

1 **Functional Characterization of Lipid Regulatory Effects of Three Genes**

2 **Using Knockout Mouse Models**

3

4

5 Chen Yao^{1,2}, Holly Savage³, Tong Hao^{4,5}, Gha Young Lee^{1,2}, Yuka Takemon³, Wenting Bian^{4,5},

6 David E Hill^{4,5}, Marc Vidal^{4,5}, Ron Korstanje^{3*} and Daniel Levy^{1,2*}

7

8 ¹The Framingham Heart Study, Framingham, MA;

9 ²The Population Sciences Branch, Division of Intramural Research, National Heart, Lung, and

10 Blood Institute, Bethesda, MD;

11 ³The Jackson Laboratory, Bar Harbor, ME;

12 ⁴Center for Cancer Systems Biology (CCSB) and Dept. of Cancer Biology, Dana-Farber Cancer

13 Institute, Boston MA, USA; ⁵Dept. of Genetics, Blavatnik Institute, Harvard Medical School,

14 Boston MA

15

16 ***Corresponding author:** Daniel Levy, MD

17 Email: levyd@nih.gov

25 **Abstract:**

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

39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

61 Introduction

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

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

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

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

97 Results

98 Identifying lipid trait-related expression quantitative loci

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

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

108 **Expanding the candidate gene network**

109 To expand the candidate gene network, we built a CVD gene network by incorporating all protein
110 coding genes containing CVD GWAS SNPs. Our CVD-centric network contained binary PPIs for
111 846 CVD-associated proteins (out of ~1,300 tested) extracted from a dataset of ~58,000 binary
112 PPIs among 10,690 human proteins obtained from both systematic binary mapping and literature
113 curations¹⁴. The 846 CVD-associated proteins and their first-degree neighbors constituted a
114 network of ~8,600 interactions among 4,336 proteins (Supplementary Table 5), including 444
115 intra-CVD PPIs among 349 lipids-associated proteins (Supplementary Table 6). This expanded
116 network is amenable to investigating disease models by analyzing other features such as shared
117 gene ontology (GO) terms, shared expression profiles, and functional similarity.

118 **Ranking candidate genes**

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

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

138 *ABCA6* Knockout

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

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

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

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

155 *ALDH2* Knockout

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

161 *ALDH2* was causally related to all three lipids in MR (Table 1). In addition, SNPs in *ALDH2*
162 were significantly associated with HDL and LDL in GWAS and gene expression in FHS
163 (Supplementary Table 3). Comparing HDL cholesterol levels between B6Dnk;B6N-
164 *Aldh2*^{tm1a(EUCOMM)Wtsi/JegRkorJ} mice (*Aldh2* knockout) and WT C57BL/6N mice, we found
165 significantly increased HDL cholesterol levels in both female and male *Aldh2* KO mice at several
166 time points compared to controls when fed the chow diet (Figure 3). On a high-fat diet, we no
167 longer observed a difference in males, but there was a significant difference in females at 18 weeks
168 of age.

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

174 *SIDT2* Knockout

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

183 *SIDT2* was significant in MR test for all three lipids traits. SNPs in *SIDT2* were significantly
184 associated with LDL in GWAS and gene expression in FHS (Supplementary Table 3). We
185 identified *SIDT2* in our human cohort to be associated primarily with triglycerides and therefore
186 compared TG levels between B6;129S5-*Sidt2*^{tm1Lex}/MmucdRkorJ mice (*Sidt2* knockout) and WT
187 control littermates. We observed a significant increase in TG levels in the female mice on the chow
188 diet at 18 and 22 weeks of age (Figure 4). Male TG levels were only increased at 8 weeks of age.

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

195 **Expression changes after gene knock out**

196 To explore transcriptome-wide alterations resulting from knocking out the target genes, we
197 conducted RNA-Seq for *ALDH2* and *ABCA6* in three male and three female knockout animals and
198 three male and three female control littermates for each of the two diets. We didn't conduct the
199 RNA-Seq for *SIDT2* in light of previously reported studies of this knockout^{16,17}. At FDR<0.05 (T
200 test, P<0.014), we identified an average of 238 differentially-expressed genes (ranging from 69 to
201 403 for eight comparisons) after knockout (Supplementary Table 7). The differentially-expressed
202 genes were analyzed through the use of Ingenuity Pathway Analysis (IPA, QIAGEN Inc.)²⁴.
203 Function enrichment analyses were conducted on the differentially expressed genes using curated
204 information from the QIAGEN Knowledge Base²⁴. Though canonical pathways of differentially
205 expressed genes from mice of different diet conditions are enriched in different metabolic and cell
206 signaling pathways, the affected diseases and molecular functions are highly consistent with regard
207 to metabolic disease and lipid metabolism (Supplementary figure 1). The top toxicological
208 functions were liver steatosis and cardiac dysfunction. Network analysis based on Ingenuity
209 Knowledge Base²⁴ revealed that *Abca6* affected cholesterol level via cytokine IFNG (interferon
210 gamma) and that *Aldh2* affected cholesterol level through enzyme GNMT (glycine N-
211 methyltransferase) and acetaldehyde (Supplementary Figure 2). Comparing RNA-seq data from
212 *Abca6* and *Aldh2* knockout of the same sex and diet, we found 19%-60% overlap of differentially
213 expressed genes. Using IPA functional enrichment at P<0.005, we found that the overlapping
214 genes are enriched in lipid metabolism (Supplementary Table 8), suggesting a common pathway
215 affected by these two very different genes.

216 **Discussion**

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

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

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

233 LDL cholesterol and total cholesterol, resulted in higher total cholesterol (which is largely LDL
234 cholesterol) and HDL cholesterol in male mice, suggesting a potential role of this gene in CVD
235 development. *ABCA6* plays a key role in macrophage lipid homeostasis by facilitating
236 phospholipid and cholesterol export from the cell^{19, 20}. *ABCA6* dysfunction may cause a buildup
237 of intracellular phospholipids and cholesterol. Knockout of *Aldh2*, which was chosen for its
238 putatively causal association with HDL cholesterol, resulted in significantly increased HDL
239 cholesterol levels in both female and male mice. Aldehyde dehydrogenase, the protein encoded by
240 *ALDH2*, is an enzyme in the oxidative pathway and plays a major role the metabolism of lipid
241 peroxidation-derived aldehydes²¹. In our knockout models, the absence of *Aldh2* resulted in higher
242 lipid levels, which supports the idea that *ALDH2* is involved in lipid degradation/clearance.
243 Knockout of *Sidt2*, which contains GWAS SNPs for HDL cholesterol, resulted in a significant
244 increase in triglyceride levels in the females on the chow diet at 18 and 22 weeks of age. *SIDT2*
245 plays a crucial role in the uptake and intracellular transport of triglycerides and cholesterol^{22, 23}.
246 Previous studies¹⁶ have shown that the absence of *Sidt2* causes increased serum triglycerides and
247 free fatty acids in mice. After analyzing the knockout results, we compared them with expression
248 data for the corresponding human genes. We discovered that while each gene knockout produced
249 some significantly different results in the female and male mice, FHS human gene expression data
250 did not reveal sex differences in lipid effects for any of the genes ($P > 0.05$, data not shown). This
251 may be due to innate differences in lipid metabolism and regulation between the two species.

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

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

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

278 dysfunction of novel CVD-associate genes is manifested at the molecular level ²⁹. Approaches
279 similar to those used in this study may not only be expanded upon in relation to CVD but also be
280 used to investigate other diseases impacted by genetic and epigenetic factors. To make future
281 studies more comprehensive and accurate, it will be important to expand eQTL databases to other
282 disease-related tissues and to include additional clinically important traits as well.

283 **Materials and Methods**

284 **Samples and Phenotypes**

285 In 1948, the FHS started recruiting participants (original cohort) from Framingham, MA to begin
286 the first round of extensive physical examinations and lifestyle surveys to investigate CVD and its
287 risk factors ³⁰. In 1971 and 2002, the FHS recruited offspring (and their spouses) and adult
288 grandchildren of the original cohort participants into the offspring and third-generation cohorts,
289 respectively ³¹⁻³³. Plasma total cholesterol, HDL cholesterol, triglycerides, and glucose were
290 measured in the morning after an eight-hour overnight fasting. Body mass index (BMI) was
291 defined as weight (kilograms) divided by height squared (meters²). Current smokers were defined
292 as those who smoked, on average, at least one cigarette per day during the year prior to the FHS
293 clinical assessment. Clinical characteristics of the study sample are summarized in Supplementary
294 Table 1. All participants from the FHS gave informed consent for participation in this study and
295 for the collection of plasma and DNA for analysis. The FHS study protocol was approved by
296 Boston Medical Center: Protocol ID: H-27984, Boston University Medical Center IRB (BUMC
297 IRB) Title: FRAMINGHAM HEART STUDY BIOMARKER PROJECT.

298 Genotype data: 5,568 SNPs that were associated with the three lipid traits – HDL cholesterol, LDL
299 cholesterol, and triglycerides – at $p \leq 5 \times 10^{-8}$ (in the GRASP database, downloaded in June 2016,
300 Supplementary Table 2) were curated and matched with the FHS 1000 Genomes Project imputed
301 genotype data ³⁴. SNPs with imputed quality score (r^2) < 0.3 and minor allele frequency (MAF)
302 < 0.01 were excluded, resulting in 4173 genome-wide significant SNPs for eQTL analysis.

303 Gene expression: Whole blood was collected in PAXgene™ tubes (PreAnalytiX, Hombrechtikon,
304 Switzerland) and frozen at -80°C . RNA was extracted using a whole blood RNA System Kit
305 (Qiagen, Venlo, Netherlands) and mRNA expression profiling was assessed using the Affymetrix
306 Human Exon 1.0 ST GeneChip platform (Affymetrix Inc, Santa Clara, CA), which contains more
307 than 5.5 million probes targeting the expression of 17,873 genes. The Robust Multi-array Average
308 (RMA) package ³⁵ was used to normalize the gene expression values and remove any technical or
309 spurious background variation. Linear regression models were used to adjust for technical
310 covariates (batch, first principal component, and residual mean of all probesets).

311 **Identification of LIPIDS-associated eQTLs**

312 The eQTLs were identified from FHS genotype data and whole blood gene expression as described
313 previously ¹³. eQTL analyses were conducted in two phases: 1) gene expression residuals were
314 generated after accounting for the effects of sex, age, platelet count, white blood cell count, and
315 imputed differential blood cell counts; analyses were performed using R version 3.0.1 with a
316 mixed-effect modeling package that adjusted for familial relationships; and 2) underlying
317 confounding factors were accounted for by 20 Probabilistic Estimation of Expression Residuals
318 (PEERs)³⁶ factors that were computed using the residualized expression data. The residualized
319 expression was fit to a linear model using the PEER factors along with sex, age, and effect allele
320 dosages. The algorithm was implemented with Graphical Processing Units (GPUs). *cis*-eQTLs

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

324 Putative causality was first tested by mediation tests, in which the underlying mechanism for a
325 relationship between two variables is tested by introducing a third explanatory variable. The
326 analysis was conducted with the Mediation Package³⁷ in R where the “exposure” variable
327 represented a SNP, the “mediator” or “explanatory variable” represented gene expression, and the
328 “outcome” variable represented the phenotype. The mediation effect was measured on a scale from
329 0-100% where a 100% mediation effect indicated that the entire relationship between a SNP and
330 a phenotype (direct effect) was explained by changes in gene expression, i.e., the “mediator” or
331 the “explanatory variable.” Significant mediation effects were selected at a permutation p-value
332 <0.005 (based on 1000 permutations).

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

343 **Integration with PPI Networks**

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

353 **Identifying Candidate Genes**

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

360 We then ranked the genes based on how many of the criteria were met and further excluded any
361 genes if the knockout was known to be lethal based on existing literature. Three genes *ABCA6*,
362 *ALDH2*, and *SIDT2* met more than three of these selection criteria and were selected for further
363 investigation.

364 Validation of Candidate Genes in Animal Models

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

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

381 At 22 weeks of age, animals were euthanized and their livers were snap-frozen. RNA was
382 isolated from liver tissue using the MagMAX mirVana Total RNA Isolation Kit (ThermoFisher)
383 and the KingFisher Flex purification system (ThermoFisher). Tissues were lysed and homogenized
384 in TRIzol Reagent (ThermoFisher). After the addition of chloroform, the RNA-containing aqueous
385 layer was removed for RNA isolation according to the manufacturer's protocol, beginning with
386 the RNA bead binding step. RNA concentration and quality were assessed using the Nanodrop
387 2000 spectrophotometer (Thermo Scientific) and the RNA 6000 Nano LabChip assay (Agilent
388 Technologies).

389 Libraries were prepared by the Genome Technologies core facility at The Jackson Laboratory
390 using the KAPA Stranded mRNA-Seq Kit (KAPA Biosystems), according to the manufacturer's
391 instructions. Briefly, the protocol entailed isolation of polyA containing mRNA using oligo-dT
392 magnetic beads, RNA fragmentation, first and second-strand cDNA synthesis, ligation of Illumina-
393 specific adapters containing a unique barcode sequence for each library, and polymerase chain
394 reaction (PCR) amplification. Libraries were checked for quality and concentration using the DNA
395 1000 LabChip assay (Agilent Technologies) and quantitative PCR (KAPA Biosystems), according
396 to the manufacturer's instructions. A total of 36 liver samples were collected (3 male and 3 female
397 *Aldh2* knockout mice, 3 male and 3 female *Abca6* knockout mice, and 3 male and 3 female
398 wildtype mice, for each of the two diets). A library was made for each sample followed by pooling
399 of 6 libraries and samples were sequenced by the Genome Technologies core facility at The
400 Jackson Laboratory, 125 bp paired-end on the HiSeq 2500 system (Illumina, Inc.; San Diego, CA)
401 using the TruSeq SBS Kit v4 reagents (Illumina, Inc.) with a minimum of 40M reads per sample.
402 A repeated measures two-way ANOVA test with post-hoc pairwise comparison was used to test
403 for differentially expressed genes between wild type and knock-out mouse. A significant
404 difference was determined at a false discovery rate <0.05 to account multiple testing.

405 Data availability

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

408

409

410 **Supplementary Materials**

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

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

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

415 Table S2. SNPs from Lipids GWAS

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

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

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

420 Table S6. Interactions between CVD proteins

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

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

424

425 **References**

426

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

- 449 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat Genet.*
450 2015;47:1121-1130.
- 451 4. Nelson CP, Hamby SE, Saleheen D, Hopewell JC, Zeng L, Assimes TL, Kanoni S, Willenborg C,
452 Burgess S, Amouyel P, Anand S, Blankenberg S, Boehm BO, Clarke RJ, Collins R, Dedoussis G, Farrall M,
453 Franks PW, Groop L, Hall AS, Hamsten A, Hengstenberg C, Hovingh GK, Ingelsson E, Kathiresan S, Kee F,
454 Konig IR, Kooner J, Lehtimaki T, Marz W, McPherson R, Metspalu A, Nieminen MS, O'Donnell CJ, Palmer
455 CN, Peters A, Perola M, Reilly MP, Ripatti S, Roberts R, Salomaa V, Shah SH, Schreiber S, Siegbahn A,
456 Thorsteinsdottir U, Veronesi G, Wareham N, Willer CJ, Zalloua PA, Erdmann J, Deloukas P, Watkins H,
457 Schunkert H, Danesh J, Thompson JR, Samani NJ and Consortium CACD. Genetically determined height
458 and coronary artery disease. *N Engl J Med.* 2015;372:1608-18.
- 459 5. Do R, Willer CJ, Schmidt EM, Sengupta S, Gao C, Peloso GM, Gustafsson S, Kanoni S, Ganna A,
460 Chen J, Buchkovich ML, Mora S, Beckmann JS, Bragg-Gresham JL, Chang HY, Demirkan A, Den Hertog
461 HM, Donnelly LA, Ehret GB, Esko T, Feitosa MF, Ferreira T, Fischer K, Fontanillas P, Fraser RM, Freitag DF,
462 Gurdasani D, Heikkila K, Hypponen E, Isaacs A, Jackson AU, Johansson A, Johnson T, Kaakinen M,
463 Kettunen J, Kleber ME, Li X, Luan J, Lytikainen LP, Magnusson PK, Mangino M, Mihailov E, Montasser
464 ME, Muller-Nurasyid M, Nolte IM, O'Connell JR, Palmer CD, Perola M, Petersen AK, Sanna S, Saxena R,
465 Service SK, Shah S, Shungin D, Sidore C, Song C, Strawbridge RJ, Surakka I, Tanaka T, Teslovich TM,
466 Thorleifsson G, Van den Herik EG, Voight BF, Volcik KA, Waite LL, Wong A, Wu Y, Zhang W, Absher D,
467 Asiki G, Barroso I, Been LF, Bolton JL, Bonnycastle LL, Brambilla P, Burnett MS, Cesana G, Dimitriou M,
468 Doney AS, Doring A, Elliott P, Epstein SE, Eyjolfsson GI, Gigante B, Goodarzi MO, Grallert H, Gravito ML,
469 Groves CJ, Hallmans G, Hartikainen AL, Hayward C, Hernandez D, Hicks AA, Holm H, Hung YJ, Illig T, Jones
470 MR, Kaleebu P, Kastelein JJ, Khaw KT, Kim E, Klopp N, Komulainen P, Kumari M, Langenberg C, Lehtimaki
471 T, Lin SY, Lindstrom J, Loos RJ, Mach F, McArdle WL, Meisinger C, Mitchell BD, Muller G, Nagaraja R,
472 Narisu N, Nieminen TV, Nsubuga RN, Olafsson I, Ong KK, Palotie A, Papamarkou T, Pomilla C, Pouta A,
473 Rader DJ, Reilly MP, Ridker PM, Rivadeneira F, Rudan I, Ruukonen A, Samani N, Scharnagl H, Seeley J,
474 Silander K, Stancakova A, Stirrups K, Swift AJ, Tirit L, Uitterlinden AG, van Pelt LJ, Vedantam S,
475 Wainwright N, Wijmenga C, Wild SH, Willemsen G, Wilsgaard T, Wilson JF, Young EH, Zhao JH, Adair LS,
476 Arveiler D, Assimes TL, Bandinelli S, Bennett F, Bochud M, Boehm BO, Boomsma DI, Borecki IB, Bornstein
477 SR, Bovet P, Burnier M, Campbell H, Chakravarti A, Chambers JC, Chen YD, Collins FS, Cooper RS, Danesh
478 J, Dedoussis G, de Faire U, Feranil AB, Ferrieres J, Ferrucci L, Freimer NB, Gieger C, Groop LC, Gudnason
479 V, Gyllensten U, Hamsten A, Harris TB, Hingorani A, Hirschhorn JN, Hofman A, Hovingh GK, Hsiung CA,
480 Humphries SE, Hunt SC, Hveem K, Iribarren C, Jarvelin MR, Jula A, Kahonen M, Kaprio J, Kesaniemi A,
481 Kivimaki M, Kooner JS, Koudstaal PJ, Krauss RM, Kuh D, Kuusisto J, Kyvik KO, Laakso M, Lakka TA, Lind L,
482 Lindgren CM, Martin NG, Marz W, McCarthy MI, McKenzie CA, Meneton P, Metspalu A, Moilanen L,
483 Morris AD, Munroe PB, Njolstad I, Pedersen NL, Power C, Pramstaller PP, Price JF, Psaty BM,
484 Quertermous T, Rauramaa R, Saleheen D, Salomaa V, Sanghera DK, Saramies J, Schwarz PE, Sheu WH,
485 Shuldiner AR, Siegbahn A, Spector TD, Stefansson K, Strachan DP, Tayo BO, Tremoli E, Tuomilehto J,
486 Uusitupa M, van Duijn CM, Vollenweider P, Wallentin L, Wareham NJ, Whitfield JB, Wolffenbuttel BH,
487 Altshuler D, Ordovas JM, Boerwinkle E, Palmer CN, Thorsteinsdottir U, Chasman DI, Rotter JI, Franks PW,
488 Ripatti S, Cupples LA, Sandhu MS, Rich SS, Boehnke M, Deloukas P, Mohlke KL, Ingelsson E, Abecasis GR,
489 Daly MJ, Neale BM and Kathiresan S. Common variants associated with plasma triglycerides and risk for
490 coronary artery disease. *Nature genetics.* 2013;45:1345-52.
- 491 6. Rockman MV and Kruglyak L. Genetics of global gene expression. *Nature reviews Genetics.*
492 2006;7:862-72.
- 493 7. Weile J, Sun S, Cote AG, Knapp J, Verby M, Mellor JC, Wu Y, Pons C, Wong C, van Lieshout N,
494 Yang F, Tasan M, Tan G, Yang S, Fowler DM, Nussbaum R, Bloom JD, Vidal M, Hill DE, Aloy P and Roth FP.
495 A framework for exhaustively mapping functional missense variants. *Mol Syst Biol.* 2017;13:957.

- 496 8. Menche J, Sharma A, Kitsak M, Ghiassian SD, Vidal M, Loscalzo J and Barabasi AL. Disease
497 networks. Uncovering disease-disease relationships through the incomplete interactome. *Science*.
498 2015;347:1257601.
- 499 9. Sharma A, Halu A, Decano JL, Padi M, Liu YY, Prasad RB, Fadista J, Santolini M, Menche J, Weiss
500 ST, Vidal M, Silverman EK, Aikawa M, Barabasi AL, Groop L and Loscalzo J. Controllability in an islet
501 specific regulatory network identifies the transcriptional factor NFATC4, which regulates Type 2 Diabetes
502 associated genes. *NPJ systems biology and applications*. 2018;4:25.
- 503 10. Yao C, Chen BH, Joehanes R, Otlu B, Zhang X, Liu C, Huan T, Tastan O, Cupples LA, Meigs JB, Fox
504 CS, Freedman JE, Courchesne P, O'Donnell CJ, Munson PJ, Keles S and Levy D. Integromic analysis of
505 genetic variation and gene expression identifies networks for cardiovascular disease phenotypes.
506 *Circulation*. 2015;131:536-49.
- 507 11. Rolland T, Tasan M, Charlotheaux B, Pevzner SJ, Zhong Q, Sahni N, Yi S, Lemmens I, Fontanillo C,
508 Mosca R, Kamburov A, Ghiassian SD, Yang X, Ghamsari L, Balcha D, Begg BE, Braun P, Brehme M, Broly
509 MP, Carvunis AR, Convery-Zupan D, Corominas R, Coulombe-Huntington J, Dann E, Dreze M, Dricot A,
510 Fan C, Franzosa E, Gebreab F, Gutierrez BJ, Hardy MF, Jin M, Kang S, Kiros R, Lin GN, Luck K, MacWilliams
511 A, Menche J, Murray RR, Palagi A, Poulin MM, Rambout X, Rasla J, Reichert P, Romero V, Ruysinck E,
512 Sahalie JM, Scholz A, Shah AA, Sharma A, Shen Y, Spirohn K, Tam S, Tejada AO, Trigg SA, Twizere JC, Vega
513 K, Walsh J, Cusick ME, Xia Y, Barabasi AL, Iakoucheva LM, Aloy P, De Las Rivas J, Tavernier J, Calderwood
514 MA, Hill DE, Hao T, Roth FP and Vidal M. A proteome-scale map of the human interactome network. *Cell*.
515 2014;159:1212-1226.
- 516 12. Dalal JJ, Padmanabhan TN, Jain P, Patil S, Vasawala H and Gulati A. LIPITENSION: Interplay
517 between dyslipidemia and hypertension. *Indian J Endocrinol Metab*. 2012;16:240-5.
- 518 13. Joehanes R, Zhang X, Huan T, Yao C, Ying SX, Nguyen QT, Demirkale CY, Feolo ML, Sharopova NR,
519 Sturcke A, Schaffer AA, Heard-Costa N, Chen H, Liu PC, Wang R, Woodhouse KA, Tanriverdi K, Freedman
520 JE, Raghavachari N, Dupuis J, Johnson AD, O'Donnell CJ, Levy D and Munson PJ. Integrated genome-wide
521 analysis of expression quantitative trait loci aids interpretation of genomic association studies. *Genome*
522 *biology*. 2017;18:16.
- 523 14. Luck K, Kim DK, Lambourne L, Spirohn K, Begg BE, Bian W, Brignall R, Cafarelli T, Campos-Laborie
524 FJ, Charlotheaux B, Choi D, Cote AG, Daley M, Deimling S, Desbuleux A, Dricot A, Gebbia M, Hardy MF,
525 Kishore N, Knapp JJ, Kovacs IA, Lemmens I, Mee MW, Mellor JC, Pollis C, Pons C, Richardson AD,
526 Schlabach S, Teeking B, Yadav A, Babor M, Balcha D, Basha O, Bowman-Colin C, Chin SF, Choi SG,
527 Colabella C, Coppin G, D'Amata C, De Ridder D, De Rouck S, Duran-Frigola M, Ennajaoui H, Goebels F,
528 Goehring L, Gopal A, Haddad G, Hatchi E, Helmy M, Jacob Y, Kassa Y, Landini S, Li R, van Lieshout N,
529 MacWilliams A, Markey D, Paulson JN, Rangarajan S, Rasla J, Rayhan A, Rolland T, San-Miguel A, Shen Y,
530 Sheykhkarimli D, Sheynkman GM, Simonovsky E, Tasan M, Tejada A, Tropepe V, Twizere JC, Wang Y,
531 Weatheritt RJ, Weile J, Xia Y, Yang X, Yeger-Lotem E, Zhong Q, Aloy P, Bader GD, De Las Rivas J, Gaudet S,
532 Hao T, Rak J, Tavernier J, Hill DE, Vidal M, Roth FP and Calderwood MA. A reference map of the human
533 binary protein interactome. *Nature*. 2020;580:402-408.
- 534 15. Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, Ganna A, Chen J,
535 Buchkovich ML, Mora S, Beckmann JS, Bragg-Gresham JL, Chang HY, Demirkan A, Den Hertog HM, Do R,
536 Donnelly LA, Ehret GB, Esko T, Feitosa MF, Ferreira T, Fischer K, Fontanillas P, Fraser RM, Freitag DF,
537 Gurdasani D, Heikkila K, Hypponen E, Isaacs A, Jackson AU, Johansson A, Johnson T, Kaakinen M,
538 Kettunen J, Kleber ME, Li X, Luan J, Lyytikainen LP, Magnusson PKE, Mangino M, Mihailov E, Montasser
539 ME, Muller-Nurasyid M, Nolte IM, O'Connell JR, Palmer CD, Perola M, Petersen AK, Sanna S, Saxena R,
540 Service SK, Shah S, Shungin D, Sidore C, Song C, Strawbridge RJ, Surakka I, Tanaka T, Teslovich TM,
541 Thorleifsson G, Van den Herik EG, Voight BF, Volcik KA, Waite LL, Wong A, Wu Y, Zhang W, Absher D,
542 Asiki G, Barroso I, Been LF, Bolton JL, Bonnycastle LL, Brambilla P, Burnett MS, Cesana G, Dimitriou M,
543 Doney ASF, Doring A, Elliott P, Epstein SE, Ingi Eyjolfsson G, Gigante B, Goodarzi MO, Grallert H, Gravito

- 544 ML, Groves CJ, Hallmans G, Hartikainen AL, Hayward C, Hernandez D, Hicks AA, Holm H, Hung YJ, Illig T,
545 Jones MR, Kaleebu P, Kastelein JJP, Khaw KT, Kim E, Klopp N, Komulainen P, Kumari M, Langenberg C,
546 Lehtimäki T, Lin SY, Lindstrom J, Loos RJF, Mach F, McArdle WL, Meisinger C, Mitchell BD, Muller G,
547 Nagaraja R, Narisu N, Nieminen TVM, Nsubuga RN, Olafsson I, Ong KK, Palotie A, Papamarkou T, Pomilla
548 C, Pouta A, Rader DJ, Reilly MP, Ridker PM, Rivadeneira F, Rudan I, Ruokonen A, Samani N, Scharnagl H,
549 Seeley J, Silander K, Stancakova A, Stirrups K, Swift AJ, Tiret L, Uitterlinden AG, van Pelt LJ, Vedantam S,
550 Wainwright N, Wijmenga C, Wild SH, Willemsen G, Wilsgaard T, Wilson JF, Young EH, Zhao JH, Adair LS,
551 Arveiler D, Assimes TL, Bandinelli S, Bennett F, Bochud M, Boehm BO, Boomsma DI, Borecki IB, Bornstein
552 SR, Bovet P, Burnier M, Campbell H, Chakravarti A, Chambers JC, Chen YI, Collins FS, Cooper RS, Danesh
553 J, Dedoussis G, de Faire U, Feranil AB, Ferrieres J, Ferrucci L, Freimer NB, Gieger C, Groop LC, Gudnason
554 V, Gyllenstein U, Hamsten A, Harris TB, Hingorani A, Hirschhorn JN, Hofman A, Hovingh GK, Hsiung CA,
555 Humphries SE, Hunt SC, Hveem K, Iribarren C, Jarvelin MR, Jula A, Kahonen M, Kaprio J, Kesaniemi A,
556 Kivimäki M, Kooner JS, Koudstaal PJ, Krauss RM, Kuh D, Kuusisto J, Kyvik KO, Laakso M, Lakka TA, Lind L,
557 Lindgren CM, Martin NG, Marz W, McCarthy MI, McKenzie CA, Meneton P, Metspalu A, Moilanen L,
558 Morris AD, Munroe PB, Njolstad I, Pedersen NL, Power C, Pramstaller PP, Price JF, Psaty BM,
559 Quertermous T, Rauramaa R, Saleheen D, Salomaa V, Sanghera DK, Saramies J, Schwarz PEH, Sheu WH,
560 Shuldiner AR, Siegbahn A, Spector TD, Stefansson K, Strachan DP, Tayo BO, Tremoli E, Tuomilehto J,
561 Uusitupa M, van Duijn CM, Vollenweider P, Wallentin L, Wareham NJ, Whitfield JB, Wolfenbutter BHR,
562 Ordovas JM, Boerwinkle E, Palmer CNA, Thorsteinsdottir U, Chasman DI, Rotter JI, Franks PW, Ripatti S,
563 Cupples LA, Sandhu MS, Rich SS, Boehnke M, Deloukas P, Kathiresan S, Mohlke KL, Ingelsson E, Abecasis
564 GR and Global Lipids Genetics C. Discovery and refinement of loci associated with lipid levels. *Nature*
565 *genetics*. 2013;45:1274-1283.
- 566 16. Chen X, Gu X and Zhang H. Sidt2 regulates hepatocellular lipid metabolism through autophagy. *J*
567 *Lipid Res*. 2018;59:404-415.
- 568 17. Meng Y, Wang L and Ling L. Changes of lysosomal membrane permeabilization and lipid
569 metabolism in sidt2 deficient mice. *Exp Ther Med*. 2018;16:246-252.
- 570 18. Dean M and Annilo T. Evolution of the ATP-binding cassette (ABC) transporter superfamily in
571 vertebrates. *Annu Rev Genomics Hum Genet*. 2005;6:123-42.
- 572 19. Kaminski WE, Wenzel JJ, Piehler A, Langmann T and Schmitz G. ABCA6, a novel a subclass ABC
573 transporter. *Biochemical and biophysical research communications*. 2001;285:1295-301.
- 574 20. Pohl A, Devaux PF and Herrmann A. Function of prokaryotic and eukaryotic ABC proteins in lipid
575 transport. *Biochimica et biophysica acta*. 2005;1733:29-52.
- 576 21. Wada M, Daimon M, Emi M, Iijima H, Sato H, Koyano S, Tajima K, Kawanami T, Kurita K, Hunt SC,
577 Hopkins PN, Kubota I, Kawata S and Kato T. Genetic association between aldehyde dehydrogenase 2
578 (ALDH2) variation and high-density lipoprotein cholesterol (HDL-C) among non-drinkers in two large
579 population samples in Japan. *J Atheroscler Thromb*. 2008;15:179-84.
- 580 22. Gao J, Zhang Y, Yu C, Tan F and Wang L. Spontaneous nonalcoholic fatty liver disease and ER
581 stress in Sidt2 deficiency mice. *Biochemical and biophysical research communications*. 2016;476:326-
582 332.
- 583 23. Mendez-Acevedo KM, Valdes VJ, Asanov A and Vaca L. A novel family of mammalian
584 transmembrane proteins involved in cholesterol transport. *Scientific reports*. 2017;7:7450.
- 585 24. Kramer A, Green J, Pollard J, Jr. and Tugendreich S. Causal analysis approaches in Ingenuity
586 Pathway Analysis. *Bioinformatics*. 2014;30:523-30.
- 587 25. Ongen H, Brown AA, Delaneau O, Panousis NI, Nica AC, Consortium GT and Dermitzakis ET.
588 Estimating the causal tissues for complex traits and diseases. *Nature genetics*. 2017;49:1676-1683.
- 589 26. Musunuru K, Strong A, Frank-Kamenetsky M, Lee NE, Ahfeldt T, Sachs KV, Li X, Li H, Kuperwasser
590 N, Ruda VM, Pirruccello JP, Muchmore B, Prokunina-Olsson L, Hall JL, Schadt EE, Morales CR, Lund-Katz
591 S, Phillips MC, Wong J, Cantley W, Racie T, Ejebe KG, Orho-Melander M, Melander O, Kotliansky V,

- 592 Fitzgerald K, Krauss RM, Cowan CA, Kathiresan S and Rader DJ. From noncoding variant to phenotype via
593 SORT1 at the 1p13 cholesterol locus. *Nature*. 2010;466:714-9.
- 594 27. Atanasovska B, Kumar V, Fu J, Wijmenga C and Hofker MH. GWAS as a Driver of Gene Discovery
595 in Cardiometabolic Diseases. *Trends in endocrinology and metabolism: TEM*. 2015;26:722-732.
- 596 28. Jones NS, Watson KQ and Rebeck GW. Metabolic Disturbances of a High-Fat Diet Are Dependent
597 on APOE Genotype and Sex. *eNeuro*. 2019;6.
- 598 29. Yadav A, Vidal M and Luck K. Precision medicine - networks to the rescue. *Curr Opin Biotechnol*.
599 2020;63:177-189.
- 600 30. Oppenheimer GM. Becoming the Framingham Study 1947-1950. *Am J Public Health*.
601 2005;95:602-10.
- 602 31. Dawber TR, Meadors GF and Moore FE, Jr. Epidemiological approaches to heart disease: the
603 Framingham Study. *American journal of public health and the nation's health*. 1951;41:279-81.
- 604 32. Feinleib M, Kannel WB, Garrison RJ, McNamara PM and Castelli WP. The Framingham Offspring
605 Study. Design and preliminary data. *Preventive medicine*. 1975;4:518-25.
- 606 33. Splansky GL, Corey D, Yang Q, Atwood LD, Cupples LA, Benjamin EJ, D'Agostino RB, Sr., Fox CS,
607 Larson MG, Murabito JM, O'Donnell CJ, Vasan RS, Wolf PA and Levy D. The Third Generation Cohort of
608 the National Heart, Lung, and Blood Institute's Framingham Heart Study: design, recruitment, and initial
609 examination. *Am J Epidemiol*. 2007;165:1328-35.
- 610 34. Benjamin AM, Suchindran S, Pearce K, Rowell J, Lien LF, Guyton JR and McCarthy JJ. Gene by sex
611 interaction for measures of obesity in the framingham heart study. *J Obes*. 2011;2011:329038.
- 612 35. Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B and Speed TP. Summaries of Affymetrix
613 GeneChip probe level data. *Nucleic acids research*. 2003;31:e15.
- 614 36. Stegle O, Parts L, Piipari M, Winn J and Durbin R. Using probabilistic estimation of expression
615 residuals (PEER) to obtain increased power and interpretability of gene expression analyses. *Nat Protoc*.
616 2012;7:500-7.
- 617 37. Imai K, Keele L and Tingley D. A general approach to causal mediation analysis. *Psychological*
618 *methods*. 2010;15:309-34.
- 619 38. Hartwig FP, Davies NM, Hemani G and Davey Smith G. Two-sample Mendelian randomization:
620 avoiding the downsides of a powerful, widely applicable but potentially fallible technique. *International*
621 *journal of epidemiology*. 2017.
- 622 39. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, Laurin C, Burgess S, Bowden J,
623 Langdon R, Tan VY, Yarmolinsky J, Shihab HA, Timpson NJ, Evans DM, Relton C, Martin RM, Davey Smith
624 G, Gaunt TR and Haycock PC. The MR-Base platform supports systematic causal inference across the
625 human phenome. *Elife*. 2018;7.
- 626 40. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT,
627 Morris JH, Bork P, Jensen LJ and Mering CV. STRING v11: protein-protein association networks with
628 increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic*
629 *Acids Res*. 2019;47:D607-D613.

630

631

632

633

634

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

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

649

650

651

652

653

654

655

656

657

658

659

660

661

662

663

664

665

666

667

668

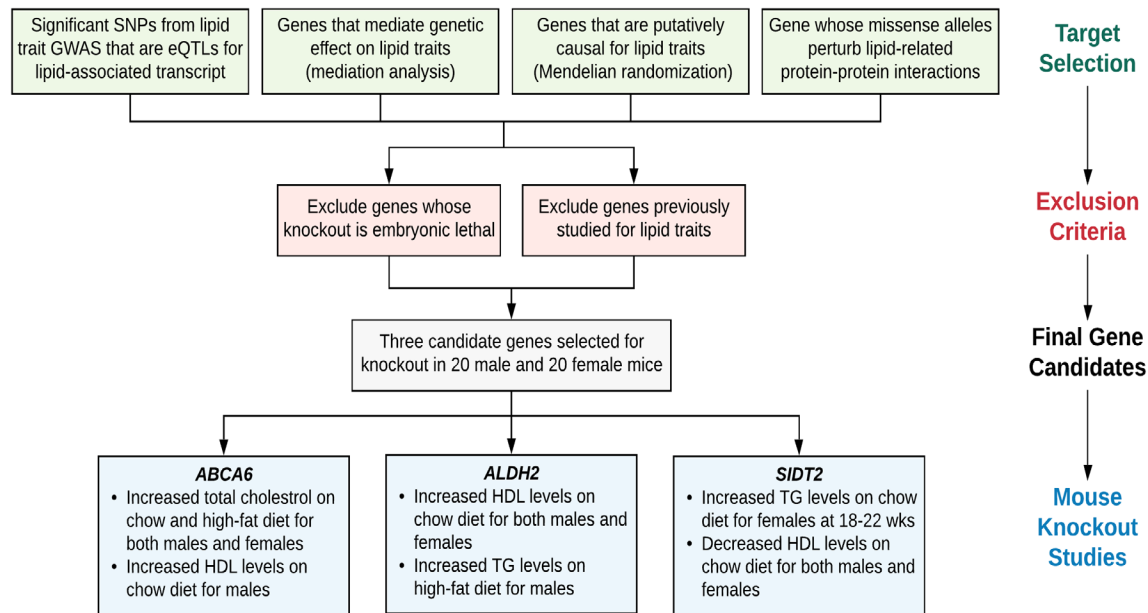
669

670

671

672 **Figures:**

673 **Figure 1. Study Design**



674

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

678

679

680

681

682

683

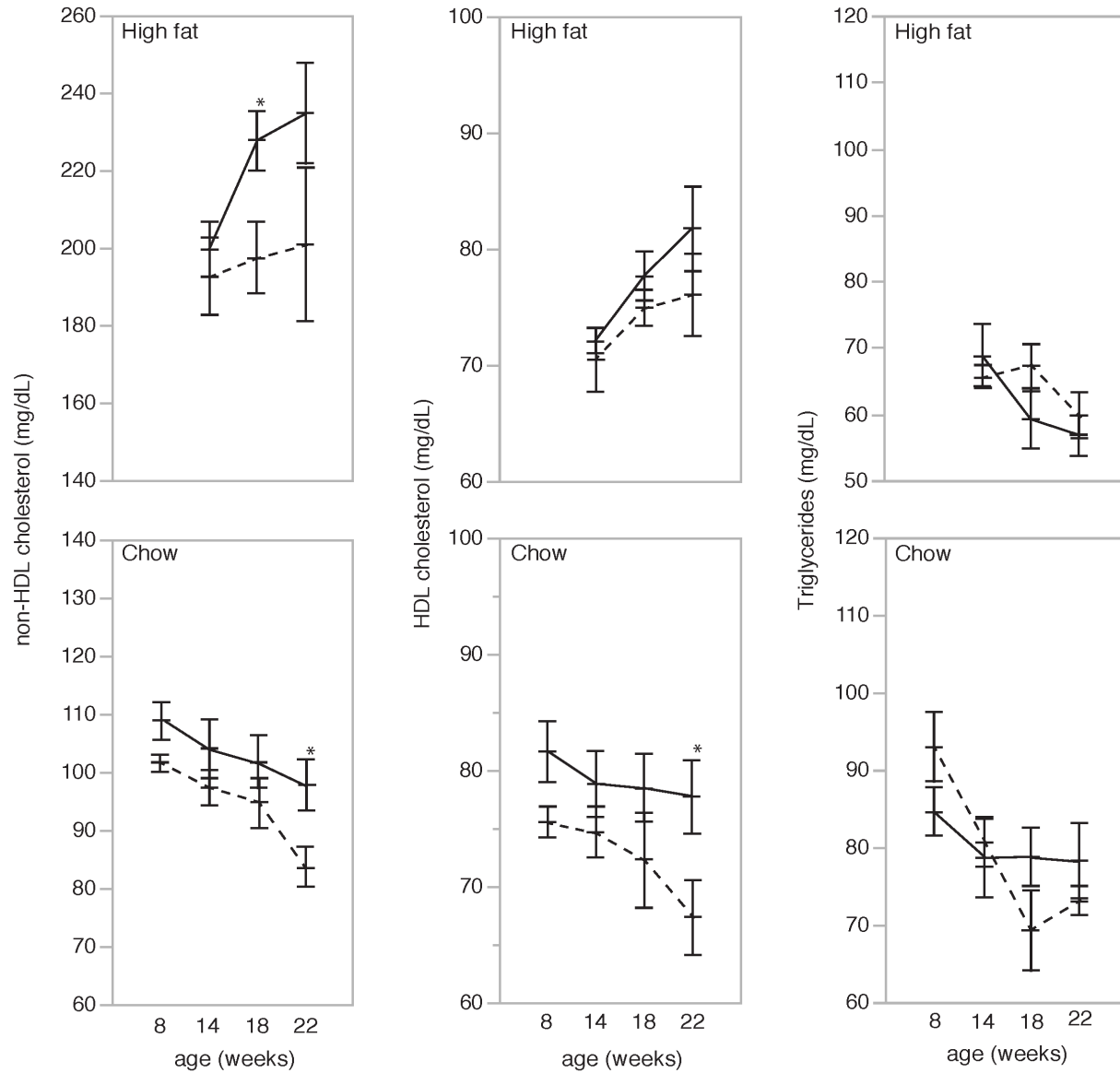
684

685

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

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

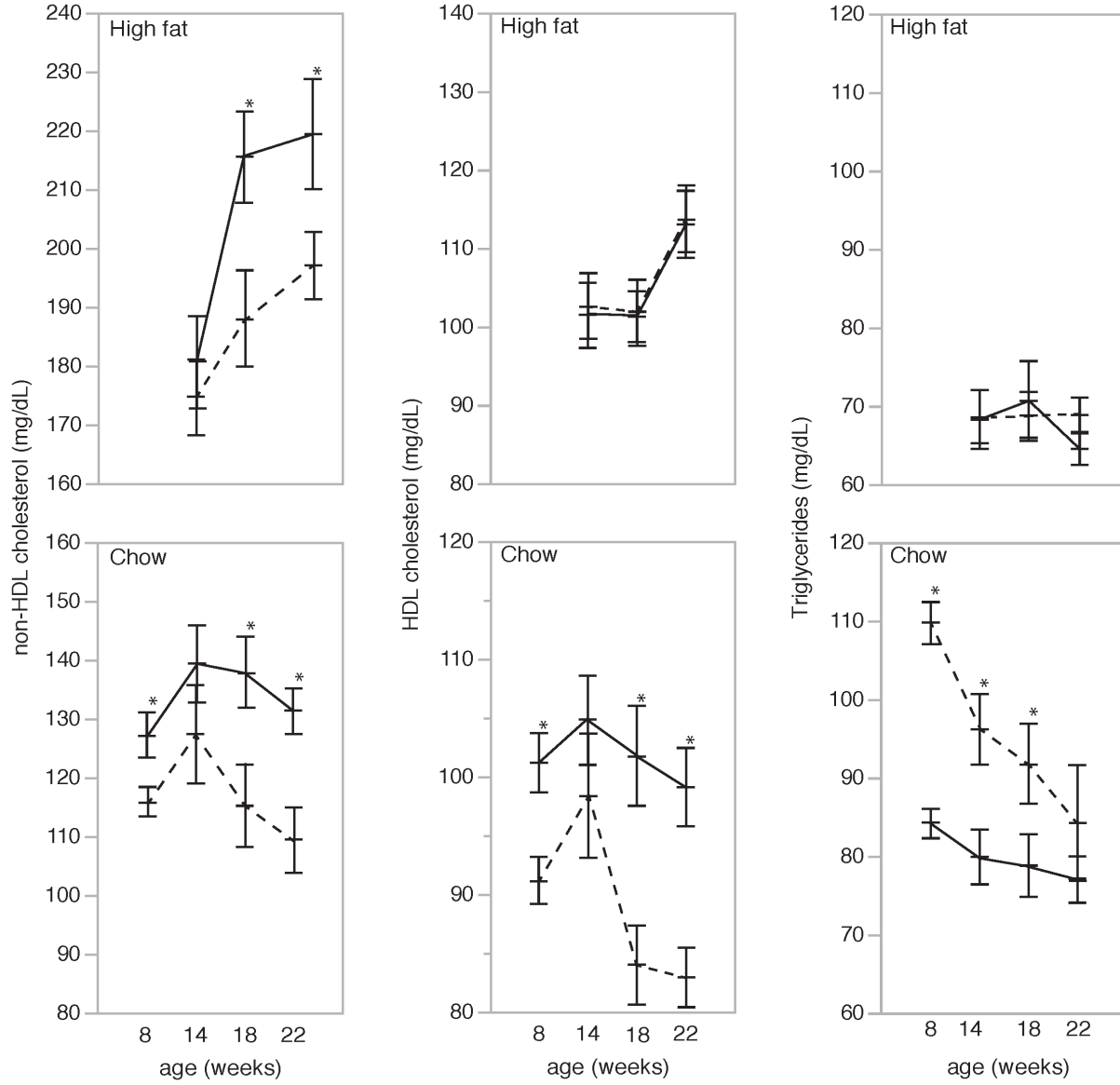
690 Female



691

692

693 Male



694

695

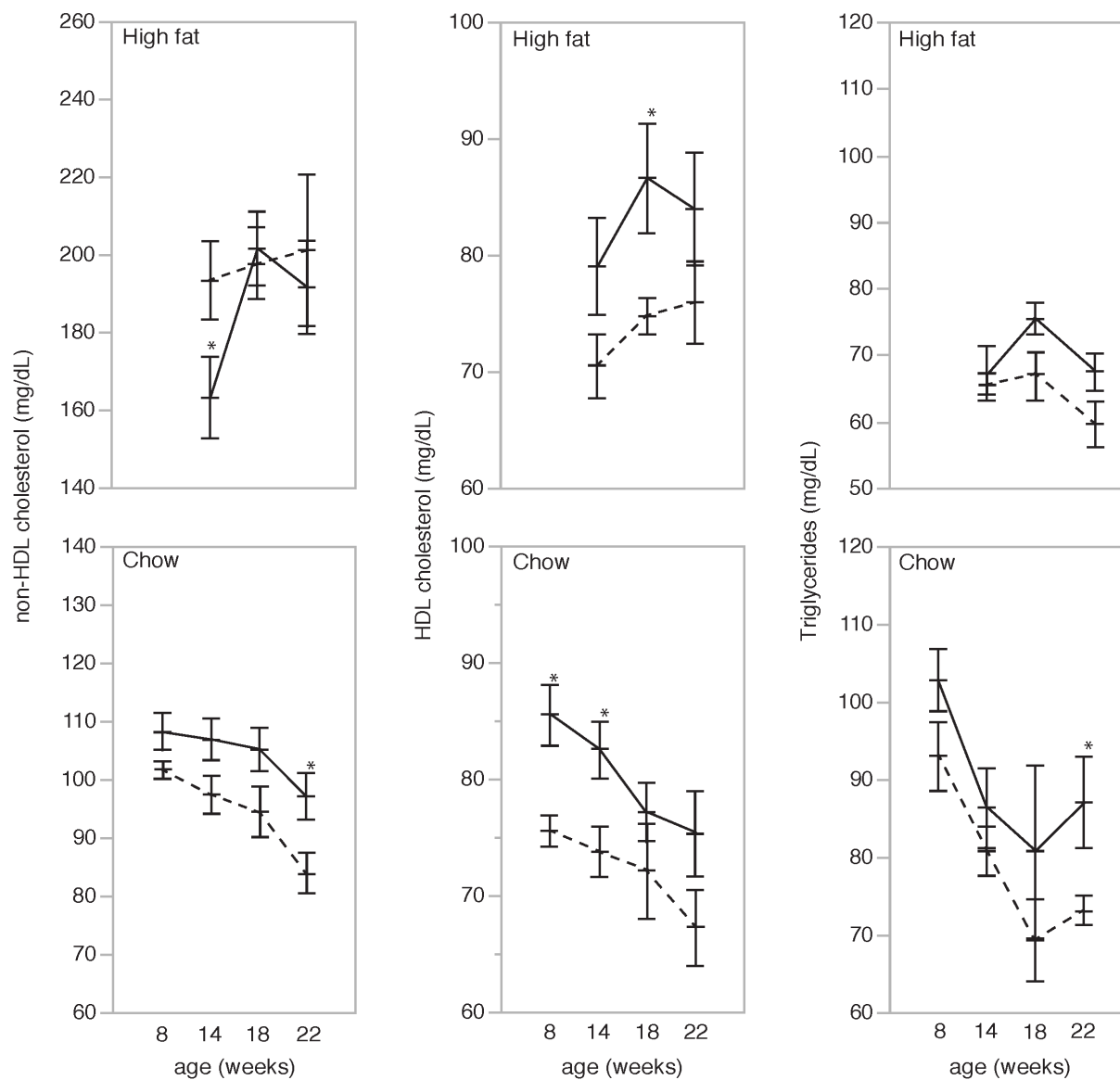
696

697

698

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

703 Female



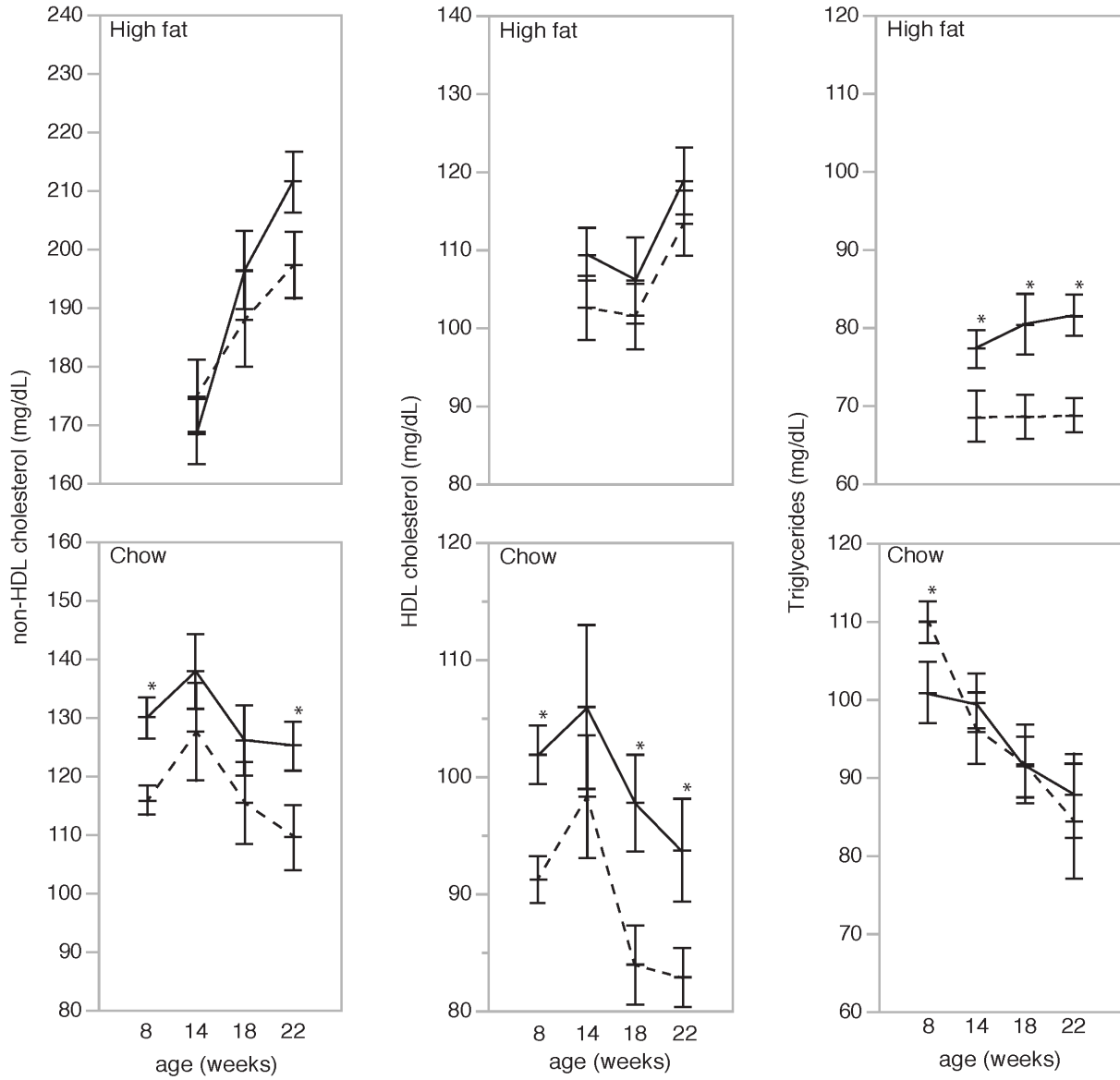
717

718

719

720

721 Male



722

723

724

725

726

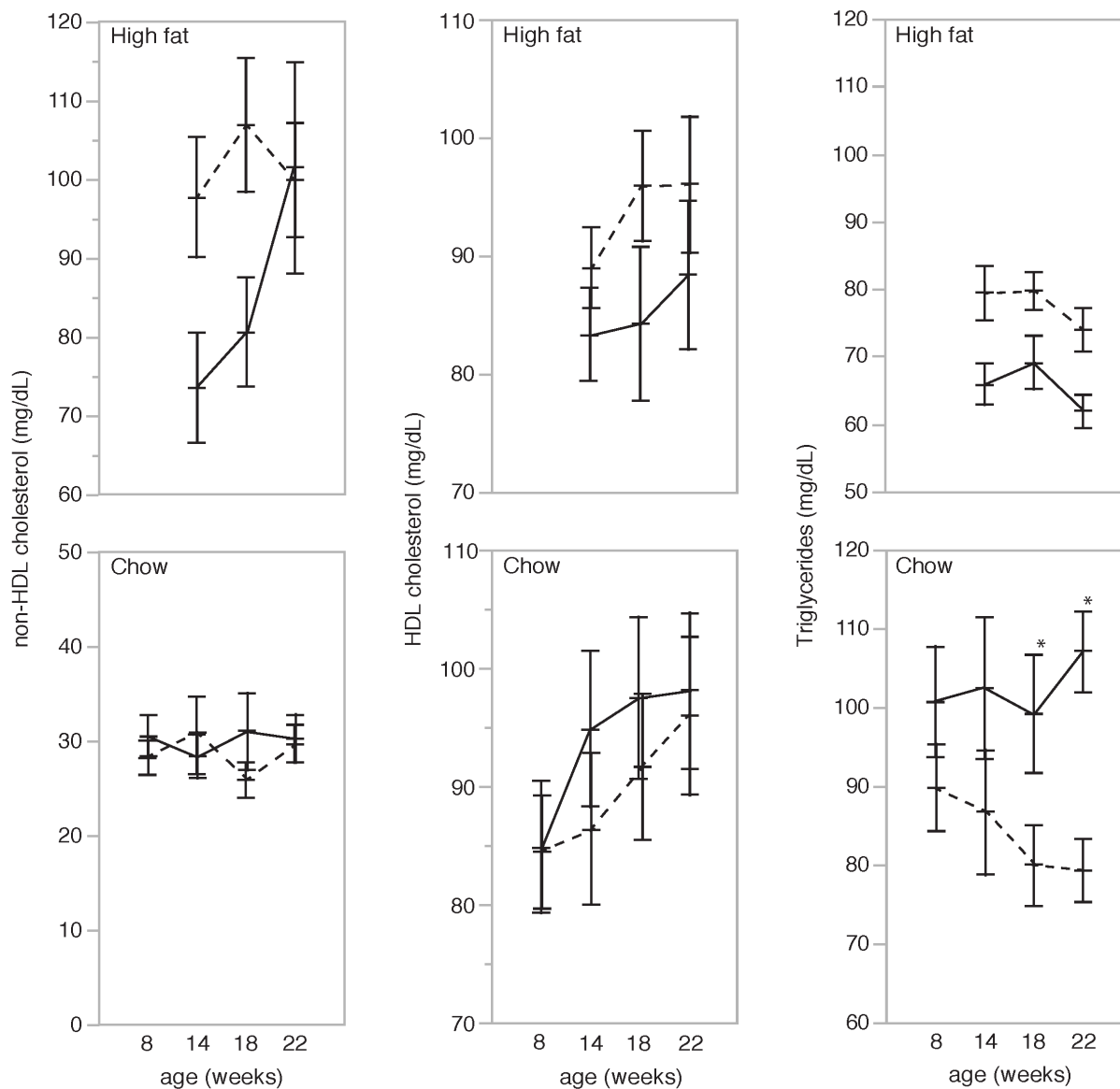
727

728

729

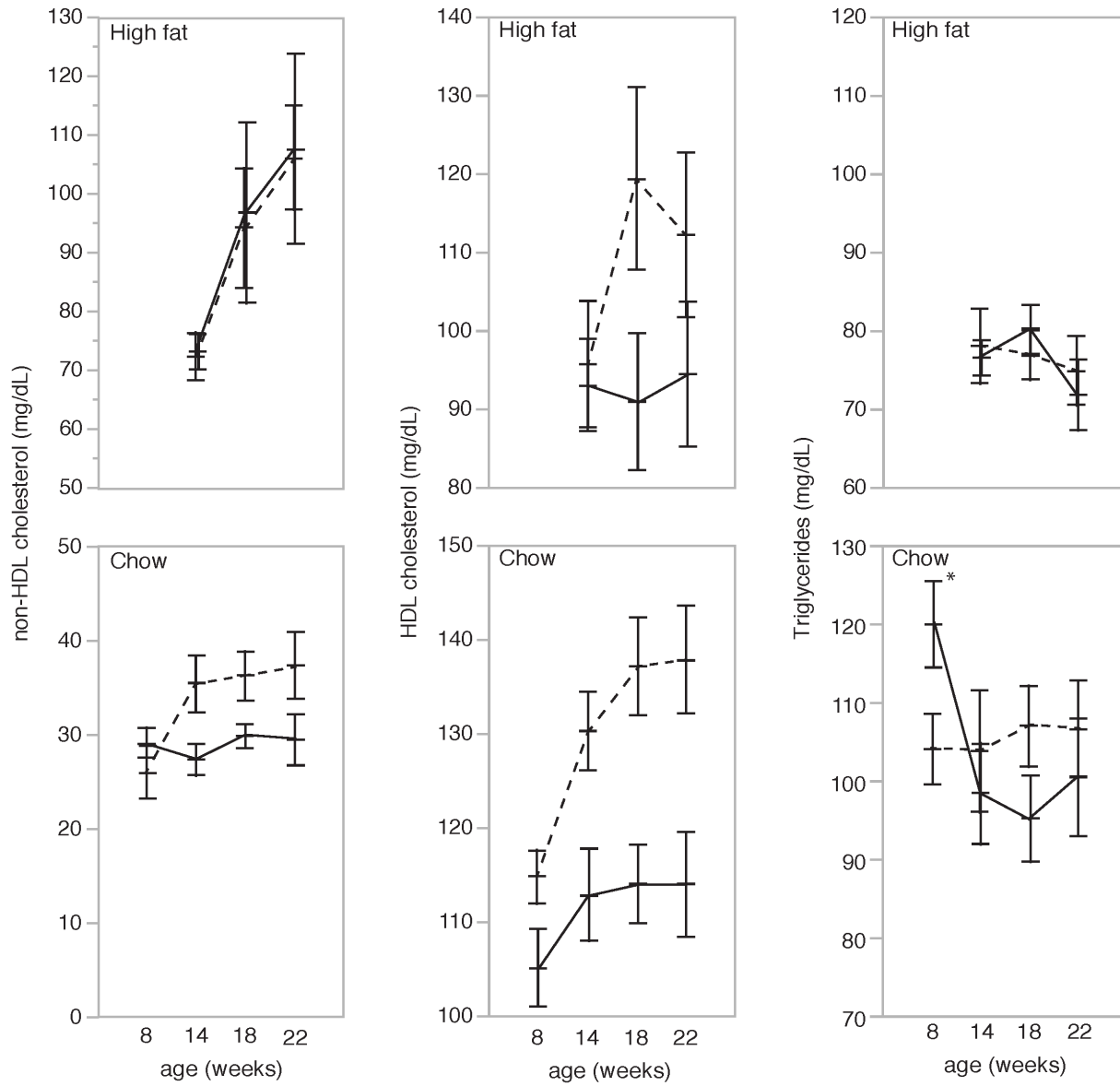
730 **Figure 4. Comparison between *Sid2* knockout (solid line) and wildtype (dotted line) mice in**
731 **males and females on regular chow and high-fat diet. Ten mice for each genotype and sex per**
732 **diet group. Asterisk indicates a significant difference ($P < 0.05$) between knockout and wildtype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

734 Female



735

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

744

745 **Table 2. Criteria of candidate gene selection**

746

	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

747

748

749

750