifestyle surveys to investigate CVD and its risk factors³⁰. In 1971 and 2002, the FHS recruited offspring (and their spouses) and adult grandchildren of the original cohort participants into the offspring and third-generation cohorts, respectively³¹⁻³³. Plasma total cholesterol, HDL cholesterol, triglycerides, and glucose were measured in the morning after an eight-hour overnight fasting. Body mass index (BMI) was defined as weight (kilograms) divided by height squared (meters²). Current smokers were defined as those who smoked, on average, at least one cigarette per day during the year prior to the FHS clinical assessment. Clinical characteristics of the study sample are summarized in Supplementary Table 1. All participants from the FHS gave informed consent for participation in this study and for the collection of plasma and DNA for analysis. The FHS study protocol was approved by Boston Medical Center: Protocol ID: H-27984, Boston University Medical Center IRB (BUMC IRB) Title: FRAMINGHAM HEART STUDY BIOMARKER PROJECT.

Genotype data: 5,568 SNPs that were associated with the three lipid traits – HDL cholesterol, LDL cholesterol, and triglycerides – at $p \leq 5 \times 10^{-8}$ (in the GRASP database, downloaded in June 2016, Supplementary Table 2) were curated and matched with the FHS 1000 Genomes Project imputed genotype data³⁴. SNPs with imputed quality score (r^2) < 0.3 and minor allele frequency (MAF) < 0.01 were excluded, resulting in 4173 genome-wide significant SNPs for eQTL analysis.

Gene expression: Whole blood was collected in PAXgeneTM tubes (PreAnalytiX, Hombrechtikon, Switzerland) and frozen at -80°C . RNA was extracted using a whole blood RNA System Kit (Qiagen, Venlo, Netherlands) and mRNA expression profiling was assessed using the Affymetrix Human Exon 1.0 ST GeneChip platform (Affymetrix Inc, Santa Clara, CA), which contains more than 5.5 million probes targeting the expression of 17,873 genes. The Robust Multi-array Average (RMA) package³⁵ was used to normalize the gene expression values and remove any technical or spurious background variation. Linear regression models were used to adjust for technical covariates (batch, first principal component, and residual mean of all probesets).

Identification of LIPIDS-associated eQTLs

The eQTLs were identified from FHS genotype data and whole blood gene expression as described previously¹³. eQTL analyses were conducted in two phases: 1) gene expression residuals were generated after accounting for the effects of sex, age, platelet count, white blood cell count, and imputed differential blood cell counts; analyses were performed using R version 3.0.1 with a mixed-effect modeling package that adjusted for familial relationships; and 2) underlying confounding factors were accounted for by 20 Probabilistic Estimation of Expression Residuals (PEERs)³⁶ factors that were computed using the residualized expression data. The residualized expression was fit to a linear model using the PEER factors along with sex, age, and effect allele dosages. The algorithm was implemented with Graphical Processing Units (GPUs). *cis*-eQTLs

were defined as SNPs that reside within 1 Mb up or downstream of the transcription start site. False discovery rate (FDR) computations for *cis*- and *trans*-eQTLs were computed separately. SNPs at FDR <0.05 were considered statistically significant eQTLs.

Putative causality was first tested by mediation tests, in which the underlying mechanism for a relationship between two variables is tested by introducing a third explanatory variable. The analysis was conducted with the Mediation Package³⁷ in R where the “exposure” variable represented a SNP, the “mediator” or “explanatory variable” represented gene expression, and the “outcome” variable represented the phenotype. The mediation effect was measured on a scale from 0-100% where a 100% mediation effect indicated that the entire relationship between a SNP and a phenotype (direct effect) was explained by changes in gene expression, i.e., the “mediator” or the “explanatory variable.” Significant mediation effects were selected at a permutation p-value <0.005 (based on 1000 permutations).

Significant mediation effects were further tested by Mendelian randomization (MR)³⁸ to determine causal associations between gene expression and phenotype. MR uses common genetic variants with well-understood effects on an exposure as instrumental variables (IV) to infer causality of an exposure to an outcome. This approach was applied to the aforementioned three lipid traits (LDL cholesterol, HDL cholesterol, and triglycerides). Here, the sentinel *cis*-eQTL (top eQTL located within 1 Mb of the tested gene), based on the lowest SNP-gene expression p-value from the 1000G GWAS, was selected as the IV for its corresponding gene in MR analysis. Based on the association between the sentinel *cis*-eQTL and the summary statistics from the three lipid GWAS of 188,577 individuals obtained from the Global Lipids Genetics Consortium (GLGC)¹⁵, the MR analysis was performed using the MR Base package³⁹ and the two-sample MR method³⁸.

Integration with PPI Networks

Binary PPIs were extracted from a systematically generated and literature-curated datasets, which in total contain ~58,000 PPIs among 10,690 human proteins^{11, 14}. Putatively causal variants from genotyping and gene expression analysis were integrated with PPI networks to assess the extent of potential network perturbations associated with disruption of PPIs or altered expression⁹. Network perturbation information on missense alleles for various lipids-associated genes, which assesses the degree to which a mutant protein exhibits an altered spectrum of PPIs relative to the WT protein and all other mutants^{11, 14}, was also included. For proteins without direct (physical) interactions, we use predicted protein-protein interactions from the STRING database to expand the network⁴⁰.

Identifying Candidate Genes

All candidate genes were assessed by each of the each of the following criteria: 1) Does the candidate gene contain a SNP(s) associated with a lipid trait in GWAS ($p < 5 \times 10^{-8}$) and is the GWAS SNP(s) is also an eQTL? 2) Is expression of the eQTL-associated gene also associated with the same lipid trait? 3) Are genetic effects on the lipid trait mediated by the expression of the eQTL-associated gene? 4) Does the gene test positive in MR ($p < 0.05$)? 5) Have any interactions with CVD related proteins?

We then ranked the genes based on how many of the criteria were met and further excluded any genes if the knockout was known to be lethal based on existing literature. Three genes *ABCA6*, *ALDH2*, and *SIRT2* met more than three of these selection criteria and were selected for further investigation.

Validation of Candidate Genes in Animal Models

To test the hypothesis that key genes identified in the LIPIDS networks can be validated in mouse models for the corresponding traits, we obtained KO mouse strains for the three genes identified as having causal effects on HDL cholesterol, LDL cholesterol, and triglycerides, respectively. C57BL/6N-*Abca6*^{tm2a(KOMP)Wtsi}/TcpRkorJ, B6Dnk;B6N-*Aldh2*^{tm1a(EUCOMM)Wtsi}/IegRkorJ, and B6;129S5-*Sidt2*^{tm1Lex}/MmucdRkorJ KO mice were generated at The Jackson Laboratory using sperm or embryos provided by the International Knockout Mouse Consortium and maintained on a C57BL/6NJ genetic background. All animals were housed at The Jackson Laboratory, which is approved by the American Association for Accreditation of Laboratory Animal Care. Animals were kept on a 12-hour (6am-6pm) light/dark cycle with a room temperature between 68 and 72°F, and either fed a chow diet (5K52, LabDiet) or a high-fat diet (TD.06414, Teklad Custom Diet).

Cohorts of 20 males and 20 females of both KO strains and C57BL/6NJ control mice were raised on the chow diet. At 12 weeks of age, ten males and ten females from each cohort were switched to the high-fat diet while the other animals continued on the chow diet. Plasma samples were collected at 8, 14, 18, and 22 weeks after a four-hour fast. Total cholesterol, HDL cholesterol, and triglycerides were measured using a Beckman Coulter Synchron CX®5 Delta autoanalyzer. Non-HDL was calculated as the difference between total cholesterol and HDL cholesterol.

At 22 weeks of age, animals were euthanized and their livers were snap-frozen. RNA was isolated from liver tissue using the MagMAX mirVana Total RNA Isolation Kit (ThermoFisher) and the KingFisher Flex purification system (ThermoFisher). Tissues were lysed and homogenized in TRIzol Reagent (ThermoFisher). After the addition of chloroform, the RNA-containing aqueous layer was removed for RNA isolation according to the manufacturer's protocol, beginning with the RNA bead binding step. RNA concentration and quality were assessed using the Nanodrop 2000 spectrophotometer (Thermo Scientific) and the RNA 6000 Nano LabChip assay (Agilent Technologies).

Libraries were prepared by the Genome Technologies core facility at The Jackson Laboratory using the KAPA Stranded mRNA-Seq Kit (KAPA Biosystems), according to the manufacturer's instructions. Briefly, the protocol entailed isolation of polyA containing mRNA using oligo-dT magnetic beads, RNA fragmentation, first and second-strand cDNA synthesis, ligation of Illumina-specific adapters containing a unique barcode sequence for each library, and polymerase chain reaction (PCR) amplification. Libraries were checked for quality and concentration using the DNA 1000 LabChip assay (Agilent Technologies) and quantitative PCR (KAPA Biosystems), according to the manufacturer's instructions. A total of 36 liver samples were collected (3 male and 3 female *Aldh2* knockout mice, 3 male and 3 female *Abca6* knockout mice, and 3 male and 3 female wildtype mice, for each of the two diets). A library was made for each sample followed by pooling of 6 libraries and samples were sequenced by the Genome Technologies core facility at The Jackson Laboratory, 125 bp paired-end on the HiSeq 2500 system (Illumina, Inc.; San Diego, CA) using the TruSeq SBS Kit v4 reagents (Illumina, Inc.) with a minimum of 40M reads per sample. A repeated measures two-way ANOVA test with post-hoc pairwise comparison was used to test for differentially expressed genes between wild type and knock-out mouse. A significant difference was determined at a false discovery rate <0.05 to account multiple testing.

Data availability

The genotype data, gene expression, phenotype data that support the findings from the FHS of this study have been deposited in dbGaP (dbGaP Study Accession: phs000363.v16.p10).

Supplementary Materials

Fig. S1. The functional analyses of differentially expressed genes generated through IPA.

Fig.S2. Interaction networks of molecules based on known relationships in the QIAGEN Knowledge

Table S1. Clinical Characteristics of the Framingham Heart Study Participants.

Table S2. SNPs from Lipids GWAS

Table S3. GWAS SNPs associated with gene expression(eQTLs) in FHS

Table S4. Mediation effect of SNPs on lipids through gene expression

Table S5. CVD proteins and their first-degree interactors from protein-protein interaction network

Table S6. Interactions between CVD proteins

Table S7. Differentially expressed genes from RNA-Seq after ALDH2 and ABCA6 knockout

Table S8. Overlap and molecular functions of differentially expressed genes between *Aldh2* and *Abca6* knockout mice

References

1. Pagidipati NJ and Gaziano TA. Estimating deaths from cardiovascular disease: a review of global methodologies of mortality measurement. *Circulation*. 2013;127:749-56.
2. Ross S, D'Mello M, Anand SS, Eikelboom J, Consortium CAD, Stewart AF, Samani NJ, Roberts R and Pare G. Effect of Bile Acid Sequestrants on the Risk of Cardiovascular Events: A Mendelian Randomization Analysis. *Circulation Cardiovascular genetics*. 2015;8:618-27.
3. Nikpay M, Goel A, Won HH, Hall LM, Willenborg C, Kanoni S, Saleheen D, Kyriakou T, Nelson CP, Hopewell JC, Webb TR, Zeng L, Dehghan A, Alver M, Armasu SM, Auro K, Bjorres A, Chasman DI, Chen S, Ford I, Franceschini N, Gieger C, Grace C, Gustafsson S, Huang J, Hwang SJ, Kim YK, Kleber ME, Lau KW, Lu X, Lu Y, Lyytikainen LP, Mihailov E, Morrison AC, Pervjakova N, Qu L, Rose LM, Salfati E, Saxena R, Scholz M, Smith AV, Tikkanen E, Uitterlinden A, Yang X, Zhang W, Zhao W, de Andrade M, de Vries PS, van Zuydam NR, Anand SS, Bertram L, Beutner F, Dedoussis G, Frossard P, Gauguier D, Goodall AH, Gottesman O, Haber M, Han BG, Huang J, Jalilzadeh S, Kessler T, Konig IR, Lannfelt L, Lieb W, Lind L, Lindgren CM, Lokki ML, Magnusson PK, Mallick NH, Mehra N, Meitinger T, Memon FU, Morris AP, Nieminen MS, Pedersen NL, Peters A, Rallidis LS, Rasheed A, Samuel M, Shah SH, Sinisalo J, Stirrups KE, Trompet S, Wang L, Zaman KS, Ardisino D, Boerwinkle E, Borecki IB, Bottinger EP, Buring JE, Chambers JC, Collins R, Cupples LA, Danesh J, Demuth I, Elosua R, Epstein SE, Esko T, Feitosa MF, Franco OH, Franzosi MG, Granger CB, Gu D, Gudnason V, Hall AS, Hamsten A, Harris TB, Hazen SL, Hengstenberg C, Hofman A, Ingelsson E, Iribarren C, Jukema JW, Karhunen PJ, Kim BJ, Kooner JS, Kullo IJ, Lehtimäki T, Loos RJF, Melander O, Metspalu A, Marz W, Palmer CN, Perola M, Quertermous T, Rader DJ, Ridker PM, Ripatti S, Roberts R, Salomaa V, Sanghera DK, Schwartz SM, Seedorf U, Stewart AF, Stott DJ, Thiery J, Zalloua PA, O'Donnell CJ, Reilly MP, Assimes TL, Thompson JR, Erdmann J, Clarke R, Watkins H, Kathiresan S, McPherson R, Deloukas P, Schunkert H, Samani NJ and Farrall M. A comprehensive 1,000

Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat Genet.* 2015;47:1121-1130.

4. Nelson CP, Hamby SE, Saleheen D, Hopewell JC, Zeng L, Assimes TL, Kanoni S, Willenborg C, Burgess S, Amouyel P, Anand S, Blankenberg S, Boehm BO, Clarke RJ, Collins R, Dedoussis G, Farrall M, Franks PW, Groop L, Hall AS, Hamsten A, Hengstenberg C, Hovingh GK, Ingelsson E, Kathiresan S, Kee F, Konig IR, Kooner J, Lehtimäki T, Marz W, McPherson R, Metspalu A, Nieminen MS, O'Donnell CJ, Palmer CN, Peters A, Perola M, Reilly MP, Ripatti S, Roberts R, Salomaa V, Shah SH, Schreiber S, Siegbahn A, Thorsteinsdottir U, Veronesi G, Wareham N, Willer CJ, Zalloua PA, Erdmann J, Deloukas P, Watkins H, Schunkert H, Danesh J, Thompson JR, Samani NJ and Consortium CACD. Genetically determined height and coronary artery disease. *N Engl J Med.* 2015;372:1608-18.

5. Do R, Willer CJ, Schmidt EM, Sengupta S, Gao C, Peloso GM, Gustafsson S, Kanoni S, Ganna A, Chen J, Buchkovich ML, Mora S, Beckmann JS, Bragg-Gresham JL, Chang HY, Demirkan A, Den Hertog HM, Donnelly LA, Ehret GB, Esko T, Feitosa MF, Ferreira T, Fischer K, Fontanillas P, Fraser RM, Freitag DF, Gurdasani D, Heikkilä K, Hyppönen E, Isaacs A, Jackson AU, Johansson A, Johnson T, Kaakinen M, Kettunen J, Kleber ME, Li X, Luan J, Lyytikäinen LP, Magnusson PK, Mangino M, Mihailov E, Montasser ME, Müller-Nurasyid M, Nolte IM, O'Connell JR, Palmer CD, Perola M, Petersen AK, Sanna S, Saxena R, Service SK, Shah S, Shungin D, Sidore C, Song C, Strawbridge RJ, Surakka I, Tanaka T, Teslovich TM, Thorleifsson G, Van den Herik EG, Voight BF, Volcik KA, Waite LL, Wong A, Wu Y, Zhang W, Absher D, Asiki G, Barroso I, Been LF, Bolton JL, Bonnycastle LL, Brambilla P, Burnett MS, Cesana G, Dimitriou M, Doney AS, Doring A, Elliott P, Epstein SE, Eyjolfsson GI, Gigante B, Goodarzi MO, Grallert H, Gravito ML, Groves CJ, Hallmans G, Hartikainen AL, Hayward C, Hernandez D, Hicks AA, Holm H, Hung YJ, Illig T, Jones MR, Kaleebu P, Kastelein JJ, Khaw KT, Kim E, Klopp N, Komulainen P, Kumari M, Langenberg C, Lehtimäki T, Lin SY, Lindstrom J, Loos RJ, Mach F, McArdle WL, Meisinger C, Mitchell BD, Müller G, Nagaraja R, Narisu N, Nieminen TV, Nsubuga RN, Olafsson I, Ong KK, Palotie A, Papamarkou T, Pomilla C, Pouta A, Rader DJ, Reilly MP, Ridker PM, Rivadeneira F, Rudan I, Ruokonen A, Samani N, Scharnagl H, Seeley J, Silander K, Stancakova A, Stirrups K, Swift AJ, Tirit L, Uitterlinden AG, van Pelt LJ, Vedantam S, Wainwright N, Wijmenga C, Wild SH, Willemsen G, Wilsgaard T, Wilson JF, Young EH, Zhao JH, Adair LS, Arveiler D, Assimes TL, Bandinelli S, Bennett F, Bochud M, Boehm BO, Boomsma DI, Borecki IB, Bornstein SR, Bovet P, Burnier M, Campbell H, Chakravarti A, Chambers JC, Chen YD, Collins FS, Cooper RS, Danesh J, Dedoussis G, de Faire U, Feranil AB, Ferrières J, Ferrucci L, Freimer NB, Gieger C, Groop LC, Gudnason V, Gyllenstein U, Hamsten A, Harris TB, Hingorani A, Hirschhorn JN, Hofman A, Hovingh GK, Hsiung CA, Humphries SE, Hunt SC, Hveem K, Iribarren C, Jarvelin MR, Jula A, Kahonen M, Kaprio J, Kesaniemi A, Kivimäki M, Kooner JS, Koudstaal PJ, Krauss RM, Kuh D, Kuusisto J, Kyvik KO, Laakso M, Lakka TA, Lind L, Lindgren CM, Martin NG, Marz W, McCarthy MI, McKenzie CA, Meneton P, Metspalu A, Moilanen L, Morris AD, Munroe PB, Njolstad I, Pedersen NL, Power C, Pramstaller PP, Price JF, Psaty BM, Quertermous T, Rauramaa R, Saleheen D, Salomaa V, Sanghera DK, Saramies J, Schwarz PE, Sheu WH, Shuldiner AR, Siegbahn A, Spector TD, Stefansson K, Strachan DP, Tayo BO, Tremoli E, Tuomilehto J, Uusitupa M, van Duijn CM, Vollenweider P, Wallentin L, Wareham NJ, Whitfield JB, Wolffenbuttel BH, Altshuler D, Ordovas JM, Boerwinkle E, Palmer CN, Thorsteinsdottir U, Chasman DI, Rotter JI, Franks PW, Ripatti S, Cupples LA, Sandhu MS, Rich SS, Boehnke M, Deloukas P, Mohlke KL, Ingelsson E, Abecasis GR, Daly MJ, Neale BM and Kathiresan S. Common variants associated with plasma triglycerides and risk for coronary artery disease. *Nature genetics.* 2013;45:1345-52.

6. Rockman MV and Kruglyak L. Genetics of global gene expression. *Nature reviews Genetics.* 2006;7:862-72.

7. Weile J, Sun S, Cote AG, Knapp J, Verby M, Mellor JC, Wu Y, Pons C, Wong C, van Lieshout N, Yang F, Tasan M, Tan G, Yang S, Fowler DM, Nussbaum R, Bloom JD, Vidal M, Hill DE, Aloy P and Roth FP. A framework for exhaustively mapping functional missense variants. *Mol Syst Biol.* 2017;13:957.

8. Menche J, Sharma A, Kitsak M, Ghiassian SD, Vidal M, Loscalzo J and Barabasi AL. Disease networks. Uncovering disease-disease relationships through the incomplete interactome. *Science*. 2015;347:1257601.
9. Sharma A, Halu A, Decano JL, Padi M, Liu YY, Prasad RB, Fadista J, Santolini M, Menche J, Weiss ST, Vidal M, Silverman EK, Aikawa M, Barabasi AL, Groop L and Loscalzo J. Controllability in an islet specific regulatory network identifies the transcriptional factor NFATC4, which regulates Type 2 Diabetes associated genes. *NPJ systems biology and applications*. 2018;4:25.
10. Yao C, Chen BH, Joehanes R, Otlu B, Zhang X, Liu C, Huan T, Tastan O, Cupples LA, Meigs JB, Fox CS, Freedman JE, Courchesne P, O'Donnell CJ, Munson PJ, Keles S and Levy D. Integromic analysis of genetic variation and gene expression identifies networks for cardiovascular disease phenotypes. *Circulation*. 2015;131:536-49.
11. Rolland T, Tasan M, Charlotiaux B, Pevzner SJ, Zhong Q, Sahni N, Yi S, Lemmens I, Fontanillo C, Mosca R, Kamburov A, Ghiassian SD, Yang X, Ghamsari L, Balcha D, Begg BE, Braun P, Brehme M, Broly MP, Carvunis AR, Convery-Zupan D, Corominas R, Coulombe-Huntington J, Dann E, Dreze M, Dricot A, Fan C, Franzosa E, Gebreab F, Gutierrez BJ, Hardy MF, Jin M, Kang S, Kiros R, Lin GN, Luck K, MacWilliams A, Menche J, Murray RR, Palagi A, Poulin MM, Rambout X, Rasla J, Reichert P, Romero V, Ruysinck E, Sahalie JM, Scholz A, Shah AA, Sharma A, Shen Y, Spirohn K, Tam S, Tejada AO, Trigg SA, Twizere JC, Vega K, Walsh J, Cusick ME, Xia Y, Barabasi AL, Iakoucheva LM, Aloy P, De Las Rivas J, Tavernier J, Calderwood MA, Hill DE, Hao T, Roth FP and Vidal M. A proteome-scale map of the human interactome network. *Cell*. 2014;159:1212-1226.
12. Dalal JJ, Padmanabhan TN, Jain P, Patil S, Vasawala H and Gulati A. LIPITENSION: Interplay between dyslipidemia and hypertension. *Indian J Endocrinol Metab*. 2012;16:240-5.
13. Joehanes R, Zhang X, Huan T, Yao C, Ying SX, Nguyen QT, Demirkale CY, Feolo ML, Sharopova NR, Sturcke A, Schaffer AA, Heard-Costa N, Chen H, Liu PC, Wang R, Woodhouse KA, Tanriverdi K, Freedman JE, Raghavachari N, Dupuis J, Johnson AD, O'Donnell CJ, Levy D and Munson PJ. Integrated genome-wide analysis of expression quantitative trait loci aids interpretation of genomic association studies. *Genome biology*. 2017;18:16.
14. Luck K, Kim DK, Lambourne L, Spirohn K, Begg BE, Bian W, Brignall R, Cafarelli T, Campos-Laborie FJ, Charlotiaux B, Choi D, Cote AG, Daley M, Deimling S, Desbuleux A, Dricot A, Gebbia M, Hardy MF, Kishore N, Knapp JJ, Kovacs IA, Lemmens I, Mee MW, Mellor JC, Pollis C, Pons C, Richardson AD, Schlabach S, Teeking B, Yadav A, Babor M, Balcha D, Basha O, Bowman-Colin C, Chin SF, Choi SG, Colabella C, Coppin G, D'Amata C, De Ridder D, De Rouck S, Duran-Frigola M, Ennajaoui H, Goebels F, Goehring L, Gopal A, Haddad G, Hatchi E, Helmy M, Jacob Y, Kassa Y, Landini S, Li R, van Lieshout N, MacWilliams A, Markey D, Paulson JN, Rangarajan S, Rasla J, Rayhan A, Rolland T, San-Miguel A, Shen Y, Sheykhkarimli D, Sheynkman GM, Simonovsky E, Tasan M, Tejada A, Tropepe V, Twizere JC, Wang Y, Weatheritt RJ, Weile J, Xia Y, Yang X, Yeger-Lotem E, Zhong Q, Aloy P, Bader GD, De Las Rivas J, Gaudet S, Hao T, Rak J, Tavernier J, Hill DE, Vidal M, Roth FP and Calderwood MA. A reference map of the human binary protein interactome. *Nature*. 2020;580:402-408.
15. Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, Ganna A, Chen J, Buchkovich ML, Mora S, Beckmann JS, Bragg-Gresham JL, Chang HY, Demirkan A, Den Hertog HM, Do R, Donnelly LA, Ehret GB, Esko T, Feitosa MF, Ferreira T, Fischer K, Fontanillas P, Fraser RM, Freitag DF, Gurdasani D, Heikkila K, Hypponen E, Isaacs A, Jackson AU, Johansson A, Johnson T, Kaakinen M, Kettunen J, Kleber ME, Li X, Luan J, Lyytikainen LP, Magnusson PKE, Mangino M, Mihailov E, Montasser ME, Muller-Nurasyid M, Nolte IM, O'Connell JR, Palmer CD, Perola M, Petersen AK, Sanna S, Saxena R, Service SK, Shah S, Shungin D, Sidore C, Song C, Strawbridge RJ, Surakka I, Tanaka T, Teslovich TM, Thorleifsson G, Van den Herik EG, Voight BF, Volcik KA, Waite LL, Wong A, Wu Y, Zhang W, Absher D, Asiki G, Barroso I, Been LF, Bolton JL, Bonnycastle LL, Brambilla P, Burnett MS, Cesana G, Dimitriou M, Doney ASF, Doring A, Elliott P, Epstein SE, Ingi Eyjolfsson G, Gigante B, Goodarzi MO, Grallert H, Gravit

ML, Groves CJ, Hallmans G, Hartikainen AL, Hayward C, Hernandez D, Hicks AA, Holm H, Hung YJ, Illig T, Jones MR, Kaleebu P, Kastelein JJP, Khaw KT, Kim E, Klopp N, Komulainen P, Kumari M, Langenberg C, Lehtimäki T, Lin SY, Lindstrom J, Loos RJF, Mach F, McArdle WL, Meisinger C, Mitchell BD, Muller G, Nagaraja R, Narisu N, Nieminen TVM, Nsubuga RN, Olafsson I, Ong KK, Palotie A, Papamarkou T, Pomilla C, Pouta A, Rader DJ, Reilly MP, Ridker PM, Rivadeneira F, Rudan I, Ruokonen A, Samani N, Scharnagl H, Seeley J, Silander K, Stancakova A, Stirrups K, Swift AJ, Tirit L, Uitterlinden AG, van Pelt LJ, Vedantam S, Wainwright N, Wijmenga C, Wild SH, Willemssen G, Wilsgaard T, Wilson JF, Young EH, Zhao JH, Adair LS, Arveiler D, Assimes TL, Bandinelli S, Bennett F, Bochud M, Boehm BO, Boomsma DI, Borecki IB, Bornstein SR, Bovet P, Burnier M, Campbell H, Chakravarti A, Chambers JC, Chen YI, Collins FS, Cooper RS, Danesh J, Dedoussis G, de Faire U, Feranil AB, Ferrieres J, Ferrucci L, Freimer NB, Gieger C, Groop LC, Gudnason V, Gyllenstein U, Hamsten A, Harris TB, Hingorani A, Hirschhorn JN, Hofman A, Hovingh GK, Hsiung CA, Humphries SE, Hunt SC, Hveem K, Iribarren C, Jarvelin MR, Jula A, Kahonen M, Kaprio J, Kesaniemi A, Kivimäki M, Kooner JS, Koudstaal PJ, Krauss RM, Kuh D, Kuusisto J, Kyvik KO, Laakso M, Lakka TA, Lind L, Lindgren CM, Martin NG, Marz W, McCarthy MI, McKenzie CA, Meneton P, Metspalu A, Moilanen L, Morris AD, Munroe PB, Njolstad I, Pedersen NL, Power C, Pramstaller PP, Price JF, Psaty BM, Quertermous T, Rauramaa R, Saleheen D, Salomaa V, Sanghera DK, Saramies J, Schwarz PEH, Sheu WH, Shuldiner AR, Siegbahn A, Spector TD, Stefansson K, Strachan DP, Tayo BO, Tremoli E, Tuomilehto J, Uusitupa M, van Duijn CM, Vollenweider P, Wallentin L, Wareham NJ, Whitfield JB, Wolfenbutter BHR, Ordovas JM, Boerwinkle E, Palmer CNA, Thorsteinsdottir U, Chasman DI, Rotter JI, Franks PW, Ripatti S, Cupples LA, Sandhu MS, Rich SS, Boehnke M, Deloukas P, Kathiresan S, Mohlke KL, Ingelsson E, Abecasis GR and Global Lipids Genetics C. Discovery and refinement of loci associated with lipid levels. *Nature genetics*. 2013;45:1274-1283.

16. Chen X, Gu X and Zhang H. Sidt2 regulates hepatocellular lipid metabolism through autophagy. *J Lipid Res*. 2018;59:404-415.

17. Meng Y, Wang L and Ling L. Changes of lysosomal membrane permeabilization and lipid metabolism in sidt2 deficient mice. *Exp Ther Med*. 2018;16:246-252.

18. Dean M and Annilo T. Evolution of the ATP-binding cassette (ABC) transporter superfamily in vertebrates. *Annu Rev Genomics Hum Genet*. 2005;6:123-42.

19. Kaminski WE, Wenzel JJ, Piehler A, Langmann T and Schmitz G. ABCA6, a novel a subclass ABC transporter. *Biochemical and biophysical research communications*. 2001;285:1295-301.

20. Pohl A, Devaux PF and Herrmann A. Function of prokaryotic and eukaryotic ABC proteins in lipid transport. *Biochimica et biophysica acta*. 2005;1733:29-52.

21. Wada M, Daimon M, Emi M, Iijima H, Sato H, Koyano S, Tajima K, Kawanami T, Kurita K, Hunt SC, Hopkins PN, Kubota I, Kawata S and Kato T. Genetic association between aldehyde dehydrogenase 2 (ALDH2) variation and high-density lipoprotein cholesterol (HDL-C) among non-drinkers in two large population samples in Japan. *J Atheroscler Thromb*. 2008;15:179-84.

22. Gao J, Zhang Y, Yu C, Tan F and Wang L. Spontaneous nonalcoholic fatty liver disease and ER stress in Sidt2 deficiency mice. *Biochemical and biophysical research communications*. 2016;476:326-332.

23. Mendez-Acevedo KM, Valdes VJ, Asanov A and Vaca L. A novel family of mammalian transmembrane proteins involved in cholesterol transport. *Scientific reports*. 2017;7:7450.

24. Kramer A, Green J, Pollard J, Jr. and Tugendreich S. Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics*. 2014;30:523-30.

25. Ongen H, Brown AA, Delaneau O, Panousis NI, Nica AC, Consortium GT and Dermitzakis ET. Estimating the causal tissues for complex traits and diseases. *Nature genetics*. 2017;49:1676-1683.

26. Musunuru K, Strong A, Frank-Kamenetsky M, Lee NE, Ahfeldt T, Sachs KV, Li X, Li H, Kuperwasser N, Ruda VM, Pirruccello JP, Muchmore B, Prokunina-Olsson L, Hall JL, Schadt EE, Morales CR, Lund-Katz S, Phillips MC, Wong J, Cantley W, Racie T, Ejebe KG, Orho-Melander M, Melander O, Kotliansky V,

- Fitzgerald K, Krauss RM, Cowan CA, Kathiresan S and Rader DJ. From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. *Nature*. 2010;466:714-9.
27. Atanasovska B, Kumar V, Fu J, Wijmenga C and Hofker MH. GWAS as a Driver of Gene Discovery in Cardiometabolic Diseases. *Trends in endocrinology and metabolism: TEM*. 2015;26:722-732.
28. Jones NS, Watson KQ and Rebeck GW. Metabolic Disturbances of a High-Fat Diet Are Dependent on APOE Genotype and Sex. *eNeuro*. 2019;6.
29. Yadav A, Vidal M and Luck K. Precision medicine - networks to the rescue. *Curr Opin Biotechnol*. 2020;63:177-189.
30. Oppenheimer GM. Becoming the Framingham Study 1947-1950. *Am J Public Health*. 2005;95:602-10.
31. Dawber TR, Meadors GF and Moore FE, Jr. Epidemiological approaches to heart disease: the Framingham Study. *American journal of public health and the nation's health*. 1951;41:279-81.
32. Feinleib M, Kannel WB, Garrison RJ, McNamara PM and Castelli WP. The Framingham Offspring Study. Design and preliminary data. *Preventive medicine*. 1975;4:518-25.
33. Splansky GL, Corey D, Yang Q, Atwood LD, Cupples LA, Benjamin EJ, D'Agostino RB, Sr., Fox CS, Larson MG, Murabito JM, O'Donnell CJ, Vasan RS, Wolf PA and Levy D. The Third Generation Cohort of the National Heart, Lung, and Blood Institute's Framingham Heart Study: design, recruitment, and initial examination. *Am J Epidemiol*. 2007;165:1328-35.
34. Benjamin AM, Suchindran S, Pearce K, Rowell J, Lien LF, Guyton JR and McCarthy JJ. Gene by sex interaction for measures of obesity in the framingham heart study. *J Obes*. 2011;2011:329038.
35. Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B and Speed TP. Summaries of Affymetrix GeneChip probe level data. *Nucleic acids research*. 2003;31:e15.
36. Stegle O, Parts L, Piipari M, Winn J and Durbin R. Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. *Nat Protoc*. 2012;7:500-7.
37. Imai K, Keele L and Tingley D. A general approach to causal mediation analysis. *Psychological methods*. 2010;15:309-34.
38. Hartwig FP, Davies NM, Hemani G and Davey Smith G. Two-sample Mendelian randomization: avoiding the downsides of a powerful, widely applicable but potentially fallible technique. *International journal of epidemiology*. 2017.
39. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, Laurin C, Burgess S, Bowden J, Langdon R, Tan VY, Yarmolinsky J, Shihab HA, Timpson NJ, Evans DM, Relton C, Martin RM, Davey Smith G, Gaunt TR and Haycock PC. The MR-Base platform supports systematic causal inference across the human phenome. *Elife*. 2018;7.
40. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, Jensen LJ and Mering CV. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*. 2019;47:D607-D613.

Acknowledgments: We gratefully acknowledge the contribution of the Reproductive Science, Histopathology, and Genome Technologies Services at The Jackson Laboratory for expert assistance with the work described in this publication. **Funding:** The Framingham Heart Study is

funded by National Institutes of Health contract N01-HC-25195. The analytical component for this investigation was funded by the Division of Intramural Research, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD (D. Levy, Principal Investigator). The laboratory work of this project was funded by American Heart Association (AHA) Cardiovascular Genome-Phenome Study (CVGPS) grant 15CVGPS23430000. **Author contributions:** Conceptualization, C.Y. and D.L. Methodology and data analysis, C.Y. Mouse knockout experiments, H.S., Y.T., R.K. Protein interaction network, T.H., W.B., D.E.H, M.V. Writing, C.Y., R.K., D.L. **Competing interests:** No competing interests to declare for all authors. **Data and materials availability:** The SNP, gene expression and protein phenotype data that Support the findings from the FHS of this study have been deposited in dbGaP (dbGaP Study Accession: phs000363.v16.p10).

Figures:

Figure 1. Study Design

Schematic diagram of stepwise approach to identify target genes by utilizing human lipid GWAS datasets, gene expression, protein interaction network and phenotype data, cross-reference with mouse knockout database, prioritize genes, and validate candidate genes.

Figure 2. Comparison between *Abca6* knockout (solid line) and wildtype (dotted line) mice in females and males on regular chow and high-fat diet. Ten mice for each genotype and sex

per diet group. Asterisk indicates a significant difference ($P < 0.05$) between knockout and wildtype at the specific time point.

Female

693 Male

694

695

696

697

698

Figure 3. Comparison between *Aldh2* knockout (solid line) and wildtype (dotted line) mice in females and males on regular chow and high-fat diet. Ten mice for each genotype and sex per diet group. Asterisk indicates a significant difference ($P < 0.05$) between knockout and wildtype at the specific time point.

Female

721 Male

722

723

724

725

726

727

728

729

Figure 4. Comparison between *Sid2* knockout (solid line) and wildtype (dotted line) mice in males and females on regular chow and high-fat diet. Ten mice for each genotype and sex per diet group. Asterisk indicates a significant difference ($P < 0.05$) between knockout and wildtype at the specific time point.

Female

736 Male

737

738

739

740

741

742

743 **Table 1 . MR tests for candidate genes**

Exposure	Outcome	Instrumental	Beta	SE	P Value
<i>DOCK7</i>	HDL cholesterol	rs10889354	0.22	0.0821	0.0067
	LDL cholesterol	rs10889354	0.75	0.086	2.22E-18
	Triglycerides	rs10889354	1.15	0.078	1.56E-48
<i>TAGLN</i>	HDL cholesterol	rs3736120	-0.18	0.067	0.0091
	LDL cholesterol	rs3736120	-0.14	0.074	0.064
	Triglycerides	rs3736120	-0.34	0.066	2.90E-07
<i>SIDT2</i>	HDL cholesterol	rs12420127	0.27	0.11	0.013
	LDL cholesterol	rs12420127	0.32	0.12	0.0065
	Triglycerides	rs12420127	0.32	0.11	0.0027
<i>ALDH2</i>	HDL cholesterol	rs10744777	0.20	0.09	0.017
	LDL cholesterol	rs10744777	0.49	0.093	1.28E-07
	Triglycerides	rs10744777	-0.23	0.083	0.0056
<i>SLC44A4</i>	HDL cholesterol	rs535586	0.079	0.055	0.15
	LDL cholesterol	rs535586	-0.23	0.059	0.00011
	Triglycerides	rs535586	-0.38	0.054	2.56E-12
<i>ABCA6</i>	HDL cholesterol	rs918167	0.10	0.093	0.27
	LDL cholesterol	rs918167	0.061	0.10	0.55
	Triglycerides	rs918167	-0.16	0.091	0.079
<i>ATG4C</i>	HDL cholesterol	rs7540030	0.14	0.19	0.46
	LDL cholesterol	rs7540030	0.17	0.20	0.40
	Triglycerides	rs7540030	0.10	0.18	0.56

Table 2. Criteria of candidate gene selection

	Causal effects (MR P<0.05)	Novelty	Availability of knockout mice	Associated with CVD- proteins in PPI network
DOCK7	3 lipids traits	No	Yes	Yes
TAGLN	2 lipids traits	No	Yes	Yes
SIDT2	2 lipids traits	No	Yes	Yes
ALDH2	3 lipids traits	Yes	Yes	Yes
SLC44A4	2 lipids traits	Yes	No	No
ABCA6	0 lipids traits	Yes	Yes	Yes
ATG4C	0 lipids traits	Yes	Yes	No