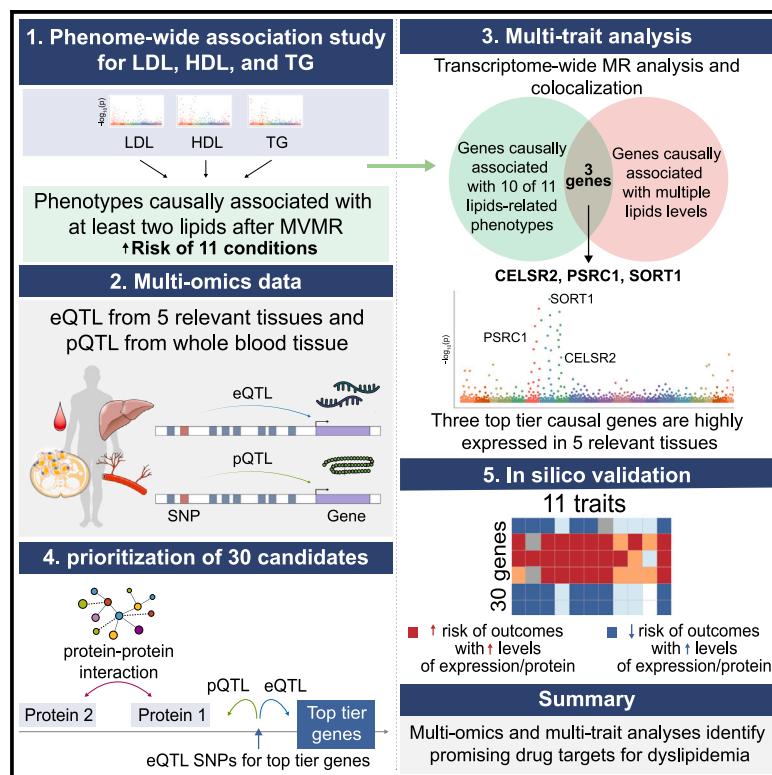


Prioritization of therapeutic targets for dyslipidemia using integrative multi-omics and multi-trait analysis

Graphical abstract



Authors

Min Seo Kim, Minku Song,
Beomsu Kim, ..., Pradeep Natarajan,
Ron Do, Hong-Hee Won

Correspondence

wonhh@skku.edu

In brief

Kim et al. identifies 30 therapeutic targets for dyslipidemia using genetic-driven causal inferences. This study demonstrates that the genetic-driven approach informs drug target prioritization, repurposing, and adverse effects of using lipid-lowering agents.

Highlights

- Multi-omics and multi-trait analyses prioritize 30 therapeutic targets for dyslipidemia
- Six out of 30 candidates are already approved or under drug investigations (i.e., PCSK9)
- Drug targets include genes that have not been detected in large GWASs for lipids levels



Article

Prioritization of therapeutic targets for dyslipidemia using integrative multi-omics and multi-trait analysis

Min Seo Kim,^{1,8} Minku Song,^{1,8} Beomsu Kim,¹ Injeong Shim,¹ Dan Say Kim,¹ Pradeep Natarajan,^{2,3,4} Ron Do,^{5,6} and Hong-Hee Won^{1,7,9,*}

¹Samsung Advanced Institute for Health Sciences & Technology (SAIHST), Sungkyunkwan University, Samsung Medical Center, Seoul, Republic of Korea

²Medical and Population Genetics and Cardiovascular Disease Initiative, Broad Institute of MIT and Harvard, Cambridge, MA, USA

³Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA, USA

⁴Department of Medicine, Harvard Medical School, Boston, MA, USA

⁵The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA

⁶Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA

⁷Samsung Genome Institute, Samsung Medical Center, Seoul, Republic of Korea

⁸These authors contributed equally

⁹Lead contact

*Correspondence: wonhh@skku.edu

<https://doi.org/10.1016/j.xcrm.2023.101112>

SUMMARY

Drug targets with genetic support are several-fold more likely to succeed in clinical trials. We introduce a genetic-driven approach based on causal inferences that can inform drug target prioritization, repurposing, and adverse effects of using lipid-lowering agents. Given that a multi-trait approach increases the power to detect meaningful variants/genes, we conduct multi-omics and multi-trait analyses, followed by network connectivity investigations, and prioritize 30 potential therapeutic targets for dyslipidemia, including SORT1, PSRC1, CELSR2, PCSK9, HMGCR, APOB, GRN, HFE2, FJX1, C1QTNF1, and SLC5A8. 20% (6/30) of prioritized targets from our hypothesis-free drug target search are either approved or under investigation for dyslipidemia. The prioritized targets are 22-fold higher in likelihood of being approved or under investigation in clinical trials than genome-wide association study (GWAS)-curated targets. Our results demonstrate that the genetic-driven approach used in this study is a promising strategy for prioritizing targets while informing about the potential adverse effects and repurposing opportunities.

INTRODUCTION

In 2019, dyslipidemia became the eighth leading cause of mortality and a major risk factor for cardiovascular morbidities and mortalities.¹ Despite a high global demand for lipid-lowering agents, the number of drugs available for statin-intolerant or refractory patients is limited.² Cardiovascular drug development has stagnated compared with other diseases. Between 2010 and 2017, the approval rate for cardiovascular drugs entering phase I was 4%, compared with a rate of 9% for anticancer drugs.^{3,4} Randomized clinical trial (RCT) failures have largely been attributed to the lack of efficacy (~50%) and adverse effects (~25%) that result in huge costs and limited access to novel agents.^{3,5}

Increasing evidence suggests that drug targets with genetic support are more likely to succeed in clinical trials and reach the market.^{3,6} However, the translational use of genetic data, such as data from genome-wide association studies (GWASs), in drug development pipelines remain challenging, as the GWAS ap-

proaches cannot reliably pinpoint causal variants or genes.⁷ Therefore, Mendelian randomization (MR), an analytic approach that uses genetic tools to mimic RCTs and assess causation, is increasingly used to identify causal and druggable targets.^{3,7,8} Recently, using the MR approach, Gaziano et al. explored therapeutic targets and repurposing opportunities for treating coronavirus disease 2019 (COVID-19). They successfully isolated *IFNAR2* and *ACE2* as causal genes and proposed them as the main targets for early COVID-19 management.⁹

As lipid-lowering efforts are aimed at preventing the risk of lipid-mediated diseases, especially cardiovascular diseases (CVDs),² we hypothesized that pleiotropic variants/genes that control both lipid levels and lipid-related traits are more likely to serve as therapeutic purposes.¹⁰ Given that a multi-trait approach may increase the detection of meaningful variants and actionable drug targets,^{10,11} we conducted a multi-trait transcriptome-wide MR analysis (tested >15,000 genes) to prioritize druggable targets that would probably prevent the composite risk of lipid perturbations and multiple lipid-related traits. We



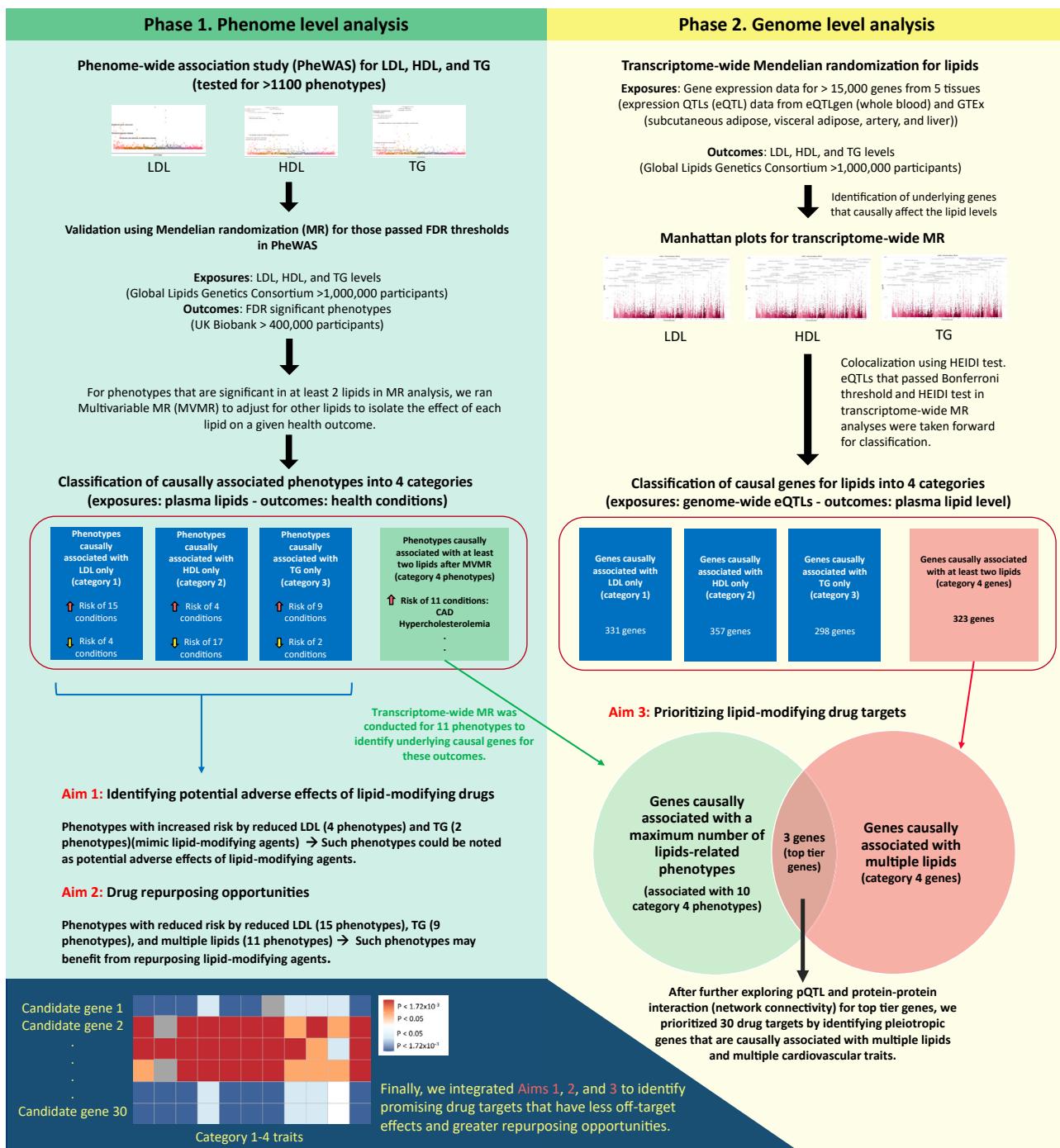


Figure 1. Study design and flow diagrams

For phase 1 (phenome-level analysis), phenome-wide association study (PheWAS) adjusted for age, gender, genotype array, and principal components was conducted for each lipid. Phenotypes that passed the false discovery rate (FDR) threshold in PheWAS were validated using Mendelian randomization (MR) analysis. Phenotypes significant in at least 2 lipids in MR analysis were further subjected to multi-variable MR (MVMR) adjusted for other lipids to isolate the effect of each lipid on a given health outcome. Finally, MR-validated phenotypes were classified into 4 categories (phenotypes causally affected by LDL [1], HDL [2], TG [3], and multiple lipids [4]), and categories 1–3 were used to implicate potential adverse effects and repurposing opportunities for lipid-modifying drugs. For phase 2 (genome-level analysis), transcriptome-wide MR for lipids was performed to identify causal genes for three lipids and 11 category 4 phenotypes (Figure 3A, steps 2–3). Subsequently, we identified strong pleiotropic causal genes (top-tier genes) that are associated with multiple lipids and multiple traits. Top-tier genes

(legend continued on next page)

aimed to explore pleiotropic targets for dyslipidemia that have a high probability of trial success and to demonstrate how our framework could integrate human genetics into the life cycle of drug development as per the guidelines³: identification of therapeutic targets, repurposing opportunities, and potential adverse effects of using lipid-lowering agents.

RESULTS

Phenome-wide associations

As lipid levels measured at specific time points may not accurately capture individuals' representative levels, lipid polygenic risk score (PRS) was used to indicate genetically predicted lipid levels for each person. Using PRS for lipids as an exposure and phecodes as an outcome (Data S1: Table S2, related to Figure 2A), we performed a phenome-wide association study (PheWAS) to identify associated phenotypes. Among 1,129 tested phenotypes, 97 phenotypes were significant with a false discovery rate (FDR) <0.05 for low-density lipoprotein (LDL) (Data S1: Table S3, related to Figure 2), 317 phenotypes were significant with an FDR <0.05 for high-density lipoprotein (HDL) (Data S1: Table S4, related to Figure 2), and 291 phenotypes were significant with an FDR <0.05 for triglyceride (TG) (Data S1: Table S5, related to Figure 2). Manhattan plots for PheWAS results are shown in Figure S1.

MR

An MR study should meet three main assumptions¹²: (1) the genetic instrument is associated with exposure, (2) the genetic instrument is not associated with outcome and should exist through the exposure, and (3) there should be no association between confounders of exposure-outcome association and genetic instrument. First, we conducted univariate two-sample MR analyses using four MR methods, namely inverse-variance weighted (IVW),¹³ weighted median,¹⁴ MR-Egger,¹⁵ and MR-PRESSO,¹⁶ based on different assumptions, and phenotypes identified as significant via at least 2 MR methods were retained (Data S1: Tables S6–S8, related to Figure 2). For phenotypes that were significant as per at least two lipids in univariate MR analysis, we performed multi-variate MR (MVMR), adjusting for other lipids to isolate the effects of each lipid on a specific health outcome (Data S1: Table S9, related to Figure 2). All MR-validated outcomes were categorized into four groups (Figure 2C; Data S1: Table S10, related to Figure 2): 19 phenotypes were causally associated with LDL alone (category 1 phenotype), 21 phenotypes were causally associated with HDL alone (category 2 phenotype), 11 phenotypes were causally linked to TG alone (category 3 phenotype), and 11 phenotypes were causally associated with at least two lipids after MVMR (category 4 phenotype).

Potential adverse effects and repurposing opportunities associated with lipid-lowering effects

We inferred potential adverse effects and repurposing opportunities during lipid-lowering agent usage (Figure 2C). Phenotypes

with increased risk due to reduced LDL (4 phenotypes) and TG (2 phenotypes) mimic lipid-modifying agents and thus can be identified as potential adverse effects of using lipid-lowering agents. Phenotypes showing reduced risk upon reducing levels of LDL (15 phenotypes), TG (9 phenotypes), and multiple lipids (11 phenotypes) may benefit from repurposing lipid-modifying agents.

Transcriptome-wide MR

The summary data-based MR (SMR) approach selects the most relevant single *cis*-expression quantitative trait locus (*cis*-eQTL) single-nucleotide polymorphism (SNP) (i.e., top *cis*-eQTL SNP) for each gene as an instrument. The eQTL data for over 15,000 genes were used to run hypothesis-free transcriptome-wide MR analysis in five tissues (whole blood, subcutaneous adipose tissue, visceral adipose tissue, arterial tissue, and liver tissue) for multiple lipids and lipid-driven outcomes; five tissues were selected based on the linkage disequilibrium score regression (LDSC) heritability presented in Figure S2. Among >15,000 genes, transcriptome-wide MR isolated 485 genes for LDL, 617 genes for HDL, and 573 genes for TG. These genes were significant in Bonferroni threshold for SMR and heterogeneity in dependent instrument (HEIDI) colocalization (Data S1: Tables S11–S13, related to Figure 3). Gene was counted once when a given gene is expressed significantly in multiple tissues. HEIDI colocalization eliminated approximately 25% of the signals from SMR; the extent to which the signals were excluded was similar to that of the previous study.¹⁷ The Manhattan plots for transcriptome-wide MR results are provided in Figure S3. After categorizing causal genes for lipid level into 4 groups (Figure 1, phase 2), 331 genes were causally associated only with the LDL (category 1 genes), 357 genes were causally associated only with the HDL (category 2 genes), 298 genes were causally associated only with the TG (category 3 genes), and 323 genes were causally associated with at least two lipids (category 4 genes) (Figure 1; Data S1: Table S14, related to Figure 3).

Drug target prioritization for lipid lowering

As exploring the pleiotropic gene could be a promising strategy for identifying druggable targets,¹⁰ we conducted eQTL MR using category 4 genes (323 genes) as exposure and 11 phenotypes (category 4) as outcomes to isolate strong pleiotropic genes (Figure 1; Data S1: Table S15, related to Figure 3). Among category 4 genes that caused multiple lipid profiles, the top three genes (SORT1, CELSR2, and PSRC1) that caused 10 lipid-driven traits were isolated as well (Data S1: Tables S16 and S17, related to Figure 3; we termed these three genes top-tier genes). The top six *cis*-eQTLs (rs12740374 [effect allele: T], rs7528419 [G], rs655246 [A], rs586254 [A], rs688386 [T], rs12081530 [A]) of the top-tier genes were retained from five tissues and were subsequently used to connect associated proteins (protein QTL [pQTL] with <1E–5; Data S1: Table S18, related to Figure 3; Figure 3A). We further curated non-seed genes that lack direct

were further subjected to drug target prioritization analysis (Figure 3A, steps 4–6). Finally, 30 candidate drug target genes for dyslipidemia were identified and their causal associations with category 1–4 traits were explored to prioritize drug targets that have fewer off-target effects (safety concerns) and greater repurposing opportunities (other health benefits).

See also Data S1: Tables S3–S14.

Table 1. Prioritized candidate therapeutic targets for lipid modification (30 candidates)

	Prioritization methods				Scoring methods							Drug development ^a		Validation
	Top-tier eQTL	Top-tier pQTL	PPI-GO	PPI-KEGG	GWAS Score	TWAS genes	MR genes	EWAS genes	PheWAS genes	Text mining	Animal experiments	Approved drug target genes	Investigational drug target genes	
APOB	✓	✓	✓	✓	7	✓	✓	✓	✓	✓	✓	✓ ^a		yes
PCSK9		✓			6	✓	✓	✓		✓	✓	✓ ^a		yes
HMGCR		✓			6	✓	✓	✓		✓	✓	✓ ^a		yes
SORT1	✓	✓			6	✓	✓	✓		✓	✓			yes
PSRC1	✓				6	✓	✓	✓		✓	✓			yes
ABCA1		✓	✓		6	✓	✓	✓		✓	✓	✓ ^a		yes
LDLR		✓			6	✓	✓		✓	✓	✓			yes
CYP7A1		✓			6	✓	✓	✓		✓	✓			yes
APOE		✓			5	✓	✓			✓	✓			yes
CELSR2	✓				5	✓	✓	✓		✓	✓			yes
APOA1		✓	✓		5	✓	✓			✓	✓	✓ ^c		no
LCAT		✓			5	✓		✓		✓	✓	✓ ^a		no
PLA2G12B	✓		✓		4	✓			✓	✓	✓			yes
TNFSF12	✓				4	✓			✓	✓	✓	✓ ^d		yes
MTTP		✓			4	✓			✓	✓	✓	✓ ^a		yes
APOA2	✓				4	✓			✓	✓	✓			no
SELE	✓				3	✓				✓	✓			yes
SREBF2		✓			3			✓		✓	✓			no
CYP11A1	✓				3	✓				✓	✓			yes
GRN	✓				3	✓			✓	✓				yes
C1QTNF1	✓				2	✓				✓				yes
FJX1	✓				2					✓	✓			yes
CAT	✓	✓			2					✓	✓			yes
NEO1	✓				2	✓		✓						yes
STC1	✓				2					✓	✓			yes
LRP1B	✓				1		✓							yes
SLC5A8	✓				1					✓				yes
HFE2	✓				1					✓				yes
CA10	✓				1					✓				yes
RGMA	✓				0									yes

Check-marked for any category that applies to the given gene. In addition to top-tier gene-based curation, further targets were identified using network connectivity information based on protein-protein and pathway molecular interactions. We prioritized 30 drug targets. It should be noted that this does not necessarily mean that all 30 genes are druggable. Prioritization simply implies a greater probability of searching for druggable targets out of many genes (approximately 20,000), focusing the resources and research efforts, and minimizing the time and cost risks associated with drug development failure. Scoring system consists of 7 parts: (1) significant genes ($p < 5E-8$) identified from GWASs of lipid levels, (2) significant genes ($p < 1E-6$) identified from TWASs of lipids or related CVDs, (3) causal genes identified from eQTL MR of our study (Data S1: Table S14, related to Figure 3), (4) significant genes ($p < 1E-7$) identified from EWASs of lipids or related CVDs, (5) genes significantly associated with lipids or related CVDs in PheWASs (passing Bonferroni threshold), (6) genes prioritized by PubMed text mining, and (7) genes that have been investigated for their associations with lipids and CVDs using animal models. The overall score indicates the number of criteria fulfilled. PPI, protein-protein interaction; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; MR, Mendelian randomization; GWAS, genome-wide association study; TWAS, transcriptome-wide association study; EWAS, epigenome-wide association study; PheWAS, phenotype-wide association study; CVD, cardiovascular diseases.

^aApproved or investigational drugs for dyslipidemia.

^bTargets that have shown statistically significant association ($p < 0.05$) at least once in MR validation 1–3 (Figure 5) or MetaXcan for hypercholesterolemia or hyperlipidemia.

^cDrug target under investigation (phase 1–3) for atherosclerosis and acute coronary syndromes.

^dDrug target under investigation for solid cancers.

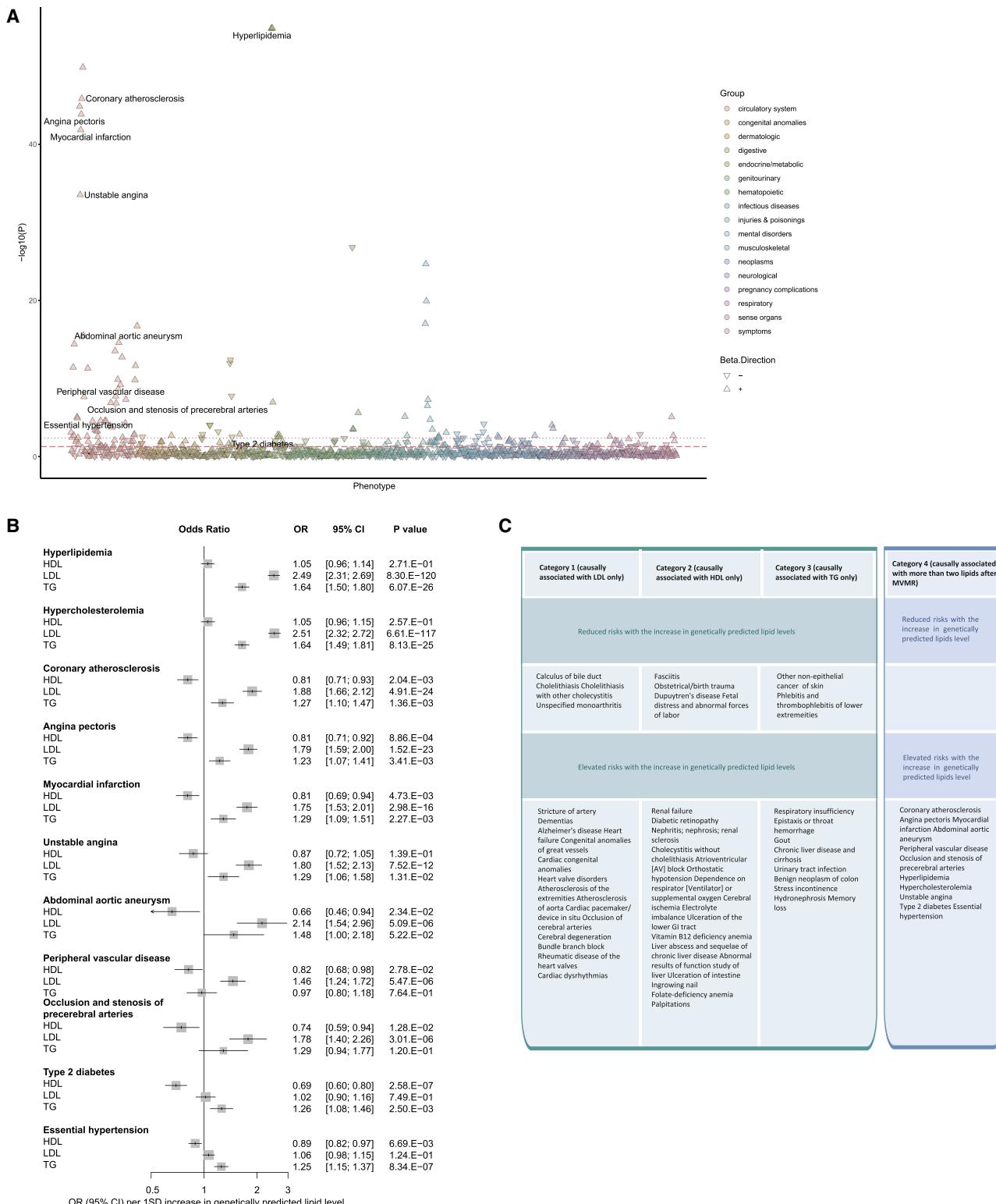


Figure 2. Phase 1: Phenome-level analysis to identify potential adverse effects and repurposing opportunities for lipid-modifying agents

(A) Manhattan plot for PheWAS of LDL. Manhattan plots for HDL and TG are provided in the [supplemental information](#). Category 4 phenotypes are indexed. For results, dot line indicates significant ($p < 0.05$), dashed line indicates significant with an FDR < 0.05 .

(B) MVMR results for 11 phenotypes that are causally associated with at least two lipids (category 4 phenotypes).

(legend continued on next page)

genetic evidence but have high potential based on the network connectivity using STITCH18. The Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment tools in STITCH were used to identify additional proteins and pathways that interacted with the proteins identified in an earlier step. The gene set with the most significant biological GO (cholesterol metabolic process; FDR < 1.26E-15) and KEGG pathway (fat digestion and absorption; FDR < 7.17E-6) were identified. Genes consisting of GO and KEGG gene sets were further accrued for prioritization. Initially, 30 candidates were listed after considering network connectivity (Table 1).

Likelihood of drug success

The ongoing drug trial evidence (phase 1–3) were obtained from the Pharmaprojects database, and drug approval evidence was obtained from the WHO Collaborating Center for Drug Statistics Methodology (WHOCC), DrugBank, and ChEMBL databases (Data S1: Tables S20 and S21, related to Figure 3). Compared with the odds of approved or investigational drug targets among GWAS-significant lipid genes (13/1,171 [1.1%]), the proportion of pleiotropic genes causally associated with multiple lipids and more than half of lipid-related phenotypes (associated with ≥6 category 4 phenotypes) (Data S1: Table S16, related to Figure 3) was 4.69-fold (Figure 4; Data S1: Table S19, related to Figure 3). GWAS variants/genes with $p < 5E-8$ were identified from the latest Global Lipids Genetics Consortium (GLGC) GWAS for lipids (Data S1: Table S22, related to Figures 2 and 3).¹⁸ Further prioritization based on top-tier genes and relevant pQTL/protein-protein interactions (Table 1) additionally increased the odds of success (odds ratio [OR] 22.27; 95% confidence interval [CI] 7.81–63.54; Figure 4). Further narrowing down to high-scoring (≥ 4) targets doubled the odds of success (OR 53.45; 95% CI, 16.92–168.84; Figure 4).

Exploration for the possible adverse effects of target genes using PheWAS

The SORT1 was one of the highest scored targets supported by diverse omics-data and experiments that is not yet under investigation (Table 1). A previous study experimentally validated rs12740374 and identified it as a driving variant of SORT1 effect.¹⁹ To explore the possible unwanted effects of SORT1, PheWAS for rs12740374 (T) of SORT1 was conducted across phenotypes with at least 200 cases (1,129 phenotype data from UK Biobank) adjusted for age, gender, principal components (PCs), and genotype array (Figure S4). The minor allele T (effect allele) for rs12740374 was associated with reduced lipid levels and multiple cardiovascular burdens, whereas no noticeable adverse effect was identified at an FDR <0.05 threshold level (no traits with the positive beta), indicating rs12740374 to be a potential therapeutic target (Data S1: Table S23, related to Figure 3). Prioritized genes were highly upregulated in the liver compared with other tissues (Figure S5).

Integration of aims 1, 2, and 3

We explored repurposing of some selected drugs as well as the potential adverse effects associated with 30 prioritized genes using MR (Figure 5; Data S1: Tables S24–S31, related to Figures 5 and 6). The decrease in apolipoprotein B-100 (APOB) was associated with the reduced risk of dyslipidemia and lipid-driven cardiometabolic traits (Figure 5A), along with other health outcomes (Figure 5B), whereas it was associated with the increased risk of cholelithiasis, chronic liver diseases, and fasciitis (Figure 5B). The decreased levels of phospholipase A2 group XIIB (PLA2G12B) was associated with the reduced risk of dyslipidemia but with the increased risk of diabetes mellitus and diabetic retinopathy (Figure 5B). We also identified targets that were associated with a wide range of health benefits with few adverse effects, including GRN, HFE2, FJX1, C1QTNF1, and SLC5A8 (Figure 5; Table 1).

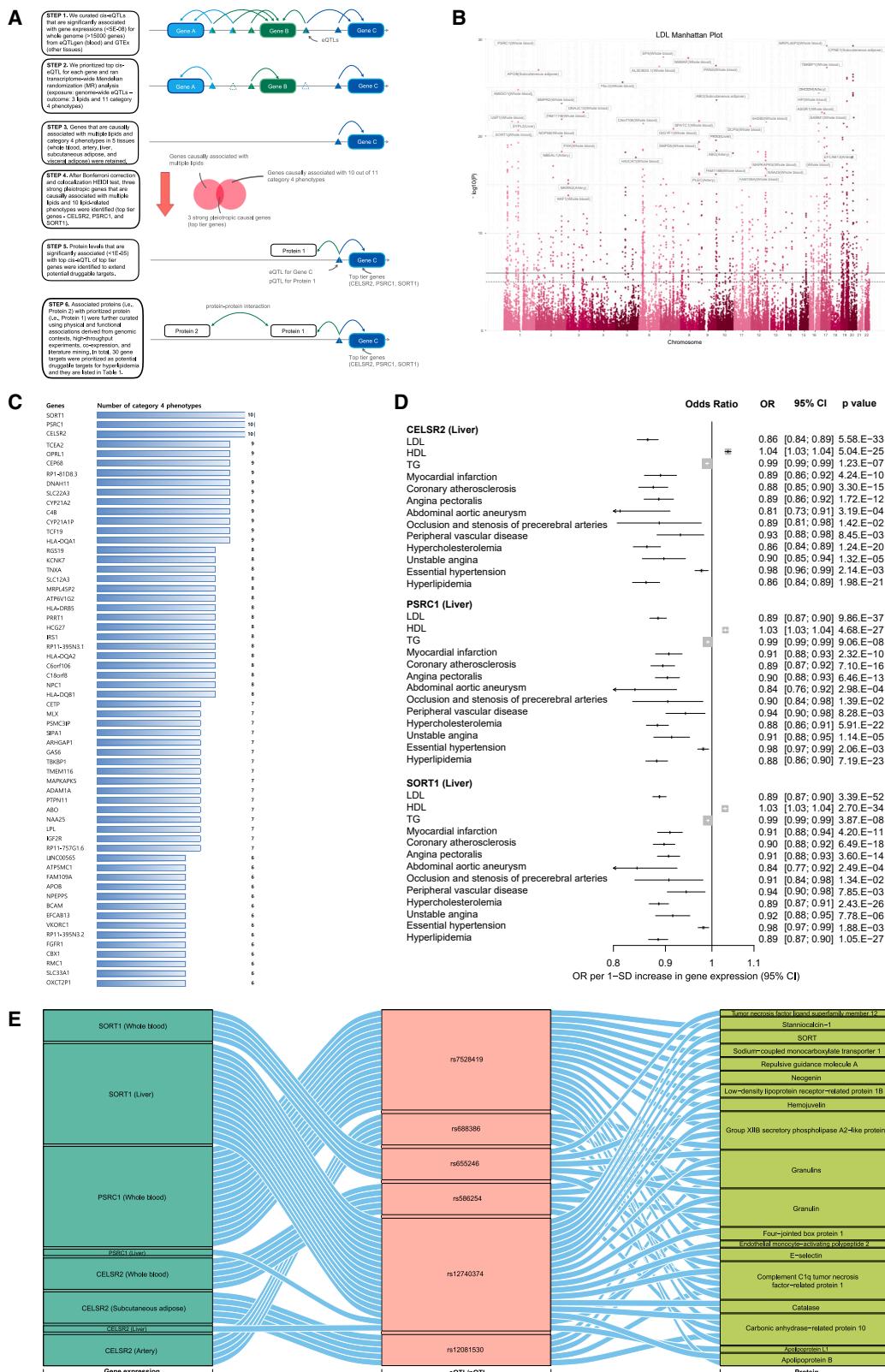
In silico validation

Validation for potential causal relationships between 29 therapeutic candidates (one target [SREBF2] is lacking eQTL/pQTL data) and category 4 phenotypes are demonstrated in Figure 6 and Data S1 (Tables S32–S37, related to Figure 6). Results of the validation using transcriptome-wide gene-based analysis (MetaXcan) are provided in Figure S6 and Data S1 (Tables S38–S41, related to Table 1). All candidates among the 30 prioritized genes were successfully validated except for APOA1, APOA2, LCAT, and SREBF2 (Table 1).

DISCUSSION

We demonstrated that integration of functional genomics and knowledge of network connectivity maximizes the utility of genetics for therapeutic target prioritization and isolated 30 target candidate genes (out of more than 15,000 genes) that are worth investigating to reduce the burden of dyslipidemia. Prioritized genes were mostly upregulated in the liver and heart compared with their expression in other tissues. Six out of 30 prioritized candidates (20%) were already approved or under investigation for dyslipidemia, implying that the integrative multi-omics and multi-trait approach successfully identified potential target genes with a higher probability of success, which was approximately 22-fold compared with that of GWAS-curated targets. Further narrowing down to high-scoring (≥ 4) targets doubled the odds of success (OR 53.45) (Figure 4). Moreover, we simultaneously assessed the repurposing opportunities and anticipated adverse effects associated with the lipid lowering (Figures 1 and 2C). Among 30 candidates, SORT1 (encoding sortilin) was one of the targets, not yet under investigation, with the most potential that is highly pleiotropic but with minimal off-target adverse effects. Other targets such as PSRC1, GRN, HFE2, and FJX1 also demonstrated a wide range of health benefits with only a few anticipated adverse effects (Figure 5). Our integrated approach provided the landscape of how genetic-driven data could be integrated and advance the life cycle of drug development.

(C) Phenotype classifications. Phenotypes with increased risk due to reduction of LDL (4 phenotypes) and TG levels (2 phenotypes) (mimic lipid-lowering agents) could be noted as potential adverse effects that are associated with lipid-modifying agents. Phenotypes with reduced risk due to reduction of levels of LDL (15 phenotypes), TG (9 phenotypes), and multiple lipids (11 phenotypes) may benefit from repurposing of lipid-modifying agents. See also Figures S1A–S1C and Data S1: Tables S3–S10.



(legend on next page)

Approved or investigational lipid modifying drug targets
 *GWAS genes
 **Pleiotropic causal genes (MR)
 Top tier genes with associated pQTL and protein–protein interactions (network connectivity)
 Final prioritized genes after scoring

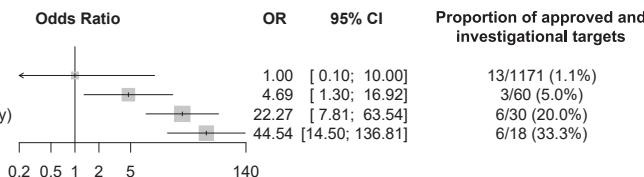


Figure 4. Target-indication pairs that entered clinical trial phase or were approved

*Genes significant in GWAS of lipids. **Genes causally associated with multiple lipids and more than half of lipid-related phenotypes (associated with ≥ 6 category 4 phenotypes). Top-tier genes and their network connectivity are provided in [Data S1 \(Table S19\)](#). A list of the final 30 genes is provided in [Table 1](#); 16 genes scored 4 or higher. The ongoing drug trial evidence (phase 1–3) was obtained from the Pharmaprojects database, and drug approval evidence was obtained from WHO Collaborating Center for Drug Statistics Methodology (WHOCC), DrugBank, and ChEMBL databases. MR, Mendelian randomization; OR, odds ratio.

See also [Data S1: Table S20](#).

The current drug development pipeline takes approximately 10 years and \$1 billion USD for a single drug to proceed from trial entry to market authorization.^{3,20} After pre-clinical investigation, the long enduring clinical trials frequently fail because of the poor translation of putative drug targets identified in animal models to the human species.²¹ As such, the conventional drug development process beginning from hypothesis-driven pre-clinical research is being challenged by the emerging paradigm of hypothesis-free, genetic-driven analysis using human genetics.²¹ The discovery of PCSK9, an important target of CVD drugs, is a result of this paradigm centered on human genetics.^{9,22} Our approach aligns with this shift, where we began from more than 15,000 human genes in a hypothesis-free manner and thus minimized the risk of failure due to disparity between animal models and humans.

The causation supported by genetic evidence (i.e., MR) increases the probability of a therapeutic agent being approved several-fold.^{3,17} For the success of the drug, Holmes et al. introduced potential roles of human genetics in the life cycle of drug development; human genetics could assist in drug target prioritization, drug repurposing, and identification of target-mediated or off-target adverse effects.³ This study could be exemplary for the comprehensive implementation of human genetics in the drug development pipeline, as we assessed all three components, i.e., target prioritization, repurposing, and safety consideration, and integrated those findings ([Figure 5](#)). Our phase 1 analysis ([Figure 1](#)) enabled the categorization of 62 lipid-driven phenotypes ([Figure 2C](#)), informed repurposing opportunities, and anticipated adverse effects of lipid-lowering agents. For example, we found that LDL cholesterol-lowering agents could be associated with an increased risk of cholelithiasis (adverse ef-

fect), whereas they could be repurposed to reduce the risk of dementia ([Figures 1](#) and [2C](#)).

We identified *SORT1*, *CELSR2*, and *PSRC1* as top-tier genes that could be translated to therapeutic targets because they have higher degree of pleiotropy, affecting all three lipid perturbations and the risks of 10 CVDs. The three genes are located in the same LD block²³ and demonstrated similar MR estimates across diverse outcomes ([Figure 3D](#)). Although only *SORT1* was mainly associated with both eQTLs and pQTLs among the three top-tier genes ([Figure 3E](#)), we could not confidently isolate the main effector, having three genes collectively explored as a potential therapeutic target. Among the three genes, *SORT1* has been most extensively investigated in experimental studies; Musunuru et al. experimentally validated that a polymorphism at the 1p13 locus (rs12740374) creates a CCAAT/enhancer-binding protein (C/EBP) transcription factor binding site and increases the hepatic expression of the *SORT1* gene.¹⁹ The increased liver *SORT1* expression was associated with decreased LDL level and myocardial infarction (MI) in mice,¹⁹ which is concordant with our MR results ([Figures 3D](#) and [5](#)). To screen unintended safety issues with *SORT1*, we performed PheWAS for rs12740374 (effect allele T), and it was determined to be safe while effectively reducing various CVD risks, such as MI, abdominal aortic aneurysm, and peripheral artery disease ([Figures 3D](#) and [S4](#)). Our genetic-driven approach supports the concept that *SORT1* could be a target for therapeutic interventions, possibly using base editing, and motivates further translational research for sortilin.

Numerous agents inhibiting PCSK9 and APOB, such as inclisiran, evolocumab, and mipomersen, exhibited promising lipid-lowering effects and improved profiles of diverse CVD outcomes with tolerable safety in clinical trials.^{2,24–27} Our findings align with

Figure 3. Phase 2: Genome-level analysis for drug target prioritization

- (A) Workflow of drug target prioritization.
- (B) Steps 1–3 of the workflow. Manhattan plot represents transcriptome-wide MR of LDL. Genes from five tissues (whole blood, liver, artery, subcutaneous adipose, and visceral adipose) that had valid eQTL instrument variables ($<5E-8$) were tested for 3 lipids and 11 phenotypes (category 4). The dotted line indicates a suggestive threshold ($<1E-5$), and the plain line indicates a Bonferroni threshold ($1.24E-6$, $<0.05/40,302$). Manhattan plots for HDL and TG are provided in the supplemental information.
- (C) Step 4 of the workflow. Three strong pleiotropic genes that are causally associated with multiple lipids and a maximum number of category 4 phenotypes (10 phenotypes) were termed top-tier genes (*CELSR2*, *PSRC1*, and *SORT1*). Their top SNPs were used to test association with proteome (pQTL and protein–protein interactions).
- (D) Step 4 of the workflow. MR estimates top-tier genes with lipids and category 4 phenotypes (inform direction of effect).
- (E) Step 5 of the workflow. Blue lines indicate strong association with eQTLs of top-tier genes ($<5E-8$ for gene expression [eQTL] and $<1E-5$ for protein [pQTL]). rs12740374 (effect allele: T) was associated with both *SORT1* gene expression and *SORT1* protein (sortilin1), indicating the robust association of *SORT1* with multiple lipids and lipid-associated outcomes.

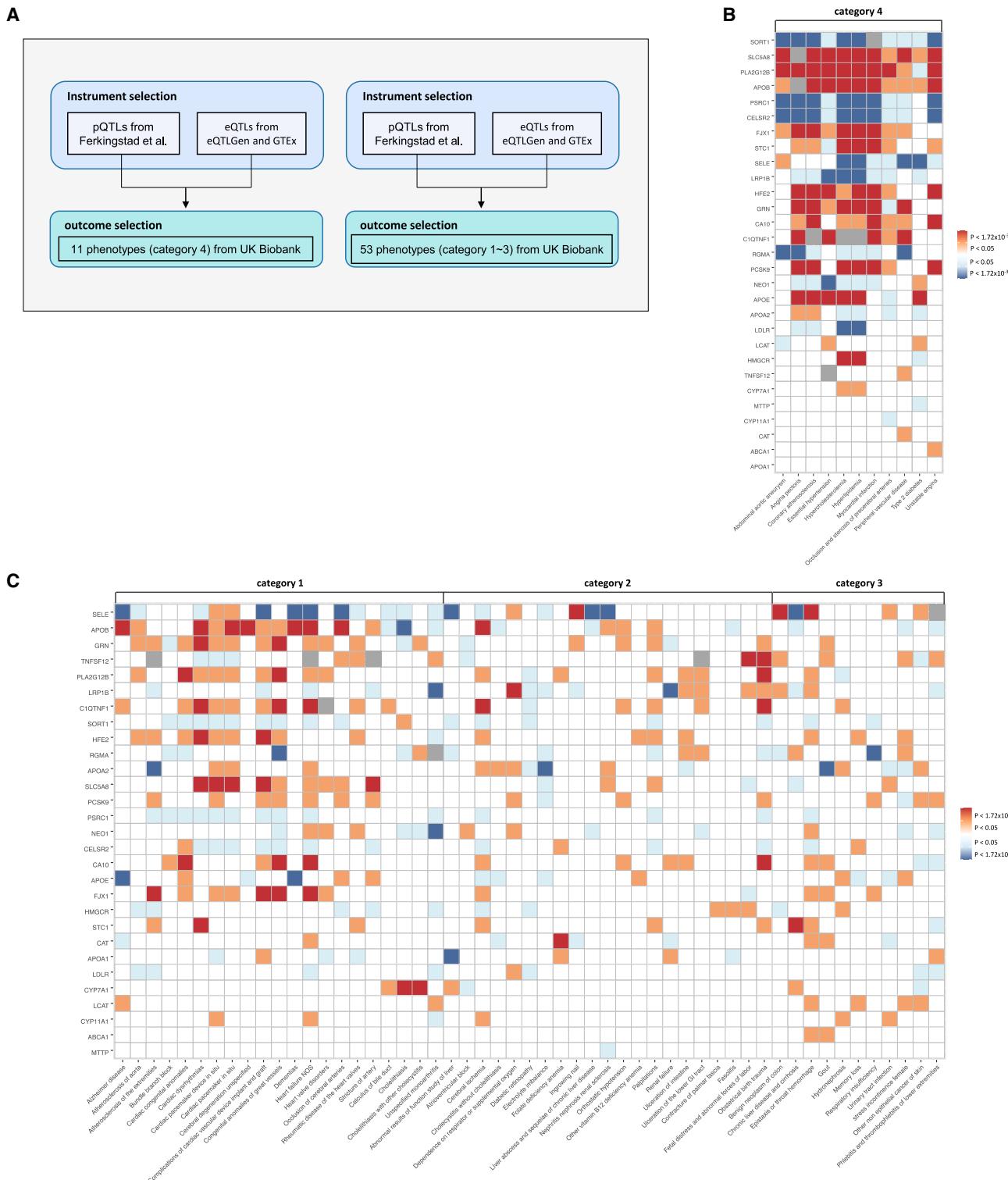


Figure 5. Causal association of prioritized candidates with category 1–4 phenotypes identifies repurposing opportunities and off-target effects of potential therapeutics

(A) Summary diagram for discovery set.

(B) The heatmap shows potential causal relationship between 29 therapeutic candidates (one target [SREBF2] lacks eQTL/pQTL data) and category 4 phenotypes from eQTL and pQTL MR and suggests the causal effect of prioritized targets on dyslipidemia and lipid-driven cardiometabolic outcomes. No association

(legend continued on next page)

trial results, as the reduction in PCSK9 and APOB expression and/or protein levels was shown to be causally associated with the decreased risk of dyslipidemia, lipid-driven cardiometabolic traits, and numerous other disorders while demonstrating a few off-target adverse outcomes (Figure 5). Any agent inhibiting PCSK9 would potentially increase the risk of electrolyte imbalance and stress incontinence, while agents antagonizing APOB may increase the risk of cholelithiasis, chronic liver diseases, and fasciitis (Figure 5B). However, these adverse events are mostly suggestive, as a majority of such associations did not reach a Bonferroni threshold. Therefore, a careful examination of unwanted off-target effects, described in this article, would still be warranted to mitigate risks and enrich benefits. We also identified signals that have not been examined in lipid GWASs, including but not limited to GRN, HFE2, FJX1, C1QTNF1, and SLC5A8, which have shown consistent results during *in silico* validation. A wide range of health benefits with few adverse effects were observed with these targets (Figure 5). Not enough data supporting the association of such targets with dyslipidemia have been reported (Table 1); therefore, further translational studies are required to validate their dyslipidemia-protective roles.

Our study has several strengths. First, we conducted transcriptome-wide ($>15,000$ genes) eQTL-based MR and colocalization for multiple tissues to identify causal genes underlying diverse lipid profiles and lipid-related traits. Such an approach overcame prior limitations of studies focusing only on druggable molecules ($\sim 1,300$ genes/proteins)^{9,17} and further elucidated the causation (direction of effect), which is crucial for therapeutic development. Second, we integrated DNA sequence variation, transcriptome, proteome, and epigenome data to improve resolution for causal gene detection and evaluation for priority among potential targets.^{3,28} Third, we conceptually combined the strength of multi-trait GWASs (increase power to detect important variants)¹¹ and MR (causation); we conducted transcriptome-wide MR for multi-trait to insulate pleiotropic causal genes that affect multiple related traits. Identifying pleiotropic variants/genes for highly selected multi-trait is useful in searching for lipid-lowering targets because the lipid-lowering efforts primarily aim to prevent the risk of lipid-mediated diseases, particularly CVDs, rather than solely reducing lipid levels.² Fourth, we systematically analyzed the phenotypes associated with levels of distinct lipid species to isolate individual lipid effects from complex lipid coperturbations. The obtained results are provided in Figure 2C and Data S1 (Table S10, related to Figure 2). These findings may guide clinicians to understand how

targeted therapy for a specific lipid would affect patients. Lastly, we demonstrated an all-inclusive genetic pipeline that can simultaneously inform drug target prioritization, repurposing, and adverse effects for a given trait.

Limitations of the study

The current study had some limitations. First, although we prioritized 30 drug targets, this does not necessarily mean that all these genes are druggable. Prioritization is needed to narrow the scope of searching for drug targets (starting with approximately 15,000–20,000 genes), focus the resources and research efforts, and minimize the time and cost risks for successful drug development. Therefore, further research efforts to validate these candidates should be followed. Second, some approved dyslipidemia drug targets are not included in our prioritized candidate genes, probably owing to a lack of eQTL and pQTL data incorporating rare variants and their suboptimal tissue-/cell-type diversity. The performance of our pipeline will likely be enhanced with the advent of larger eQTL/pQTL datasets that include rare variants at diverse tissues. Third, the heatmap plot (Figures 5 and 6) should be interpreted with caution. No association (white box) shown in the heatmap does not necessarily preclude genuine effects between studied drug targets, and outcomes because the analyses were based exclusively on five tissues (whole blood, subcutaneous adipose, visceral adipose, artery, and liver). The small sample size from non-blood tissues in GTEx may have reduced the statistical power to detect the true effect. In addition, the association should be interpreted specific to a tissue and eQTL/pQTL (Data S1: Tables S24–S31, related to Figures 5 and 6). Lastly, although several MR tools and methods were used to account for the horizontal pleiotropy in our analysis, they may be insufficient to exclude that the effects of the genetic variants on the outcome are not exclusively through the exposure.¹²

Conclusion

This study suggests how human genetic data can be integrated into the drug development pipeline to enhance the likelihood of a clinical trial's success. The phenotype-wide approach (phase 1) helped understand the therapeutic landscape across repurposing and adverse effects, and the transcriptome-wide approach (phase 2) prioritized potential drug targets. Integration of inclusive pipelines (prioritization, repurposing, and safety profiling) identified pleiotropically beneficial targets with minimal adverse effects. We demonstrated that exploring pleiotropic variants/genes for highly selected multi-trait using MR and

shown in the heatmap (white box) does not necessarily preclude the genuine effect of drug candidate because the analyses were based exclusively on five tissues (whole blood, subcutaneous adipose, visceral adipose, artery, and liver). Small sample size of non-blood tissues in GTEx may have reduced statistical power to detect the true effect or the effect could exist in unexplored tissues.

(C) The heatmap shows the potential causal relationships between 29 therapeutic candidates and category 1–3 phenotypes from eQTL and pQTL MR. (B and C) The orange color indicates that the therapeutic candidate has a positive causal association with a disease (increased level of RNA or protein of a target gene increases the risk of the corresponding outcome) at a nominal significance ($p < 0.05$); the red color indicates that the therapeutic candidate has a positive causal association with a disease at $p < 1.72E-3$ ($p = 0.05/29$; Bonferroni's correction for candidates); the light blue color indicates that the therapeutic candidate has a negative causal association with a disease at a nominal significance ($p < 0.05$); and the blue color indicates that the therapeutic candidate has a negative causal association with a disease at $p < 1.72E-3$ ($p = 0.05/29$; Bonferroni's correction for candidates). The target-outcome pair showing contradictory results between eQTL and pQTL or across tissues are marked in gray. Tissue-specific results are found in the supplemental information.

See also Data S1: Tables S24–S31.

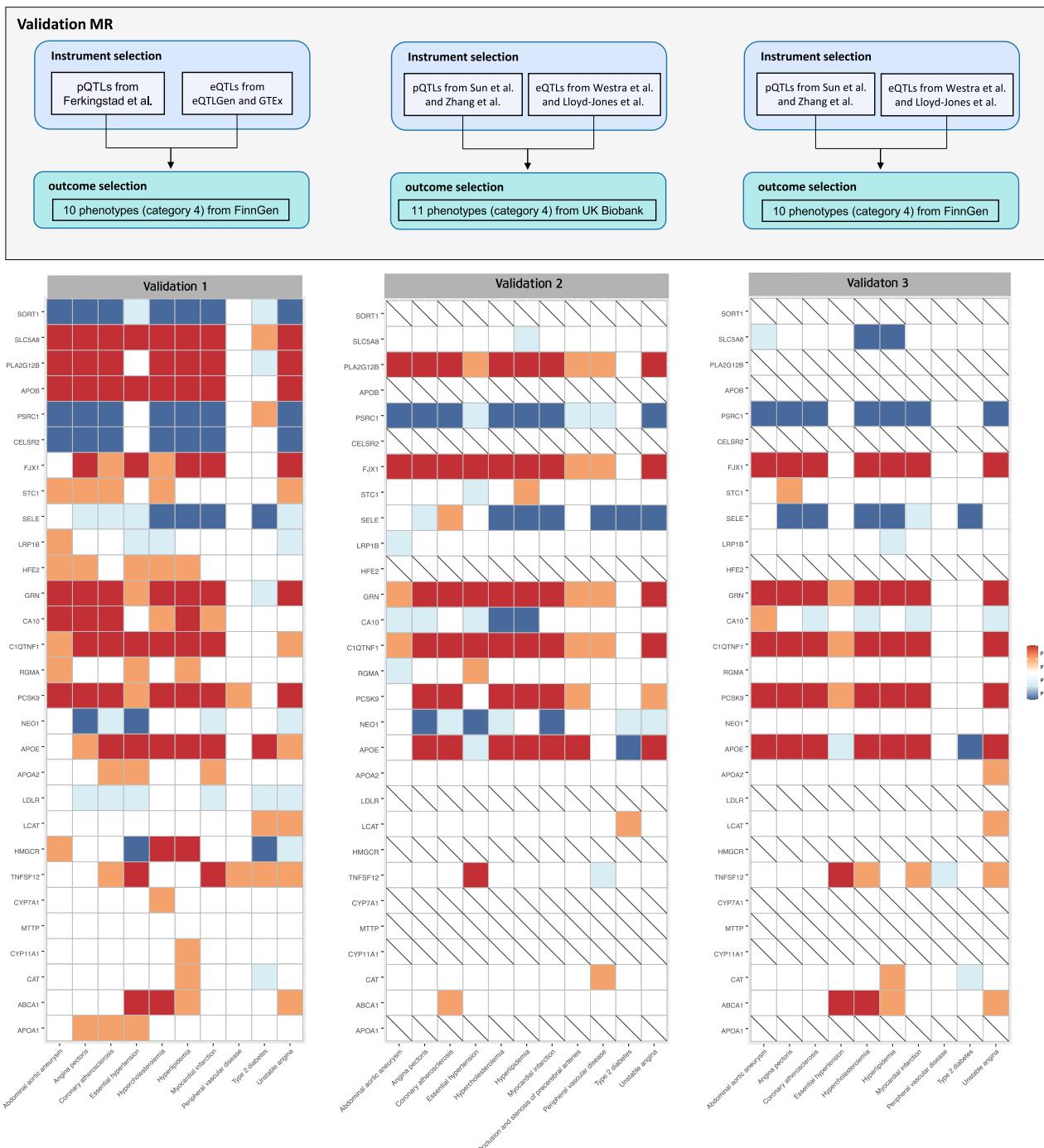


Figure 6. In silico validation for prioritized candidates

The heatmap shows potential causal relationships between 29 therapeutic candidates (one target [SREBF2] lacks eQTL/pQTL data) and category 4 phenotypes. Different sets of exposure and outcome data were used to validate findings from a discovery set. Validations 1 and 2 used different outcome (FinnGen) and exposure (eQTL/pQTL) datasets, respectively. Conversely, validation 3 utilized different exposure and outcome sets from those used for discovery. Owing to the utilization of only five tissue types, the absence of association in the heatmap (white box) does not necessarily preclude the genuine effect of drug candidates. In addition, the small sample size of non-blood tissues in GTEx may have reduced the statistical power to detect the true effect, which may exist in unexplored tissues. Here, stripes were used to mark therapeutic candidates for which results did not exist due to a lack of eQTL/pQTL data. Orange indicates that the therapeutic candidate has a positive causal association with a disease (increased RNA or protein levels of a target gene increase the risk of the corresponding outcome) at $p < 0.05$; red indicates that the therapeutic candidate has a positive causal association with a disease at $p < 1.72 \times 10^{-3}$ ($p = 0.05/29$; Bonferroni's correction for candidates); light blue indicates that the therapeutic candidate has a negative causal association with a disease ($p < 0.05$); and dark blue indicates

(legend continued on next page)

colocalization, followed by an exploration of network connectivity, is a promising strategy for prioritizing therapeutic targets. The drug targets prioritized using this strategy were 22-fold higher in the likelihood of approval than GWAS-curated targets.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **RESOURCE AVAILABILITY**
 - Lead contact
 - Materials availability
 - Data and code availability
- **EXPERIMENTAL MODEL AND SUBJECT DETAILS**
 - Study population
- **METHOD DETAILS**
 - Study design
 - Polygenic risk scores
 - Phenome-wide associations
 - Mendelian randomization
 - Transcriptome-wide Mendelian randomization
 - Identification of top tier pleiotropic genes
 - Proteome scan and protein-protein interactions
 - Phenome-wide MR analysis of category 1–4 phenotypes
 - In silico validation
- **QUANTIFICATION AND STATISTICAL ANALYSIS**

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.xcrm.2023.101112>.

ACKNOWLEDGMENTS

This study was supported by a National Research Foundation of Korea (NRF) grant funded by the Ministry of Science and Information and Communication Technologies, South Korea (2022R1A2C2009998), and a grant from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (HJ22C0042). R.D. is supported by the National Institute of General Medical Sciences of the NIH (R35-GM124836) and the National Heart, Lung, and Blood Institute of the NIH (R01-HL139865 and R01-HL155915). P.N. is supported by grants from the National Heart, Lung, and Blood Institute (R01HL142711, R01HL127564, R01HL148050, R01HL151283, R01HL148565, R01HL135242, and R01 HL151152), the National Institute of Diabetes and Digestive and Kidney Diseases (R01DK125782), the National Human Genome Research Institute (U01HG011719), the Fondation Leducq (TNE-18CVD04), and Massachusetts General Hospital (Paul and Phyllis Fireman Endowed Chair in Vascular Medicine). H.-H.W. had full access to all data in the study. H.-H.W. made the final decision to submit the manuscript for publication.

that the therapeutic candidate has a negative causal association with a disease at $p < 1.72E-3$ ($p = 0.05/29$; Bonferroni's correction for candidates). The target-outcome pairs showing contradictory results between eQTL and pQTL or across tissues are marked by gray squares. Tissue-specific results are found in the supplemental information.

See also Data S1: Tables S32–S37.

AUTHOR CONTRIBUTIONS

Conceptualization, M.S.K., M.S., and H.-H.W.; formal analysis, M.S. and M.S.K.; resources, B.K., I.S., and H.-H.W.; writing – original draft, M.S.K. and M.S.; writing – review & editing, M.S.K., M.S., P.N., R.D., and H.-H.W.; funding acquisition, H.-H.W.; supervision, H.-H.W.; all authors approved the final version of the manuscript.

DECLARATION OF INTERESTS

P.N. reports investigator-initiated grants from Amgen, Apple, AstraZeneca, Boston Scientific, and Novartis; personal fees from Apple, AstraZeneca, Blackstone Life Sciences, Foresite Labs, Novartis, and Roche/Genentech; being a cofounder of TenSixteen Bio; being a scientific advisory board member of Esperion Therapeutics, geneXwell, and TenSixteen Bio; and having spousal employment at Vertex, all unrelated to the present work. R.D. reports receiving grants from AstraZeneca; grants and non-financial support from Goldfinch Bio; being a scientific cofounder and equity holder for Pensieve Health (pending); and being a consultant for Variant Bio, all unrelated to this work.

Received: November 17, 2022

Revised: December 22, 2022

Accepted: June 19, 2023

Published: August 14, 2023

REFERENCES

1. Pirillo, A., Casula, M., Olmastroni, E., Norata, G.D., and Catapano, A.L. (2021). Global epidemiology of dyslipidaemias. *Nat. Rev. Cardiol.* 18, 689–700. <https://doi.org/10.1038/s41569-021-00541-4>.
2. Ray, K.K., Corral, P., Morales, E., and Nicholls, S.J. (2019). Pharmacological lipid-modification therapies for prevention of ischaemic heart disease: current and future options. *Lancet* 394, 697–708. [https://doi.org/10.1016/S0140-6736\(19\)31950-6](https://doi.org/10.1016/S0140-6736(19)31950-6).
3. Holmes, M.V., Richardson, T.G., Ference, B.A., Davies, N.M., and Davey Smith, G. (2021). Integrating genomics with biomarkers and therapeutic targets to invigorate cardiovascular drug development. *Nat. Rev. Cardiol.* 18, 435–453. <https://doi.org/10.1038/s41569-020-00493-1>.
4. Dowden, H., and Munro, J. (2019). Trends in clinical success rates and therapeutic focus. *Nat. Rev. Drug Discov.* 18, 495–496. <https://doi.org/10.1038/d41573-019-00074-z>.
5. Harrison, R.K. (2016). Phase II and phase III failures: 2013–2015. *Nat. Rev. Drug Discov.* 15, 817–818. <https://doi.org/10.1038/nrd.2016.184>.
6. Gill, D., Georgakis, M.K., Walker, V.M., Schmidt, A.F., Gkatzionis, A., Freitag, D.F., Finan, C., Hingorani, A.D., Howson, J.M.M., Burgess, S., and Davies, N.M. (2021). Mendelian randomization for studying the effects of perturbing drug targets. *Wellcome Open Res.* 6, 16. <https://doi.org/10.12688/wellcomeopenres.16544.2>.
7. Storm, C.S., Kia, D.A., Almramhi, M.M., Bandres-Ciga, S., Finan, C., International Parkinson's Disease Genomics Consortium IPDG; Hingorani, A.D., and Wood, N.W. (2021). Finding genetically-supported drug targets for Parkinson's disease using Mendelian randomization of the druggable genome. *Nat. Commun.* 12, 7342. <https://doi.org/10.1038/s41467-021-26280-1>.
8. Kim, M.S., Song, M., Shin, J.I., and Won, H.H. (2023). How to interpret studies using Mendelian randomisation. *BMJ*. <https://doi.org/10.1136/bmjebm-2022-112149>.
9. Gaziano, L., Giambartolomei, C., Pereira, A.C., Gaulton, A., Posner, D.C., Swanson, S.A., Ho, Y.L., Iyengar, S.K., Kosik, N.M., Vujkovic, M., et al.; VA

- Million Veteran Program; COVID-19 Science Initiative (2021). Actionable druggable genome-wide Mendelian randomization identifies repurposing opportunities for COVID-19. *Nat. Med.* 27, 668–676. <https://doi.org/10.1038/s41591-021-01310-z>.
10. Julianne, H., Laville, V., McCaw, Z.R., He, Z., Guillemot, V., Lasry, C., Ziyatdinov, A., Nerin, C., Vaysse, A., Lechat, P., and Aschard, H. (2021). Multitrait GWAS to connect disease variants and biological mechanisms. *PLoS Genet.* 17, e1009713. <https://doi.org/10.1371/journal.pgen.1009713>.
 11. Turley, P., Walters, R.K., Maghzian, O., Okbay, A., Lee, J.J., Fontana, M.A., Nguyen-Viet, T.A., Wedow, R., Zacher, M., Furlotte, N.A., and Benjamin, D.J. (2018). Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nat. Genet.* 50, 229–237. <https://doi.org/10.1038/s41588-017-0009-4>.
 12. Davies, N.M., Holmes, M.V., and Davey Smith, G. (2018). Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. *BMJ* 362, k601. <https://doi.org/10.1136/bmj.k601>.
 13. Hartwig, F.P., Davey Smith, G., and Bowden, J. (2017). Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. *Int. J. Epidemiol.* 46, 1985–1998. <https://doi.org/10.1093/ije/dyx102>.
 14. Bowden, J., Davey Smith, G., Haycock, P.C., and Burgess, S. (2016). Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genet. Epidemiol.* 40, 304–314. <https://doi.org/10.1002/gepi.21965>.
 15. Bowden, J., Davey Smith, G., and Burgess, S. (2015). Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int. J. Epidemiol.* 44, 512–525. <https://doi.org/10.1093/ije/dyv080>.
 16. Verbanck, M., Chen, C.Y., Neale, B., and Do, R. (2018). Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat. Genet.* 50, 693–698. <https://doi.org/10.1038/s41588-018-0099-7>.
 17. Zheng, J., Haberland, V., Baird, D., Walker, V., Haycock, P.C., Hurle, M.R., Gutteridge, A., Erola, P., Liu, Y., Luo, S., and Gaunt, T.R. (2020). Phenome-wide Mendelian randomization mapping the influence of the plasma proteome on complex diseases. *Nat. Genet.* 52, 1122–1131. <https://doi.org/10.1038/s41588-020-0682-6>.
 18. Wouters, O.J., McKee, M., and Luyten, J. (2020). Estimated research and development investment needed to bring a new medicine to market, 2009–2018. *JAMA* 323, 844–853. <https://doi.org/10.1001/jama.2020.1166>.
 19. Van Der Laan, S.W., Harshfield, E.L., Hemerich, D., Stacey, D., Wood, A.M., and Asselbergs, F.W. (2018). From lipid locus to drug target through human genomics. *Cardiovasc. Res.* 114, 1258–1270. <https://doi.org/10.1093/cvr/cvy120>.
 20. Cohen, J.C., Boerwinkle, E., Mosley, T.H., Jr., and Hobbs, H.H. (2006). Sequence variations in PCSK9, low LDL, and protection against coronary heart disease. *N. Engl. J. Med.* 354, 1264–1272. <https://doi.org/10.1056/NEJMoa054013>.
 21. Kjolby, M., Andersen, O.M., Breiderhoff, T., Fjorback, A.W., Pedersen, K.M., Madsen, P., Jansen, P., Heeren, J., Willnow, T.E., and Nykjaer, A. (2010). Sort1, encoded by the cardiovascular risk locus 1p13.3, is a regulator of hepatic lipoprotein export. *Cell Metab.* 12, 213–223. <https://doi.org/10.1016/j.cmet.2010.08.006>.
 22. Ray, K.K., Kallend, D., Leiter, L.A., Raal, F.J., Koenig, W., Jaros, M.J., Schwartz, G.G., Landmesser, U., Garcia Conde, L., and Wright, R.S.; ORION-11 Investigators (2022). Effect of inclisiran on lipids in primary prevention: the ORION-11 trial. *Eur. Heart J.* 43, 5047–5057. <https://doi.org/10.1093/eurheartj/ehac615>.
 23. Raal, F.J., Kallend, D., Ray, K.K., Turner, T., Koenig, W., Wright, R.S., Wijngaard, P.L.J., Curcio, D., Jaros, M.J., Leiter, L.A., et al. (2020). Inclisiran for the treatment of heterozygous familial hypercholesterolemia. *N. Engl. J. Med.* 382, 1520–1530. <https://doi.org/10.1056/NEJMoa1913805>.
 24. Guedeney, P., Giustino, G., Sorrentino, S., Claessen, B.E., Camaj, A., Kalkman, D.N., Vogel, B., Sartori, S., De Rosa, S., Baber, U., et al. (2019). Efficacy and safety of alirocumab and evolocumab: a systematic review and meta-analysis of randomized controlled trials. *Eur. Heart J.* 43, e17-e25. <https://doi.org/10.1093/eurheartj/ehz430>.
 25. Raal, F.J., Santos, R.D., Blom, D.J., Marais, A.D., Charng, M.J., Cromwell, W.C., Lachmann, R.H., Gaudet, D., Tan, J.L., Chasan-Taber, S., and Crooke, S.T. (2010). Mipomersen, an apolipoprotein B synthesis inhibitor, for lowering of LDL cholesterol concentrations in patients with homozygous familial hypercholesterolemia: a randomised, double-blind, placebo-controlled trial. *Lancet* 375, 998–1006. [https://doi.org/10.1016/S0140-6736\(10\)60284-X](https://doi.org/10.1016/S0140-6736(10)60284-X).
 26. Robinson, J.W., Battram, T., Baird, D.A., Haycock, P., Zheng, J., Hemani, G., Chen, C.-Y., and Gaunt, T.R. (2022). Evaluating the potential benefits and pitfalls of combining protein and expression quantitative trait loci in evidencing drug targets. Preprint at bioRxiv. <https://doi.org/10.1101/2022.03.15.484248>.
 27. Sun, B.B., Maranville, J.C., Peters, J.E., Stacey, D., Staley, J.R., Blackshaw, J., Burgess, S., Jiang, T., Paige, E., Surendran, P., and Butterworth, A.S. (2018). Genomic atlas of the human plasma proteome. *Nature* 558, 73–79. <https://doi.org/10.1038/s41586-018-0175-2>.
 28. Zhang, J., Dutta, D., Köttgen, A., Tin, A., Schlosser, P., Grams, M.E., Harvey, B., CKDGen Consortium; Yu, B., Boerwinkle, E., and Chatterjee, N. (2022). Plasma proteome analyses in individuals of European and African ancestry identify cis-pQTLs and models for proteome-wide association studies. *Nat. Genet.* 54, 593–602. <https://doi.org/10.1038/s41588-022-01051-w>.
 29. Ge, T., Chen, C.Y., Ni, Y., Feng, Y.C.A., and Smoller, J.W. (2019). Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat. Commun.* 10, 1776. <https://doi.org/10.1038/s41467-019-09718-5>.
 30. Zhu, Z., Zhang, F., Hu, H., Bakshi, A., Robinson, M.R., Powell, J.E., Montgomery, G.W., Goddard, M.E., Wray, N.R., Visscher, P.M., and Yang, J. (2016). Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* 48, 481–487. <https://doi.org/10.1038/ng.3538>.
 31. Finucane, H.K., Bulik-Sullivan, B., Gusev, A., Trynka, G., Reshef, Y., Loh, P.R., Anttila, V., Xu, H., Zang, C., Farh, K., and Price, A.L. (2015). Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* 47, 1228–1235. <https://doi.org/10.1038/ng.3404>.
 32. Staley, J.R., Blackshaw, J., Kamat, M.A., Ellis, S., Surendran, P., Sun, B.B., Paul, D.S., Freitag, D., Burgess, S., Danesh, J., and Butterworth, A.S. (2016). PhenoScanner: a database of human genotype–phenotype associations. *Bioinformatics* 32, 3207–3209. <https://doi.org/10.1093/bioinformatics/btw373>.
 33. Kuhn, M., Szklarczyk, D., Pletscher-Frankild, S., Blicher, T.H., Von Mering, C., Jensen, L.J., and Bork, P. (2014). STITCH 4: integration of protein–chemical interactions with user data. *Nucleic Acids Res.* 42, D401–D407. <https://doi.org/10.1093/nar/gkt1207>.
 34. Cao, C., Wang, J., Kwok, D., Cui, F., Zhang, Z., Zhao, D., Li, M.J., and Zou, Q. (2022). webTWAS: a resource for disease candidate susceptibility genes identified by transcriptome-wide association study. *Nucleic Acids Res.* 50, D1123–D1130. <https://doi.org/10.1093/nar/gkab957>.
 35. Musunuru, K., Strong, A., Frank-Kamenetsky, M., Lee, N.E., Ahfeldt, T., Sachs, K.V., Li, X., Li, H., Kuperwasser, N., Ruda, V.M., and Rader, D.J. (2010). From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. *Nature* 466, 714–719. <https://doi.org/10.1038/nature09266>.
 36. Ochoa, D., Hercules, A., Carmona, M., Suveges, D., Gonzalez-Uriarte, A., Malangone, C., Miranda, A., Fumis, L., Carvalho-Silva, D., Spitzer, M., and McDonagh, E.M. (2021). Open Targets Platform: supporting systematic drug–target identification and prioritisation. *Nucleic Acids Res.* 49, D1302–D1310. <https://doi.org/10.1093/nar/gkaa1027>.
 37. Buniello, A., MacArthur, J.A.L., Cerezo, M., Harris, L.W., Hayhurst, J., Malangone, C., McMahon, A., Morales, J., Mountjoy, E., Sollis, E., and Parkinson, H. (2019). The NHGRI-EBI GWAS Catalog of published genome-wide

- association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* 47, D1005-D1012. <https://doi.org/10.1093/nar/gky1120>.
38. Li, M., Zou, D., Li, Z., Gao, R., Sang, J., Zhang, Y., Li, R., Xia, L., Zhang, T., Niu, G., and Zhang, Z. (2019). EWAS Atlas: a curated knowledgebase of epigenome-wide association studies. *Nucleic Acids Res.* 47, D983-D988. <https://doi.org/10.1093/nar/gky1027>.
 39. Battram, T., Yousefi, P., Crawford, G., Prince, C., Sheikhalibabaei, M., Sharp, G., Hatcher, C., Vega-Salas, M.J., Khodabakhsh, S., Whitehurst, O., and Suderman, M. (2022). The EWAS Catalog: a database of epigenome-wide association studies. *Wellcome Open Res.* 7, 41. <https://doi.org/10.12688/wellcomeopenres.17598.2>.
 40. Legault, M.A., Perreault, L.P.L., Tardif, J.C., and Dubé, M.P. (2022). ExPheWas: a platform for cis-Mendelian randomization and gene-based association scans. *Nucleic Acids Res.* 50, W305-W311. <https://doi.org/10.1093/nar/gkac289>.
 41. Watanabe, K., Taskesen, E., Van Bochoven, A., and Posthuma, D. (2017). Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* 8, 1826. <https://doi.org/10.1038/s41467-017-01261-5>.
 42. Barbeira, A.N., Dickinson, S.P., Bonazzola, R., Zheng, J., Wheeler, H.E., Torres, J.M., Torstenson, E.S., Shah, K.P., Garcia, T., Edwards, T.L., et al. (2018). Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *Nat. Commun.* 9, 1825. <https://doi.org/10.1038/s41467-018-03621-1>.
 43. Manichaikul, A., Mychaleckyj, J.C., Rich, S.S., Daly, K., Sale, M., and Chen, W.M. (2010). Robust relationship inference in genome-wide association studies. *Bioinformatics* 26, 2867–2873. <https://doi.org/10.1093/bioinformatics/btq559>.
 44. Kurki, M.I., Karjalainen, J., Palta, P., Sipilä, T.P., Kristiansson, K., Donner, K.M., Reeve, M.P., Laivuori, H., Aavikko, M., Kaunisto, M.A., et al. (2023). FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature* 613, 508–518. <https://doi.org/10.1038/s41586-022-05473-8>.
 45. Graham, S.E., Clarke, S.L., Wu, K.H.H., Kanoni, S., Zajac, G.J.M., Ramdas, S., Surakka, I., Ntalla, I., Vedantam, S., Winkler, T.W., et al. (2021). The power of genetic diversity in genome-wide association studies of lipids. *Nature* 600, 675–679. <https://doi.org/10.1038/s41586-021-04064-3>.
 46. Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L.T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., and Marchini, J. (2018). The UK Biobank resource with deep phenotyping and genomic data. *Nature* 562, 203–209. <https://doi.org/10.1038/s41586-018-0579-z>.
 47. Denny, J.C., Bastarache, L., Ritchie, M.D., Carroll, R.J., Zink, R., Mosley, J.D., Field, J.R., Pulley, J.M., Ramirez, A.H., Bowton, E., and Roden, D.M. (2013). Systematic comparison of genome-wide association study of electronic medical record data and genome-wide association study data. *Nat. Biotechnol.* 31, 1102–1110. <https://doi.org/10.1038/nbt.2749>.
 48. Verma, A., Bradford, Y., Dudek, S., Lucas, A.M., Verma, S.S., Pendergrass, S.A., and Ritchie, M.D. (2018). A simulation study investigating power estimates in genome-wide association studies. *BMC Bioinformatics* 19, 120. <https://doi.org/10.1186/s12859-018-2135-0>.
 49. Puhr, R., Heinze, G., Nold, M., Lusa, L., and Geroldinger, A. (2017). Firth's logistic regression with rare events: accurate effect estimates and predictions? *Stat. Med.* 36, 2302–2317. <https://doi.org/10.1002/sim.7273>.
 50. Bush, W.S., Oetjens, M.T., and Crawford, D.C. (2016). Unravelling the human genome–phenome relationship using genome-wide association studies. *Nat. Rev. Genet.* 17, 129–145. <https://doi.org/10.1038/nrg.2015.36>.
 51. Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Stat. Soc. B* 57, 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>.
 52. Hemani, G., Bowden, J., and Davey Smith, G. (2018). Evaluating the potential role of pleiotropy in Mendelian randomization studies. *Hum. Mol. Genet.* 27, R195-R208. <https://doi.org/10.1093/hmg/ddy163>.
 53. Fang, H., ULTRA-DD Consortium; De Wolf, H., Knezevic, B., Burnham, K.L., Osgood, J., Sanniti, A., Lledó Lara, A., Kasela, S., De Cesco, S., and Knight, J.C. (2019). A genetics-led approach defines the drug target landscape of 30 immune-related traits. *Nat. Genet.* 51, 1082–1091. <https://doi.org/10.1038/s41588-019-0456-1>.
 54. Okada, Y., Wu, D., Trynka, G., Raj, T., Terao, C., Ikari, K., Kochi, Y., Ohmura, K., Suzuki, A., Yoshida, S., and Plenge, R.M. (2014). Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* 506, 376–381. <https://doi.org/10.1038/nature12873>.
 55. Ferkingstad, E., Sulem, P., Atlason, B.A., Sveinbjornsson, G., Magnusson, M.I., Styrmsdottir, E.L., Gunnarsdottir, K., Helgason, A., Oddsson, A., Halldorsson, B.V., and Stefansson, K. (2021). Large-scale integration of the plasma proteome with genetics and disease. *Nat. Genet.* 53, 1712–1721. <https://doi.org/10.1038/s41588-021-00978-w>.
 56. Kim, M.S., Song, M., Kim, S., Kim, B., Kang, W., Kim, J.Y., Myung, W., Lee, I., Do, R., Khera, A.V., and Won, H.H. (2023). Causal effect of adiposity on the risk of 19 gastrointestinal diseases: a Mendelian randomization study. *Obesity* 31, 1436–1444. <https://doi.org/10.1002/oby.23722>.
 57. Lloyd-Jones, L.R., Holloway, A., McRae, A., Yang, J., Small, K., Zhao, J., Zeng, B., Bakshi, A., Metspalu, A., Dermitzakis, M., and Powell, J.E. (2017). The genetic architecture of gene expression in peripheral blood. *Am. J. Hum. Genet.* 100, 228–237. <https://doi.org/10.1016/j.ajhg.2016.12.008>.
 58. Westra, H.J., Peters, M.J., Esko, T., Yaghootkar, H., Schurmann, C., Kettenunen, J., Christiansen, M.W., Fairfax, B.P., Schramm, K., Powell, J.E., and Franke, L. (2013). Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat. Genet.* 45, 1238–1243. <https://doi.org/10.1038/ng.2756>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
eQTLGen Consortium (eQTL) - Blood	eQTLGen	https://www.eqtlgen.org/phase1.html ; SMR Format Data: https://cnsgenomics.com/software/smr/#DataResource
GTExV8 (eQTL) – Artery, Liver, Subcutaneous adipose, Visceral adipose, Blood	The Geno-type-Tissue Expression (GTEx) Project	https://www.gtexportal.org/home/datasets ; SMR Format Data: https://cnsgenomics.com/software/smr/#DataResource
pQTL atlas (Icelandic Cancer Project and deCODE genetics) - Blood	deCODE	https://www.decode.com/summarydata/
pQTL atlas (INTERVAL study) - Blood	Sun et al. ²⁹	http://www.phpc.cam.ac.uk/ceu/proteins/
pQTL atlas (ARIC study) - Blood	Zhang et al. ³⁰	http://nilanjanchatterjeelab.org/pwas/
GLGC lipid summary statistics – LDL, HDL, TG	GLGC consortium	http://csg.sph.umich.edu/willer/public/glgc-lipids2021/
UK Biobank phenotype summary statistics	leelab	https://www.leelabsg.org/resources
UK Biobank	Biobank UK	https://biobank.ndph.ox.ac.uk/showcase/
Software and algorithms		
PRS-CS	Ge et al. ³¹	https://github.com/getian107/PRScs
TwoSample MR	TwoSample MR	https://github.com/MRCIEU/TwoSampleMR
MR-PRESSO	Verbanck et al. ¹⁶	https://github.com/rondolab/MR-PRESSO
Summary-data-based Mendelian Randomization (SMR)	Zhu et al. ³²	https://cnsgenomics.com/software/smr/#Overview
LDSC	Finucane et al. ³³	https://github.com/bulik/ldsc
R	The R Project for Statistical Computing	https://www.r-project.org/version4.0.1
PhenoScanner	Staley et al., 2016 ³⁴	http://www.phenoscanner.medschl.cam.ac.uk/
STITCH	Kuhn et al. ³⁵	http://stitch.embl.de/cgi/download.pl
WebTWAS	Cao et al. ³⁶	http://www.webtwas.net/#/genes
Open Target platform	Ochoa et al. ³⁷	https://platform.opentargets.org/downloads
GWAS catalog	Bunielo et al. ³⁸	https://www.ebi.ac.uk/gwas/downloads
EWAS atlas	Li et al. ³⁹	https://ngdc.cncb.ac.cn/ewas/downloads
EWAS catalog	Battram et al. ⁴⁰	http://ewascatalog.org/download/
ExPheWas	Legault et al. ⁴¹	https://exphewas.ca/v1/gene
FUMA	Watanabe et al. ⁴²	https://fuma.ctglab.nl/
MetaXcan	Barbeira et al. ⁴³	https://github.com/hakyimlab/MetaXcan

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Hong-Hee Won (wonhh@skku.edu).

Materials availability

This study did not generate new reagents.

Data and code availability

- Data: The data supporting the findings of this study are available in the following resources. Data from UK Biobank are available on application to its site (UKB, <https://biobank.ndph.ox.ac.uk/showcase/>). The eQTL data used in the analyses described here are freely accessible on eQTLgen (<https://www.eqtogen.org/phase1.html>) and GTEx (<https://www.gtexportal.org/home/datasets>). The GWAS summary statistics for lipids from the GLGC consortium are available on the website (<http://csg.sph.umich.edu/willer/public/glgc-lipids2021/>) and 1129 phecodes are available here: <https://www.leelabsg.org/resources>. Proteome scan and protein interaction information are available on PhenoScanner v2 (<http://www.phenoscaner.medschl.cam.ac.uk/>) and STITCH (<http://stitch.embl.de/cgi/download.pl>). Target genes were scored based on the data from WebTWAS (<http://www.webtwas.net/#genes>), Open Target platform (<https://platform.opentargets.org/downloads>), GWAS catalog (<https://www.ebi.ac.uk/gwas/downloads>), EWAS atlas (<https://ngdc.cncb.ac.cn/ewas/downloads>), EWAS catalog (<http://ewascatalog.org/download/>), and ExPheWas (<https://exphewas.ca/v1/gene>). Evidence from the ongoing drug trials (phases 1–3) was procured from the commercial Pharmaprojects database, with limited access obtained through a data transfer agreement (<https://pharmaintelligence.informa.com/products-and-services/clinical-planning/pharmaprojects>) and drug approval evidence were obtained from WHOCC (https://www.whocc.no/atc_ddd_index/?code=C10), DrugBank (<https://go.drugbank.com/targets>), and ChEMBL databases (https://www.ebi.ac.uk/chembl/g/#search_results/targets). Differential gene expression and tissue specificity were examined with FUMA (<https://fuma.ctglab.nl/>).
- Code: Previously developed pipelines were used to produce the results for the current study. No custom code was developed. Please see the Supplementary Information for details on the software URLs and data used.
- Any additional information required to reanalyze the data reported in this work paper is available from the [lead contact](#) upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Study population

Large-scale prospective individual-level genetic and phenotypic data were obtained from the UK Biobank, which includes over 500,000 participants (47% males) of 40–69 years age (<https://www.ukbiobank.ac.uk/>). Genotyping was performed on 487,409 participants using two closely related arrays: the UK BiLEVE and the UK Biobank Axiom. Participants who were defined as having a self-reported non-white British ancestry, non-European ancestry in genetic PCA, and whose kinship coefficient denoted a third-degree or closer relationships were removed; 276,249 participants qualified the quality control.⁴⁴ Individual disease diagnosis codes were collected through the International Classification of Diseases (ICD) 10 and ICD 9 codes and mapped to the phenotype codes (phecodes). The UK Biobank was approved by the North West Multi-Centre Research Ethics Committee (June 17, 2011, extended on May 10, 2016 [RES ref. 16/NW/0274]), and written informed consent using samples and data for medical research purposes was obtained from all Participants. The present research using the UK Biobank Resource was approved under Application Number 33002. This study was approved by Sungkyunkwan University Institutional Review Board (2022-05-013).

METHOD DETAILS

Study design

A description of the analytical workflow and study design is shown in (Figure 1). Our analysis consisted of two parts: a genome-wide analysis (phase 1) to inform repurposing opportunities and adverse effects, and a genome-level analysis (phase 2) to prioritize targets, of which two pipelines are interconnected (Figure 1).

Polygenic risk scores

Plasma lipid (LDL cholesterol, HDL cholesterol, and TG)-associated genetic variants were selected from the most recent GWAS of the GLGC consortium with over 1 million participants.¹⁸ PRSs were generated for each plasma lipid using PRS–Continuous Shrinkage (PRS-CS) on UK Biobank genotyped individuals of European descent.^{31,45} PRS-CS computes SNP effect sizes by a high-dimensional Bayesian regression framework using GWAS summary statistics and an external LD reference panel, the 1000 Genomes Project Phase 3 European sample.

Phenome-wide associations

Phecodes, provided by Lee lab (<https://www.leelabsg.org/resources>), were used to perform a PheWAS (Data S1: Table S2, related to Figure 2A). To ensure sufficient statistical power, phecodes with fewer than 200 cases were excluded and 1129 phenotypes were retained for the analysis.^{46,47} The PheWAS was performed using plasma lipid-PRS by fitting a Firth's bias-reduced logistic regression model,^{48,49} while adjusting for the age, gender, top 10 PCs of ancestry, and genotyping arrays. FDR method was used for correcting the multiple testing problems using the *p.adjust* function in R ($q < 0.05$).⁵⁰ The phecodes significantly associated with plasma lipid-PRS in PheWAS (FDR $Q < 0.05$) were taken forward for MR analysis to test causality.

Mendelian randomization

In the MR analysis, plasma lipid levels (LDL, HDL, and TG levels) were placed into the exposure and phenotypes important for PheWAS were placed into the outcome. We used lipid-associated SNPs at $p < 5 \times 10^{-8}$ from the summary statistics (GLGC) as instrumental variables, and LD clumped ($r^2 < 0.001$, window within 10,000 base pairs based on the European 1000 Genomes Project reference panel) for each lipid. Univariate two-sample MR analyses were conducted, and the phenotypes significant in at least 2 MR methods among IVW (random-effects model), weighted median, MR-Egger, and MR-PRESSO were retained. For phenotypes that were significant in at least two lipids as per univariate MR analysis, MVMR was performed to adjust the other lipids to isolate the effect of each lipid on a given health outcome. Phenotypes causally affected by the lipids (MR-validated) were finally categorized into 4 groups (Figure 1 – phase 1; Figure 2C).

Transcriptome-wide Mendelian randomization

SMR method was used to integrate GWAS and eQTL data to explore causal inference.³² We used *cis*-eQTLs that were associated with common SNP (minor allele frequency $>0.1\%$), as *cis*-variants are more likely to have specific effects on the gene of interest.⁶ To curate other tissues that are highly relevant to serum lipids, we calculated linkage disequilibrium score regression (LDSC) heritability and isolated five significant tissues³³ (Figure S2). Gene expression data for whole blood were obtained from the eQTLGen consortium, and gene expression data for the other four tissues were obtained from Genotype-Tissue Expression (GTEx) database v8. HEIDI was tested for colocalization.^{32,51} We set the Bonferroni correction threshold of $<1.24E-6$ ($<0.05/40302$ – a total number of genes tested for 5 tissues) for both SMR and HEIDI. We carried out transcriptome-wide MR for three lipids (LDL, HDL, and TG) and categorized responsible genes for lipid levels into 4 groups (Figure 1 – phase 2; Data S1: Table S1, related to Figure 1).

Identification of top tier pleiotropic genes

The analytic pipeline for drug target prioritization is shown in (Figure 3A). Considering that multi-trait (correlated) approaches would increase the power to detect actionable targets and that exploring the pleiotropic gene could be a promising strategy,^{10,11} we used the category 4 phenotypes and category 4 genes to isolate pleiotropic genes (Figure 1). Among the category 4 genes responsible for multiple lipid profiles, genes that were also causal to a maximum number of lipid-driven traits were identified and termed “top tier” genes (Figure 1 – phase 2; Figure 3A).

Proteome scan and protein-protein interactions

The consolidation of evidence from eQTL and protein QTL (pQTL) is helpful in drug target investigations.^{3,28} For this reason, we performed a proteome scan for top *cis*-eQTLs of top tier genes using PhenoScanner.³⁴ We identified proteins that were significantly associated with the top SNPs ($p < 1 \times 10^{-5}$) of the top tier genes, and in absence of top SNPs, we used LD proxies ($r^2 > 0.8$). The non-seed genes that lack direct genetic evidence were curated because Fang et al. corroborated the merit of extending drug target to non-seed genes that lack direct genetic evidence but are highly potential based on network connectivity.⁵² Thereby, we further curated associated proteins and molecular pathways that are interacting with prioritized proteins using physical and functional associations derived from genomic contexts, high-throughput experiments, co-expression, and literature mining (Figure 3A). The GO and KEGG enrichment tools in STITCH were used to specify proteins and pathway molecular interactions, and the genes within the most significant biological GO and KEGG gene sets were identified.

Scoring prioritized targets

To rank the “novelty” and “potential” of the prioritized targets, we introduced a scoring system, similar to that used in previous studies.^{52,53} The scoring system consists of 7 components and the total score is the sum of the following fulfilled criteria: 1) significant genes ($P < 5E-08$) identified from GWAS of lipid levels, 2) significant genes ($P < 1E-06$) identified from transcriptome-wide association study (TWAS) of lipids or related CVDs, 3) causal genes identified from eQTL MR of our study (Data S1: Table S14, related to Figure 3), 4) significant genes ($P < 1E-07$) identified from epigenome-wide association study (EWAS) of lipids or related CVDs, 5) genes significantly associated with lipids or related CVDs in PheWAS (passing Bonferroni threshold), 6) genes prioritized by PubMed text mining, and 7) genes that have been investigated for their associations with lipids and CVDs using animal models. Target genes were assessed using WebTWAS,³⁶ Open Target platform,³⁷ GWAS catalog,³⁸ EWAS atlas,³⁹ EWAS catalog,⁴⁰ and ExPheWas.⁴¹ The ExPheWas used the PCA-based association model to perform a gene-based PheWAS to generate a condensed representation capturing genetic variability within gene regions.⁴¹ Targets scored 4 or higher were deemed *highly potential targets*, whereas those scored less than 4 were considered *relatively novel targets* that have not been extensively explored.

Gene expression and tissue specificity

We explored gene expression and tissue specificity of 30 prioritized gene targets via FUMA and 53 tissue-type data from GTEx.⁴² A heatmap was drawn using the average of the normalized gene expression per tissue per gene that allowed comparison of the expression levels across tissue types within a gene.⁴² To assess tissue specificity, differentially expressed gene (DEG; genes that are significantly more or less expressed in a given tissue compared to others) sets were investigated,⁴² wherein two-sided Student's *t*-tests were performed per gene per tissue against all other tissues.⁴² The *p* values for DEG are smaller for genes whose expression in a particular tissue showed the largest discrepancy compared to expression in all other tissues.⁴² The sign of the *t*-score was used to examine gene upregulation and downregulation in specific tissues.

Phenome-wide MR analysis of category 1–4 phenotypes

We performed eQTL/pQTL MR analysis on category 1–3 phenotypes to identify drug repurposing opportunities and potential adverse effects associated with target genes, and on category 4 phenotypes to validate the potential causal effect of 30 therapeutic targets on the dyslipidemia and lipid-driven cardiometabolic outcomes. However, among the 30 therapeutic targets, SREBF2 was excluded from the eQTL/pQTL MR analysis because of lack of eQTL and pQTL data. MR analysis was performed on category 1–4 phenotypes utilizing SMR methods and eQTL data of five tissues. When eQTL data from multiple tissues were available for a given target gene-outcome pair, the results from the most significant tissue were presented in the heatmap (Figure 5). In addition, we obtained pQTL data from Egil et al.⁵⁴ and conducted conventional MR analysis for category 1–4 phenotypes using genetic variants associated with plasma protein levels as genetic instruments. We used four MR methods (inverse variance weighted [random-effects model], weighted median, MR-Egger, and MR-PRESSO) to test the consistency of the results across varying assumptions of heterogeneity and pleiotropy effects. For plasma proteins with only one genetic instrument, we conducted Wald ratio analysis. We presented the best causal estimations for each target-outcome pair in the heatmap by selecting the estimates from the robust MR method to the pleiotropy and heterogeneity of each pair. A detailed flow diagram of selecting the best causal estimation is described in (Figure S1) of a previous study by Kim et al.⁵⁵ When both eQTL and pQTL data were available for a given target-outcome pair, the results from the most significant QTL were presented in the heatmap. We defined $p < 0.05$ as nominal statistical significance, selected a robust significant threshold, and corrected the number of therapeutic candidates, as our threshold for follow-up analyses (number of therapeutic candidates = 29; $p < 1.72 \times 10^{-3}$).

In silico validation

We replicated the phenome-wide MR analysis for Category 4 phenotypes using different datasets for exposure (eQTL^{56,57}/pQTL^{29,30}) and outcomes. We developed three validation sets, with the first using a different outcome dataset (FinnGen⁵⁸). Validation 2 employed different exposure datasets (eQTL/pQTL) whereas Validation 3 utilized different exposure and outcome datasets from those used for discovery. Moreover, we used the MetaXcan framework to perform transcriptome-wide gene-based analysis that integrates large-scale transcriptome data (eQTLs) with the summary statistics of GWAS to validate the findings using a different methodology; we tested whether the predicted expression levels of the 30 prioritized genes were associated with dyslipidemia.⁴³ We explored potential causal genes in the prespecified five tissues (whole blood, subcutaneous adipose, visceral adipose, arterial, and liver tissues) that were selected based on LDSC heritability (Figure S2). The 30 therapeutic candidates from the five tissues were used to determine the correlation between predicted expression levels and dyslipidemia. Here, we defined $p < 0.05$ as the nominal statistical significance and set the Bonferroni correction threshold at $<1.67E-3$ ($<0.05/30$ therapeutic candidates). The targets were considered validated when they showed statistically significant association at least once in three MR validations or with MetaXcan for hypercholesterolemia or hyperlipidemia.

QUANTIFICATION AND STATISTICAL ANALYSIS

The MR analysis was conducted using TwoSampleMR R package (<https://github.com/MRCIEU/TwoSampleMR/>) and MR-PRESSO (<https://github.com/rondolab/MR-PRESSO>). Statistical analyses and graph generation were performed using R v4.0.1. PRS-CS was used to generate PRS on UK Biobank individuals using plasma lipid (LDL cholesterol, HDL cholesterol, and TG)-associated genetic variants. LDSC was used to estimate partitioned heritability for plasma lipids. The MetaXcan framework was used to perform transcriptome-wide gene-based analysis. The WebTWAS, Open Target platform, GWAS catalog, EWAS atlas, EWAS catalog, and ExPheWas were used to assess our prioritized target genes. We implemented the SMR and HEIDI colocalization analysis using the SMR software (<https://yanglab.westlake.edu.cn/software/smr/>). Detailed descriptions of further statistical tests are specified in the results section and the figure legends.

Supplemental information

**Prioritization of therapeutic targets
for dyslipidemia using integrative
multi-omics and multi-trait analysis**

Min Seo Kim, Minku Song, Beomsu Kim, Injeong Shim, Dan Say Kim, Pradeep Natarajan, Ron Do, and Hong-Hee Won

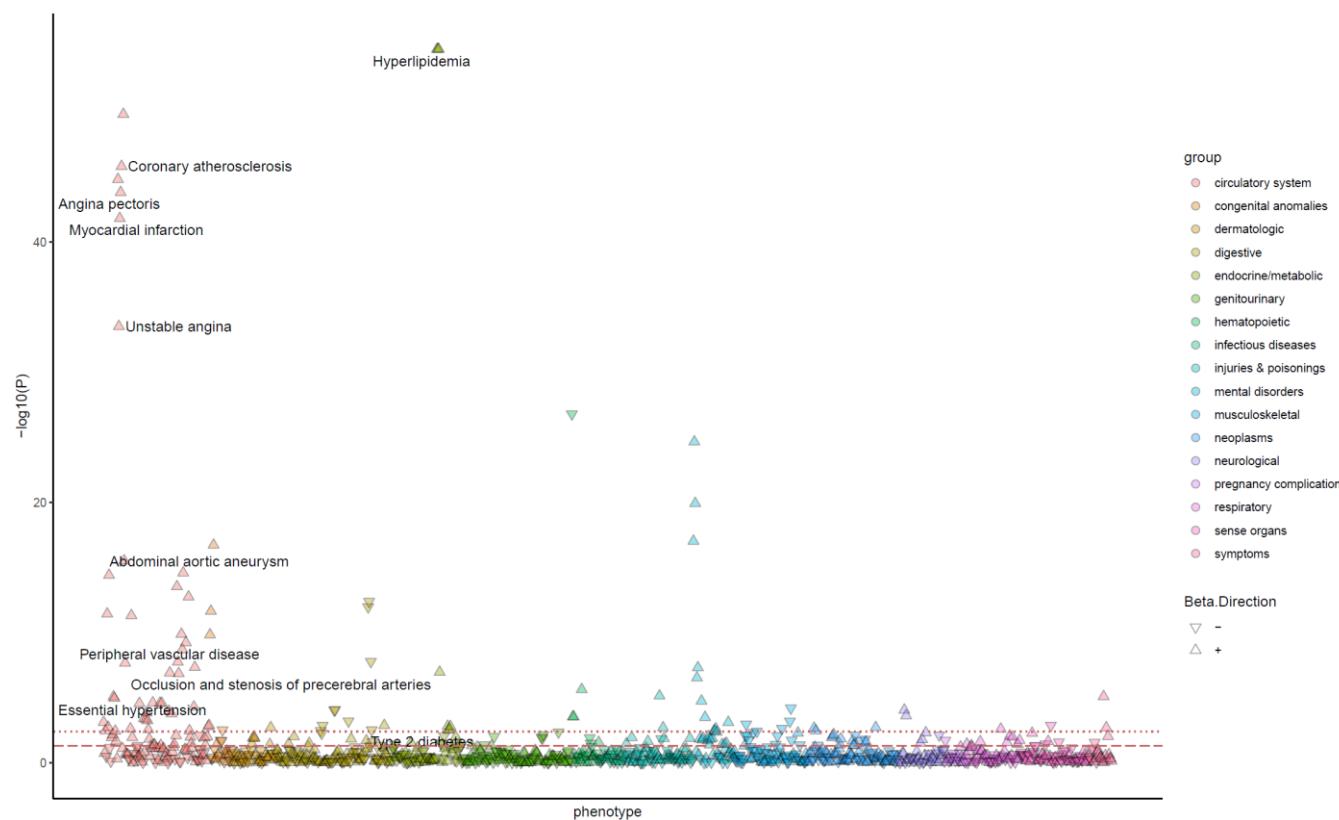
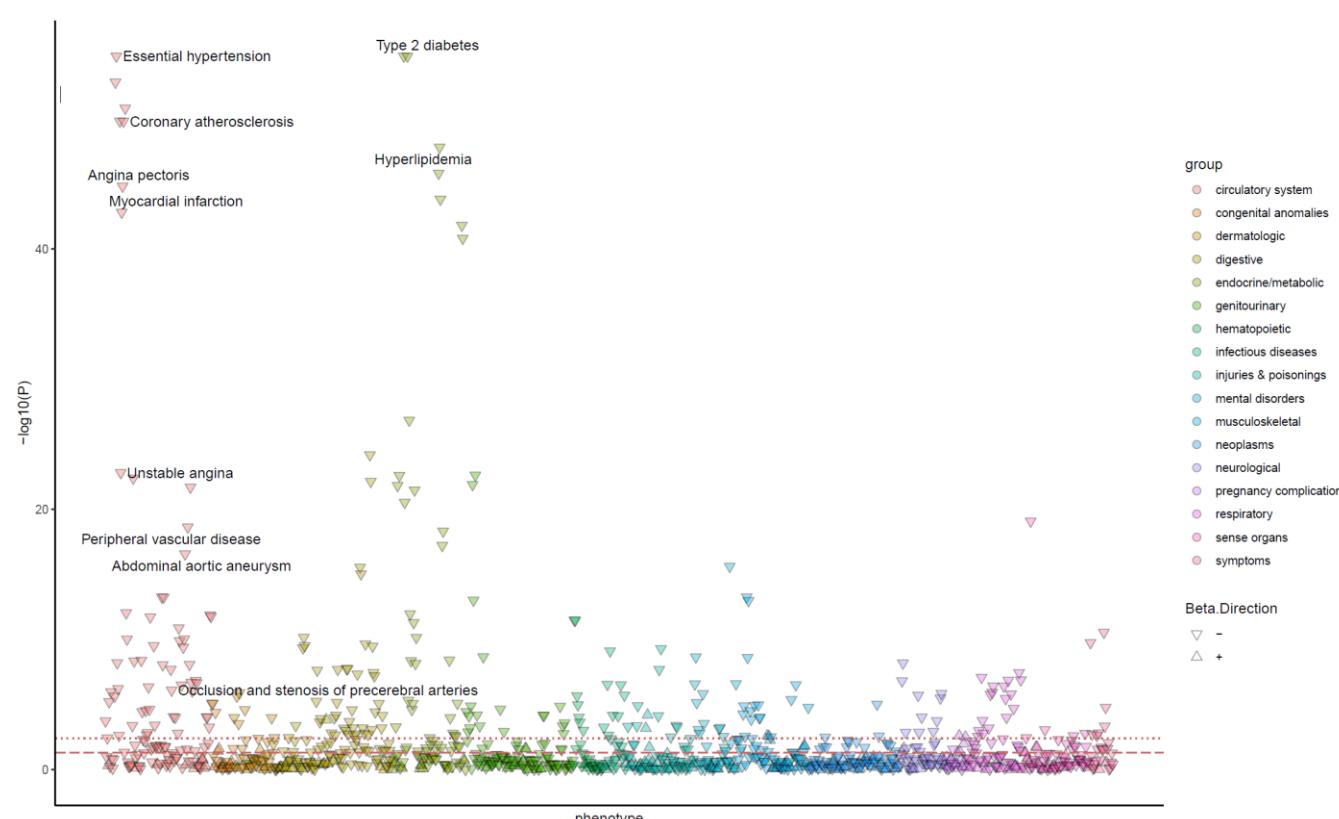
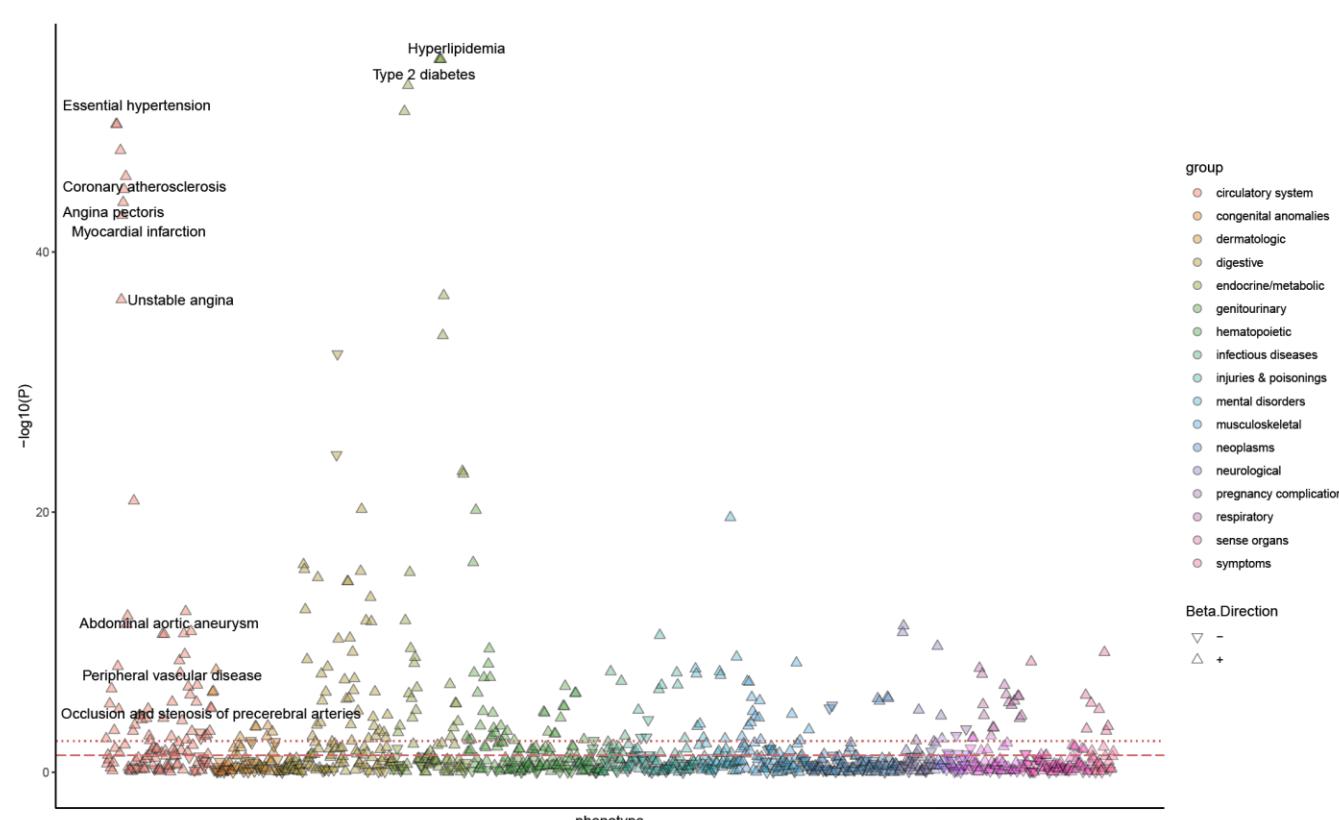
A**B****C**

Figure S1. Manhattan plots for PheWAS of LDL (A), HDL (B), and TG (C). Related Figure 2A.

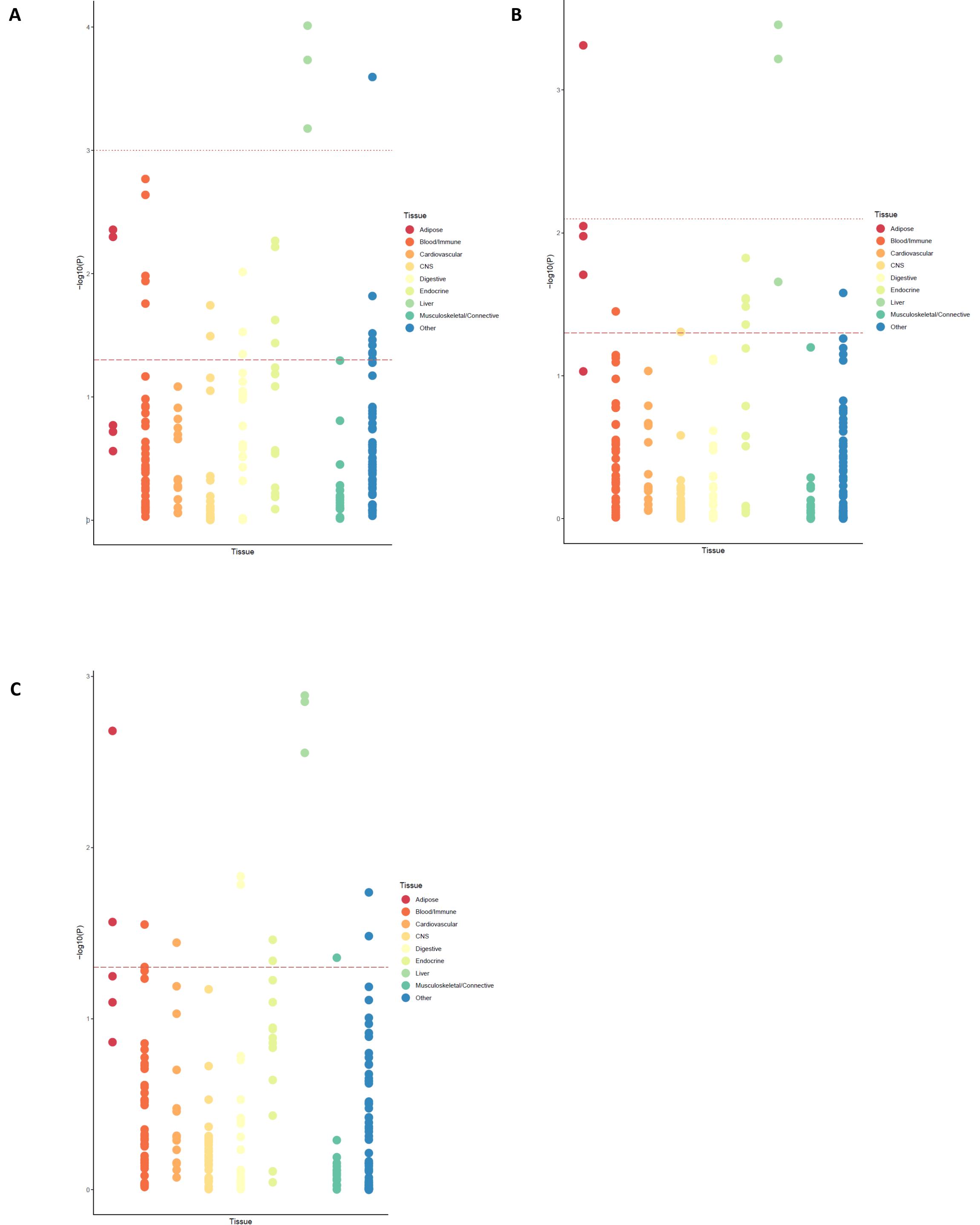


Figure S2. LDSC heritability of LDL (A), HDL (B), and TG (C). Related to Figure 3.

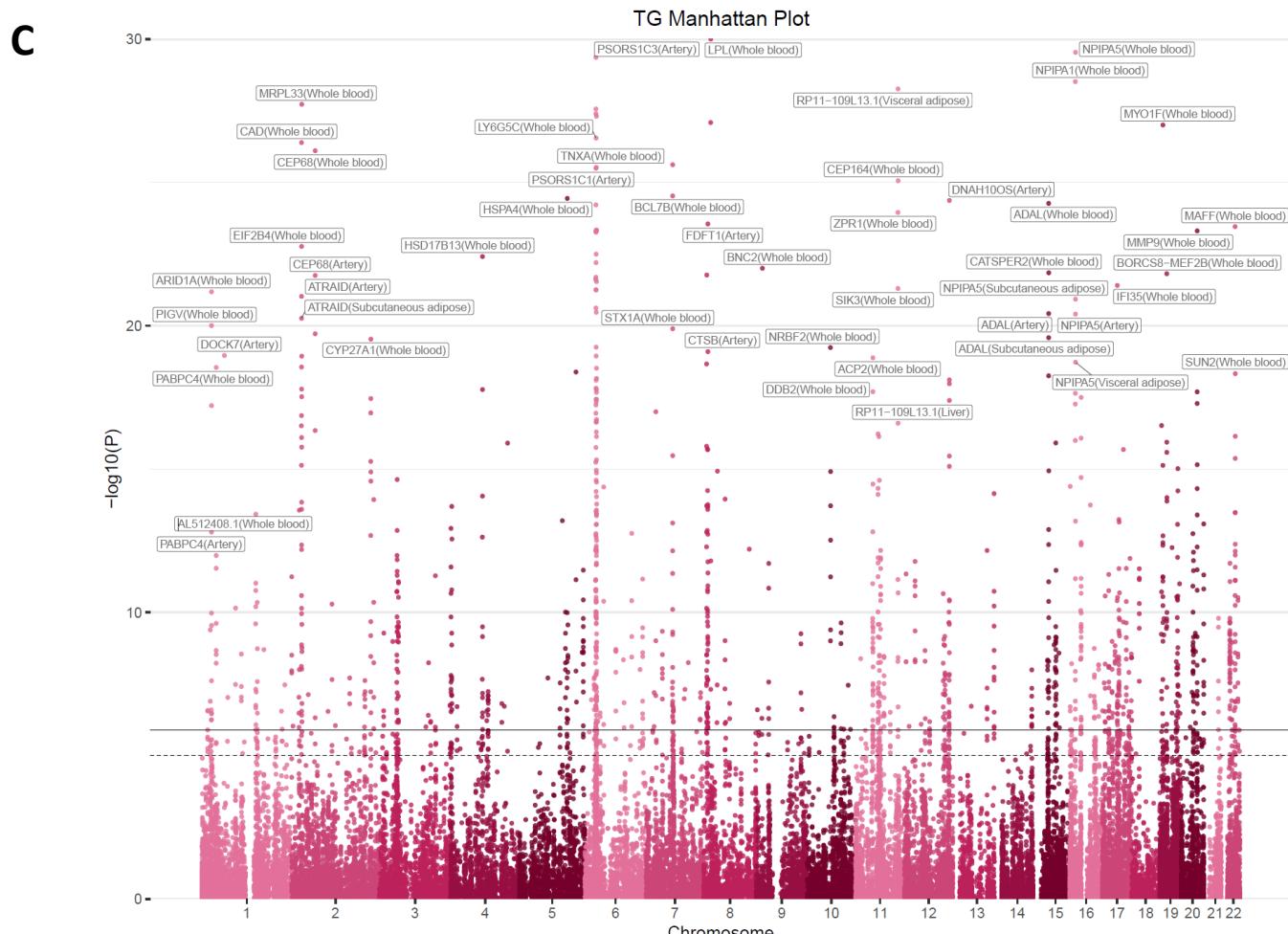
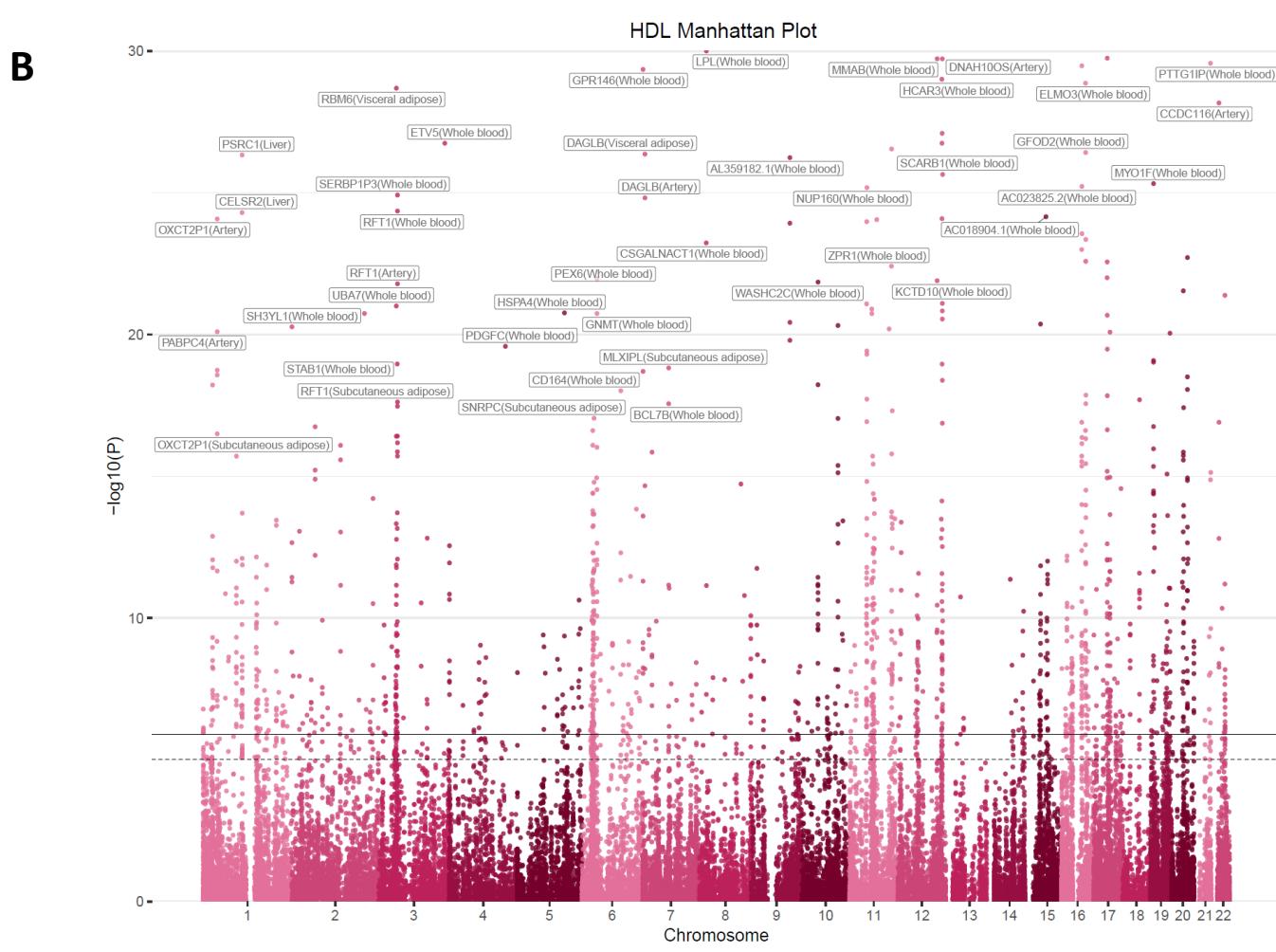
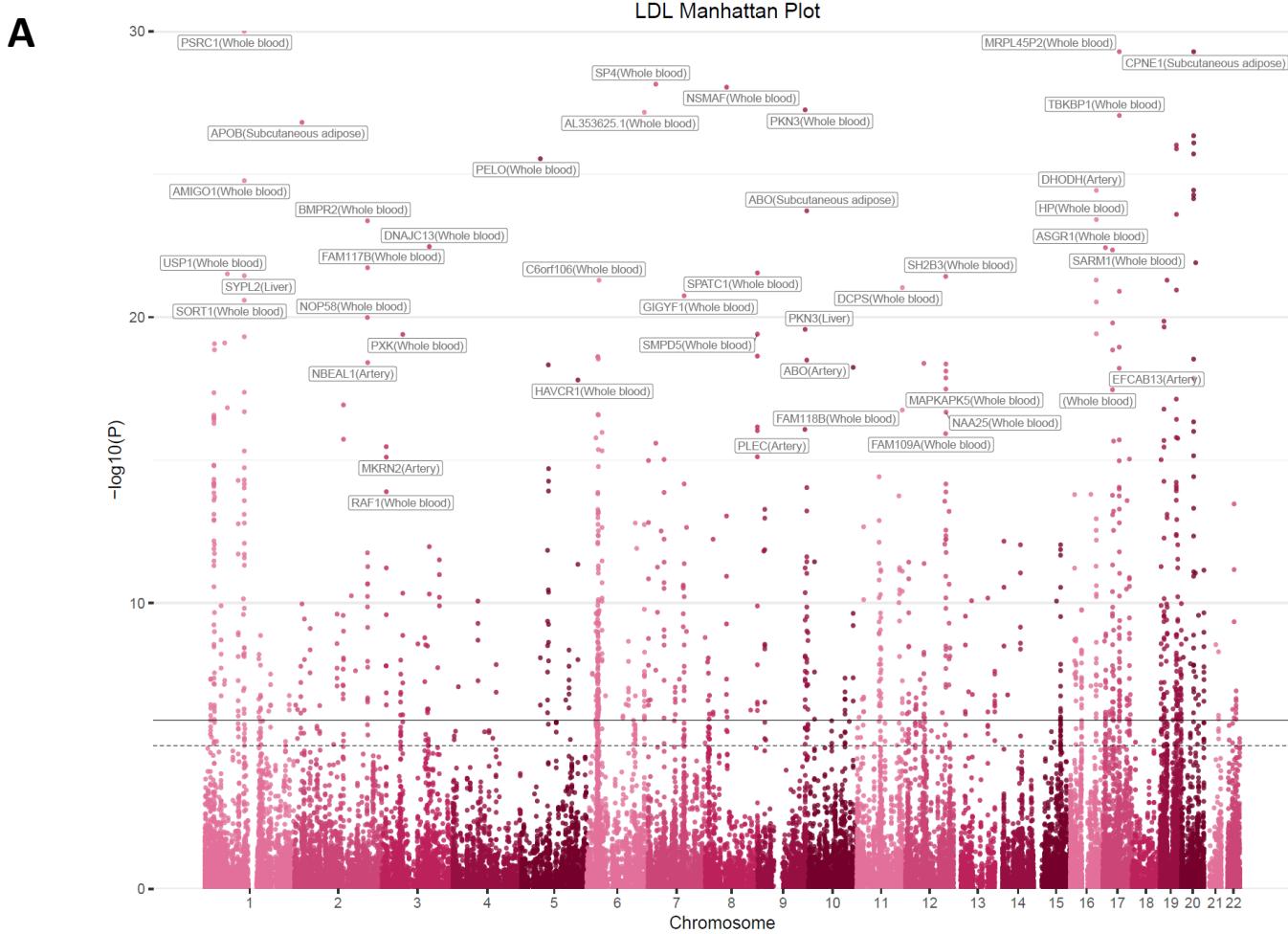


Figure S3. Manhattan plots for transcriptome-wide MR of LDL (A), HDL (B), and TG (C). Related to Figure 3.

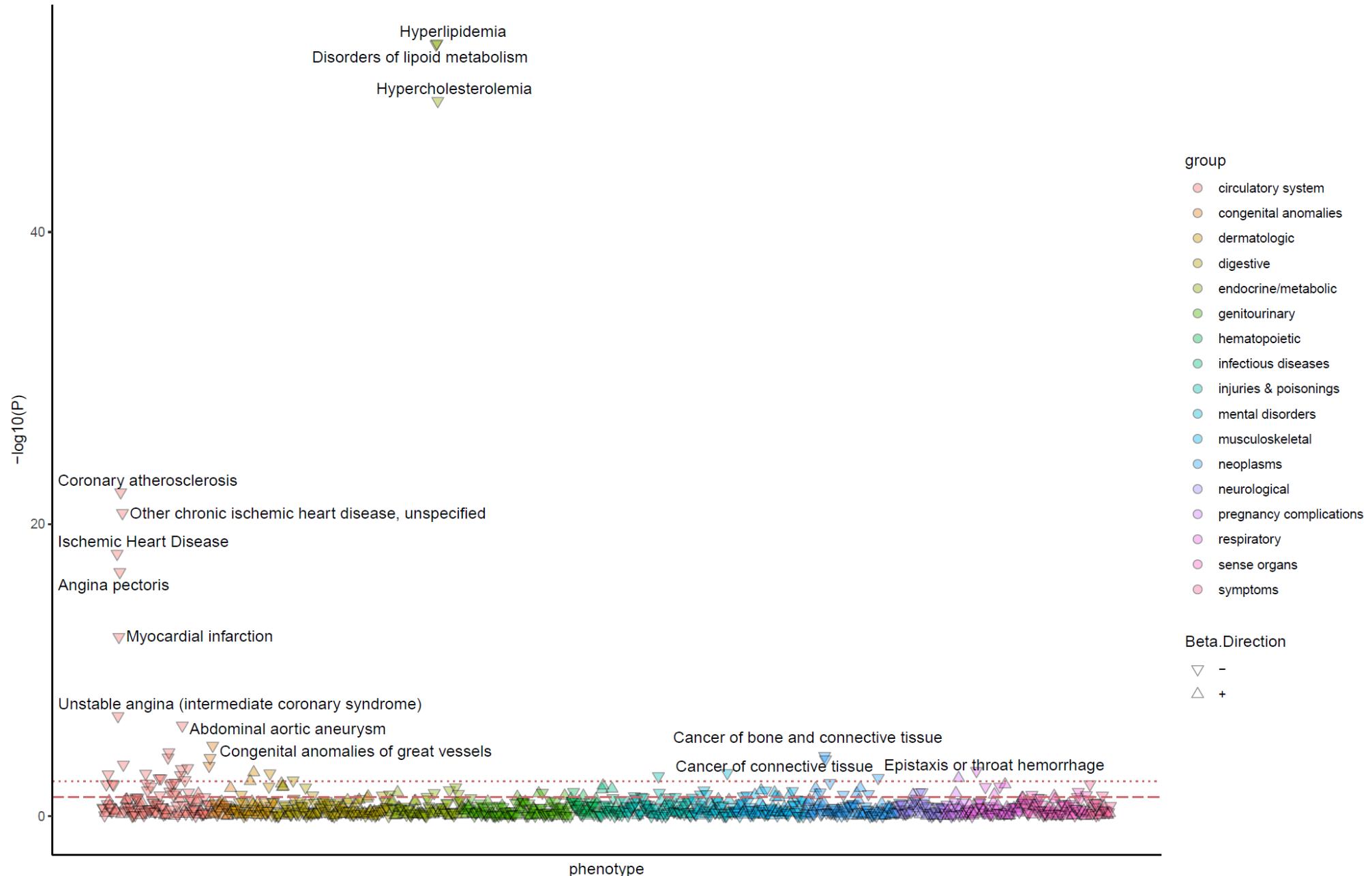


Figure S4. PheWAS of SORT1 (rs12740374). Related to Figure 3.

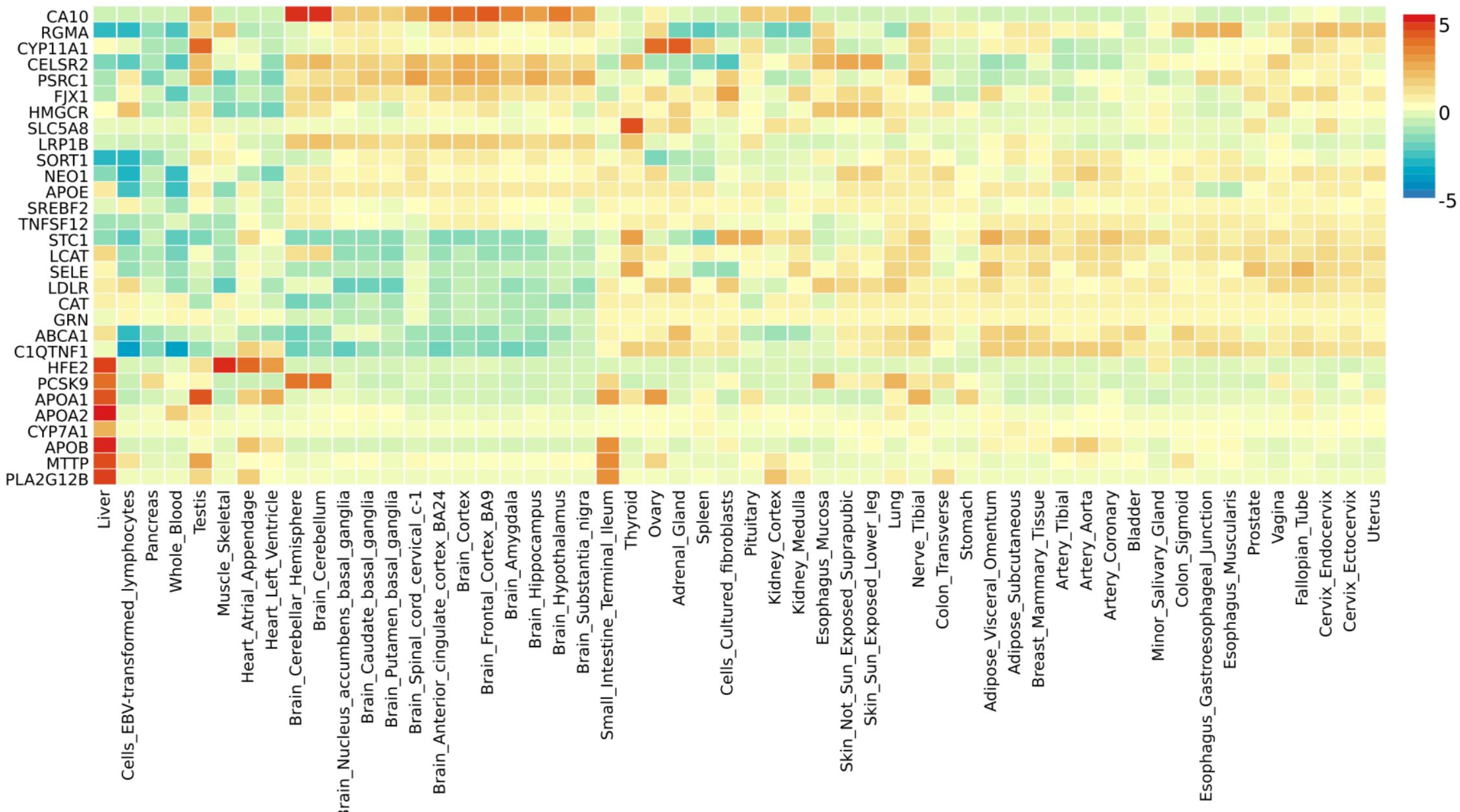
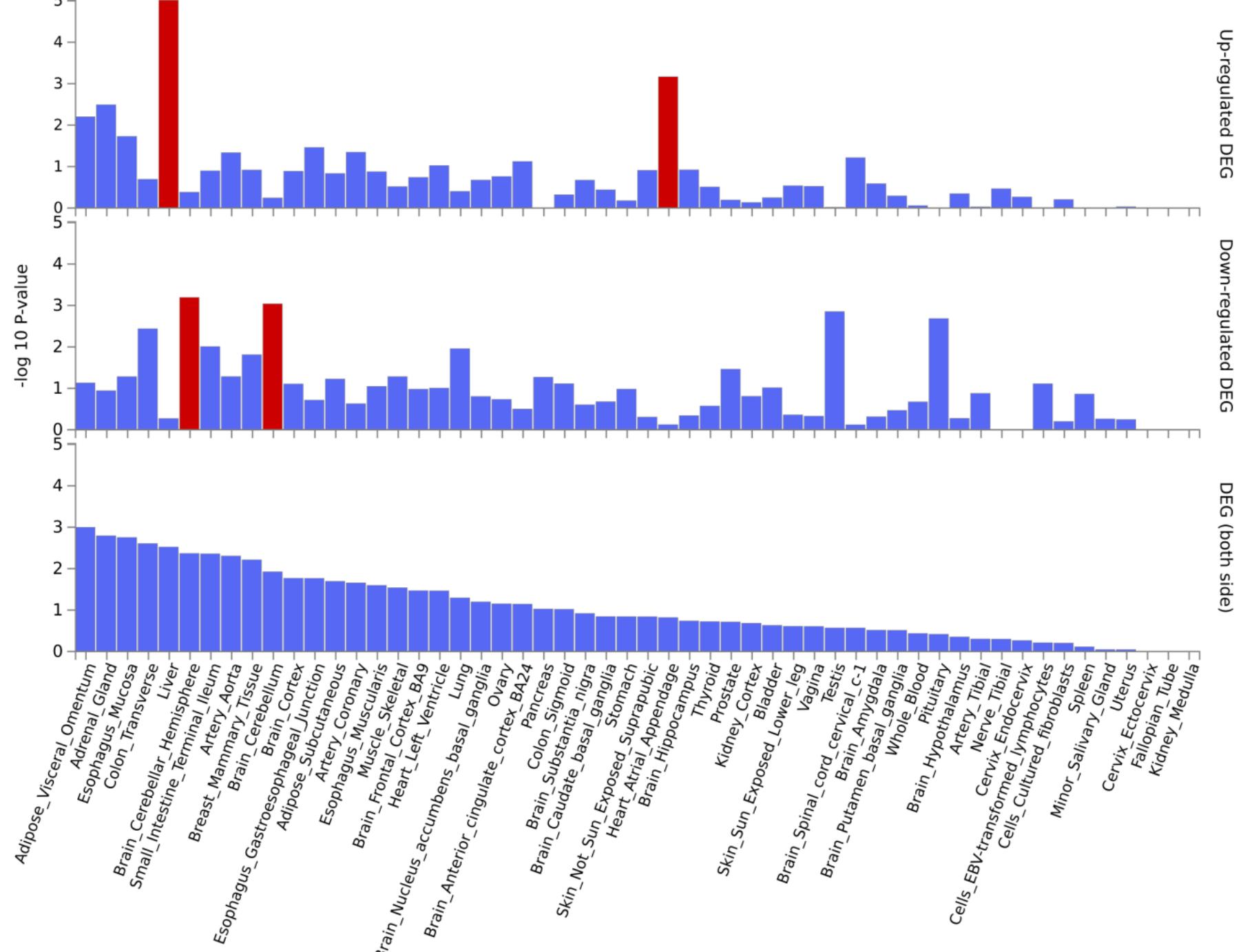
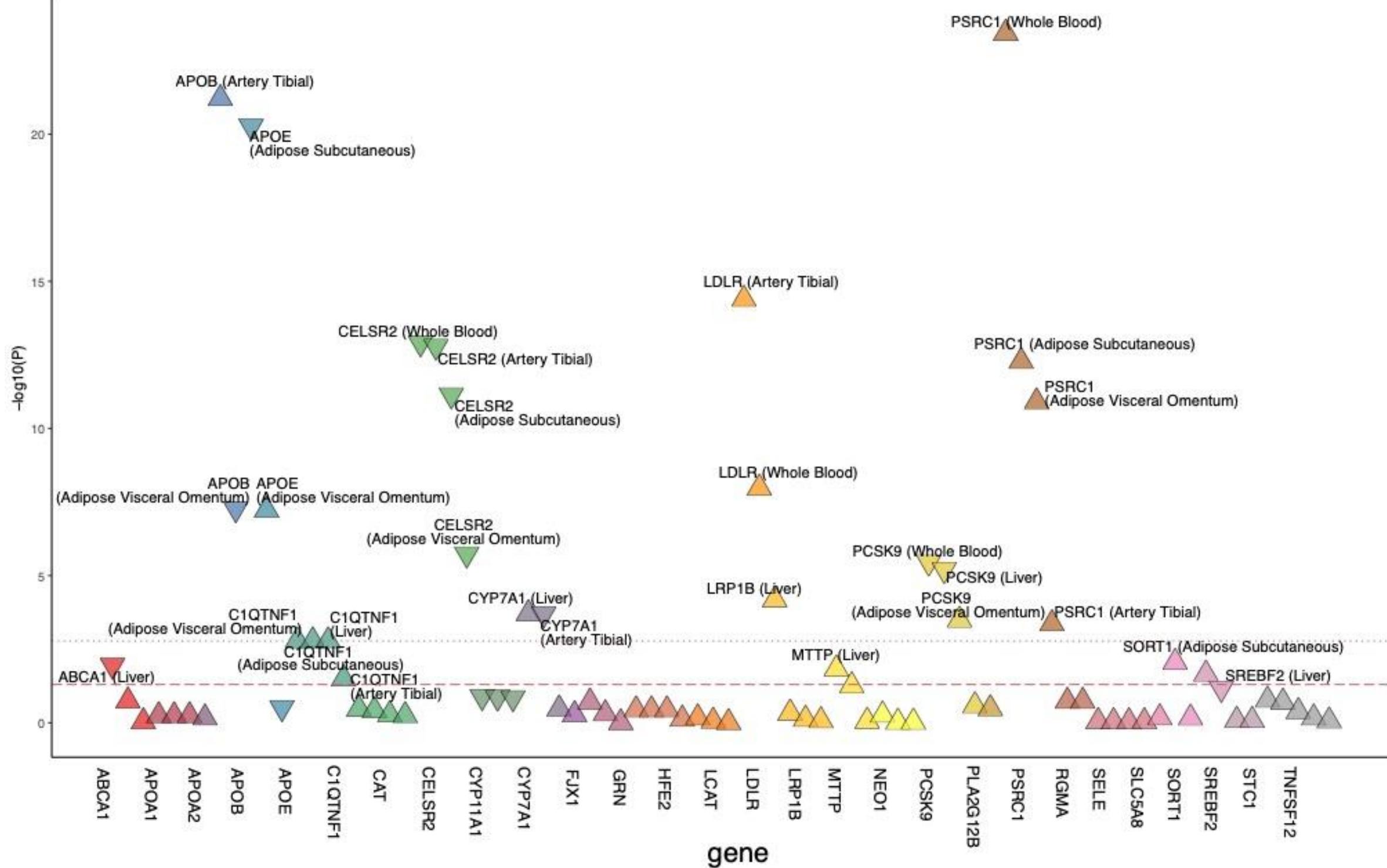
A**B**

Figure S5. Gene expression and tissue specificity of 30 prioritized genes. (A) Interactive heatmap of gene expression. (B) Tissue specificity measured using differentially expressed gene (DEG; genes that are significantly more or less expressed in a given tissue compared with others) sets for each of the 53 tissue types. Related to Figures 5 and 6.

A



B

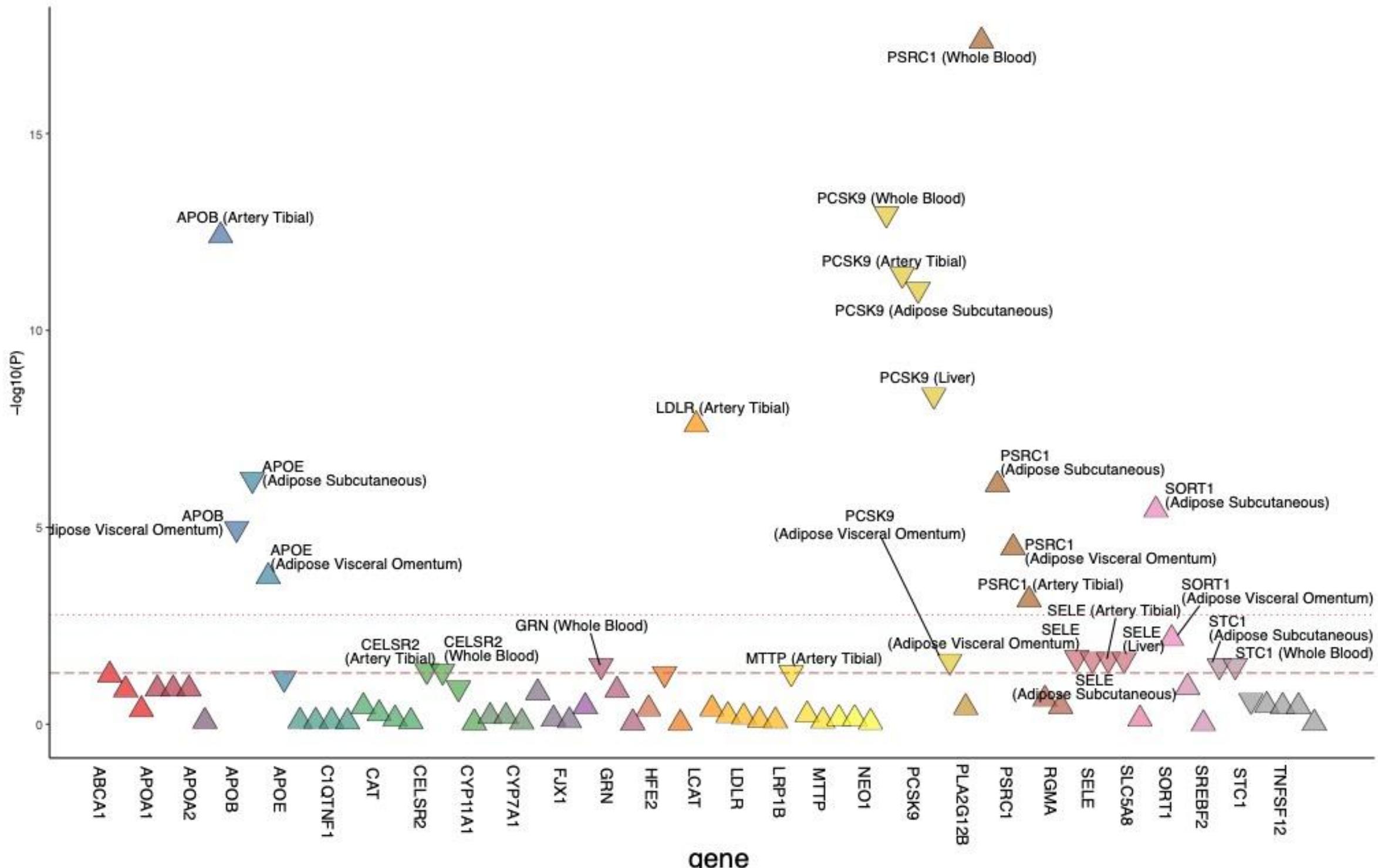


Figure S6. The metaXcan analyses for hyperlipidemia. (A) Manhattan plot for results of metaXcan for hyperlipidemia from UK Biobank. (B) Manhattan plot for results of metaXcan for hyperlipidemia from FinnGen. Related to Figures 5 and 6.

URLs

UK Biobank quality control documentation, https://www.ukbiobank.ac.uk/wp-content/uploads/2014/04/imputation_documentation_May2015.pdf; UK Biobank, <https://biobank.ndph.ox.ac.uk/showcase/>; LDSC v1.0.1, <https://github.com/bulik/ldsc>, <https://alkesgroup.broadinstitute.org/LDSCORE>; GTEx, <http://www.gtexportal.org/home/datasets>; TwoSample MR, <https://mrcieu.github.io/TwoSampleMR>; MR-PRESSO v1, <https://github.com/rondolab/MR-PRESSO>; PRS-CS, <https://github.com/getian107/PRScs>; Summary-data-based Mendelian Randomization (SMR), <https://cnsgenomics.com/software/smr/#Overview>; PhenoScanner, <http://www.phenoscaner.medschl.cam.ac.uk/>; STITCH, <http://stitch.embl.de/cgi/download.pl>; WebTWAS, <http://www.webtwas.net/#/genes>; Open Target platform, <https://platform.opentargets.org/downloads/>; EWAS atlas, <https://ngdc.cncb.ac.cn/ewas/downloads>; ExPheWas, <https://exphewas.ca/v1/gene>; FUMA, <https://fuma.ctglab.nl/>; MetaXcan, <https://github.com/hakyimlab/MetaXcan>