Proteome-Wide Association Studies for Blood Lipids and Comparison with Transcriptome-Wide Association Studies

Daiwei Zhang[1,2], Boran Gao[2], Qidi Feng[3,2], Ani Manichaikul[6], Gina M. Peloso[11], Russell P. Tracy[7], Peter Durda[8], Kent D. Taylor[5], Yongmei Liu[9], W. Craig Johnson[10], Stacey Gabriel[13], Namrata Gupta[13], Joshua D. Smith[14], Francois Aguet[3], Kristin G. Ardlie[3], Thomas W. Blackwell[15], Robert E. Gerszten[13], Stephen S. Rich[6], Jerome I. Rotter[5], Laura J. Scott[2,*], Xiang Zhou[2,*], Seunggeun Lee[4,2,*]

Abstract

Blood lipid traits are treatable and heritable risk factors for heart disease, a leading cause of mortality worldwide. Although genome-wide association studies (GWAS) have discovered hundreds of variants associated with lipids in humans, most of the causal mechanisms of lipids remain unknown. To better understand the biological processes underlying lipid metabolism, we investigated the associations of plasma protein levels with total cholesterol (TC), triglycerides (TG), high-density lipoprotein cholesterol (HDL), and low-density lipoprotein cholesterol (LDL)

[1]Department of Biostatistics, Epidemiology and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA USA. [2]Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI USA. [3]Broad Institute of Massachusetts Institute of Technology and Harvard, Cambridge, MA USA. [4]Graduate School of Data Science, Seoul National University, Seoul, Republic of Korea. [5]The Institute for Translational Genomics and Population Sciences, Department of Pediatrics, The Lundquist Institute for Biomedical Innovation at Harbor-UCLA Medical Center, Torrance, CA USA. [6]Center for Public Health Genomics, University of Virginia, Charlottesville, VA USA. [7]Departments of Pathology & Laboratory Medicine, and Biochemistry, Larner College of Medicine, University of Vermont, Burlington, VT USA. [8]Departments of Pathology & Laboratory Medicine, Larner College of Medicine, The University of Vermont, Burlington, VT USA. [9]Department of Medicine, Divisions of Cardiology and Neurology, Duke University Medical Center, Durham, NC USA. [10]Department of Biostatistics, University of Washington, Seattle, WA USA. [11]Department of Biostatistics, Boston University School of Public Health, Boston, MA USA. [12]Division of Cardiovascular Medicine, Beth Israel Deaconess Medical Center, Boston, MA USA. [13]Genomics Platform, Broad Institute of Massachusetts Institute of Technology and Harvard, Cambridge, MA USA. [14]Department of Genome Sciences, Human Genetics and Translational Genomics, The University of Washington, Seattle, WA, USA. [15]Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, St. Louis, MO USA. *Corresponding authors.

in blood. We trained protein prediction models based on samples in the Multi-Ethnic Study of Atherosclerosis (MESA) and applied them to conduct proteome-wide association studies (PWAS) for lipids using the Global Lipids Genetics Consortium (GLGC) data. Of the 749 proteins tested, 42 were significantly associated with at least one lipid trait. Furthermore, we performed transcriptome-wide association studies (TWAS) for lipids using 9,714 gene expression prediction models trained on samples from peripheral blood mononuclear cells (PBMCs) in MESA and 49 tissues in the Genotype-Tissue Expression (GTEx) project. We found that although PWAS and TWAS can show different directions of associations in an individual gene, 40 out of 49 tissues showed a positive correlation between PWAS and TWAS signed p-values across all the genes, which suggests a high-level consistency between proteome-lipid associations and transcriptome-lipid associations.

Introduction

Blood lipid levels, including levels of total cholesterol (TC), triglycerides (TG), high-density lipoprotein cholesterol (HDL), and low-density lipoprotein cholesterol (LDL), are heritable risk factors (Pilia et al., 2006) for coronary heart disease and stroke (Kannel et al., 1961; Willer & Mohlke, 2012), which are leading causes of death in the U. S. and other nations (Ahmad & Anderson, 2021; Roger et al., 2011). Genome-wide association studies (GWAS) have identified hundreds of loci that are significantly associated with at least one lipid trait in humans (Chen et al., 2013; de Vries et al., 2019; Graham et al., 2021; Hoffmann et al., 2018). Variant alleles associated with higher concentration of LDL are more abundant among subjects with coronary artery disease than those without (Willer et al., 2008). In addition, GWAS on lipids have

facilitated the discovery of biological processes involved in lipoprotein metabolism (Burkhardt et al., 2010; Kozlitina et al., 2014; Musunuru et al., 2010).

Although GWAS have been successful in identifying loci associated with lipids, they only explain a small proportion of the heritability (Manolio et al., 2009), estimated to be 35% to 60% for TG, HDL, and LDL (Kathiresan et al., 2007). Moreover, most of these variants are located in non-coding regions with unclear functional roles (Willer et al., 2013). Because of population stratification and linkage disequilibrium, it is difficult to pinpoint the exact causal variants (Visscher et al., 2012). In addition, the large number of candidate variants severely limits the statistical power of GWAS (Brandes et al., 2020; Wang et al., 2016).

To boost the statistical power of GWAS and provide biologically meaningful interpretations, it is important to analyze downstream "omic" molecules, which include epigenetic, transcriptomic, and proteomic measurements, and then test their associations with phenotypes of interest. Recent multi-omic studies have elucidated the molecular mechanism of complex diseases (Arneson et al., 2017; Hasin et al., 2017; Leon-Mimila et al., 2019; Ramazzotti et al., 2018; Xiao et al., 2018). When downstream omic measurements are not available, which is true for many of the trait- and disease-based GWAS, the genetically expected omic values can be imputed using prediction models built upon omic and genetic data from a separate study (Gamazon et al., 2015; Gusev et al., 2016; Hu et al., 2019). An association test is then conducted on each gene between the GWAS trait and the imputed omic level. For example, based on imputed gene expression measurements, transcriptome-wide association studies (TWAS) (Cao et al., 2021; Wainberg et al., 2019; Zhu & Zhou, 2020) have been performed for various diseases and clinical characteristics, such as schizophrenia (Gusev et al., 2018), breast cancer (Bhattacharya et al., 2020), and structural neuroimaging traits (Zhao et al., 2021).

In addition to transcriptomics, proteomics provide further information for understanding complex diseases, since protein levels are downstream products of gene expression and can be more directly related to biological processes (A. P. Wingo et al., 2021). Compared to TWAS, fewer proteome-wide association studies (PWAS), imputation-based or not, have been performed. Existing PWAS have investigated the associations between proteins and colorectal cancer (Brandes et al., 2020), stroke (B.-S. Wu et al., 2022), Alzheimer's disease (A. P. Wingo et al., 2021), depression (T. S. Wingo et al., 2021), post-traumatic stress disorder (T. S. Wingo et al., 2022), and other psychiatric disorders (J. Liu et al., 2021). Regarding blood lipids, although TWAS have identified hundreds of genes associated with them (Feng et al., 2021; Veturi et al., 2021; Yang et al., 2020), to the best of our knowledge, only one PWAS has been conducted for blood lipid traits (Schubert et al., 2022).

In this work, we investigated the association of blood protein abundance with blood lipid levels to identify proteins significantly associated with lipid variability. To conduct imputation-based PWAS, we trained genotype-based protein prediction models for protein levels measured from whole blood samples from the Multi-Ethnic Study of Atherosclerosis (MESA) (Bild et al., 2002; Burke et al., 2016). The prediction models were then applied to the GWAS data of the Global Lipids Genetics Consortium (GLGC) (Willer et al., 2013) to identify proteins that are significantly associated with at least one of TC, TG, HDL, and LDL. Moreover, to study the relationship between PWAS and TWAS for lipids, we conducted imputation-based TWAS for blood lipid traits using gene expression prediction models trained on samples from MESA peripheral blood mononuclear cells (PBMCs) and samples from 49 Genotype-Tissue Expression (GTEx) project tissues (Lonsdale et al., 2013). When comparing the TWAS and PWAS directions of association with lipid across all the genes on each of the 49 tissues, for most tissues,

85 we found a positive correlation between the predicted PWAS and TWAS effects. However, for
86 individual genes, we often observed opposite predicted PWAS and TWAS directions of effects.
87
88 **Methods**
89 ***Ethics statement***
90 This work was approved by the Health Sciences and Behavioral Sciences Institutional Review
91 Board of the University of Michigan (IRB ID: HUM00152975). All data in this work were
92 collected previously and analyzed anonymously.
93
94 ***Subjects***
95 The Multi-Ethnic Study of Atherosclerosis (MESA), a part of the Trans-Omics for Precision
96 Medicine program (TOPMed) (Kowalski et al., 2019; Taliun et al., 2021), investigates
97 characteristics of subclinical cardiovascular diseases, i.e. those that are detected non-invasively
98 before the onset of clinical signs and symptoms. The study aims to identify risk factors that can
99 predict the progression of subclinical cardiovascular disease into clinically overt cardiovascular
100 disease. The diverse, population-based sample includes 6,814 male and female subjects who are
101 asymptomatic and aged between 45 and 84. The recruited participants consist of 38 percent
102 White, 28 percent Black, 22 percent Hispanic, and 12 percent Asian (predominantly Chinese)
103 individuals. In addition to genomic, transcriptomic, proteomic, and lipid data, the study also
104 collected physiological, disease, demographic, lifestyle, and psychological factors (Bild et al.,
105 2002; Burke et al., 2016).
106
107 ***Preprocessing of MESA genotypes, proteomics, and transcriptomics***

For the genotypes, we used the sequencing data from TOPMed (Kowalski et al., 2019; Taliun et al., 2021). We removed variants with minor allele frequency (MAF) of 0.05 or less among the TOPMed subjects, leaving 12,744,944 variants. Among the subjects who had genotypes, lipid levels, and demographic information, 1,438 of them were included in MESA. Samples with degrees of relatedness up to 2, as determined by KING (Manichaikul et al., 2010), were removed, which resulted in 1,403 subjects.

A total of 1,281 proteins were measured from 984 subjects. Protein levels were measured using a SOMAscan HTS Assay 1.3K for plasma proteins. The SOMAscan Assay is an aptamer-based multiplex protein assay. It measures protein levels by the number of protein-specific aptamers that successfully bind to their target protein, though some proteins may be targeted by multiple aptamers (Gold et al., 2010; Raffield et al., 2020; Schubert et al., 2022). In our analysis, targets that corresponded to multiple proteins were removed, which resulted in 1,212 proteins. As part of the TOPMed MESA Multi-Omics project, the 984 participants were selected for proteomic measurement based on the following criteria. First, participant samples were restricted to those already included in the TOPMed Whole Genome Sequencing effort (Taliun et al., 2021). Second, the race and ethnicity reflected that of participants in the parent MESA cohort. Third, participants were chosen to maximize the amount of overlapping omic data. Fourth, a substantial proportion of participants had biospecimens from MESA Exams 1 and 5.

Among these participants, 935 individuals with protein levels had blood lipid measurements, genotypes, and covariate information. After inversely normalizing the protein levels, we computed the top 10 protein principal component (PC) scores and top 10 surrogate values (Lee et al., 2017) to detect outliers and adjust for unobserved factors that might adversely affect the analysis. Samples with p-values less than 0.001 for the chi-squared statistics of either

the PC scores or the surrogate values were removed, leaving 918 samples (See Table S1 for sample characteristics). The inversely normalized protein levels were then adjusted for age, sex, self-reported race and ethnicity, usage of lipid-lowering medications, top 4 genetic PCs, and top 10 surrogate values. The residuals of the protein levels were used for the subsequent analyses.

RNA-seq was previously performed on MESA peripheral blood mononuclear cells (PBMCs) (Brown et al., 2019; Y. Liu et al., 2013). We used the reads per kilobase of transcript per million reads mapped (RPKM) of each gene in our analysis. After applying the same preprocessing pipeline as for the proteomics (i.e. sample matching, inverse normalization, outlier removal, and adjustment for the same set of covariates), we had 1,021 samples for 22,791 genes, which covered 1,167 out of the 1,212 genes in the proteomic data.

Protein and gene expression prediction models based on MESA

We performed imputation-based PWAS for lipids by using SPrediXcan (Barbeira et al., 2018) to achieve higher statistical power. SPrediXcan builds an elastic net (Zou & Hastie, 2005) prediction model of the omic measurements of each gene using its cis-SNPs as predictors. These prediction models are then combined with external GWAS summary statistics to predict the associations between the omic levels and the phenotypes of interest. Intuitively, this approach can be understood as an association study between observed phenotypes and predicted omic levels. Figure 1(a) illustrates the workflow of SPrediXcan. In our analysis, we trained the elastic nets on the MESA data to predict the preprocessed protein levels from the cis-SNPs within a window extending one mega-base (MB) upstream and 1 MB downstream of the protein's gene body (from the transcription starting site (TSS) to the transcription ending site (TES)). During model training, we restricted candidate predictive SNPs to those that are included in the GWAS.

The optimal elastic net penalty weights were selected by cross-validation as recommended for SPrediXcan (Barbeira et al., 2018). We used the same procedure to build the predictive models for the transcriptomic data. After model training on the MESA data, we obtained non-trivial (i.e. at least one cis-SNP has a non-zero weight) prediction models for 749 out of 1212 proteins and 886 out of 1167 gene expressions, with an intersection of 562 genes that have both a non-trivial protein prediction model and a non-trivial gene expression prediction model.

Gene expression prediction models based on the Genotype-Tissue Expression project

The Genotype-Tissue Expression (GTEx) project (Lonsdale et al., 2013) investigated the influence of regions in the human genome on gene expression and regulation in different tissues. Genotypes and gene expression levels were collected in 49 tissues from 900 post-mortem donors, and the sample size for each tissue ranged from 73 to 706. In our analysis, we downloaded gene expression prediction models pre-trained using the GTEx data by the authors of SPrediXcan, all of which had a predictive p-value less than 0.05. We applied the models to the GWAS summary statistics via the SPrediXcan framework to obtain tissue-specific TWAS results.

Imputation-based PWAS and TWAS using the Global Lipids Genetics Consortium

After training the elastic nets on the MESA data, we applied the prediction models to the GWAS summary statistics from the Global Lipids Genetics Consortium (GLGC) (Willer et al., 2013). GLGC examined the associations between the genotypes and the lipid levels of 188,577 individuals of European ancestry. GWAS effect sizes and their standard errors were obtained for more than 2 million SNPs. For each blood lipid trait, we applied the protein prediction models trained on the MESA data and the tissue-specific gene expression prediction models trained on

177 both MESA and GTEx data to the GLGC summary statistics and computed the association
178 between the lipid and the gene's protein and gene expression levels.
179
180 **Results**
181 ***Overview of PWAS results***
182 Since our PWAS is imputation-based, we first assessed the prediction power of the cis-SNPs for
183 the protein levels. Figure 1(b) shows the prediction p-values for the 749 proteins that have at
184 least one predictive cis-SNP with a non-zero weight. The cumulative distribution function (CDF)
185 of the predictive r^2 is shown in Figure S1. With the false discovery rate (FDR) controlled at 0.05
186 (Ferreira & Zwinderman, 2006), 469 (63%) of the 749 proteins were significantly predictable
187 (Figure 1 (b), Figure S1), and the predictive r^2 of these proteins ranged from 0.01 to 0.80
188 (Figure S1). This result indicates the significance of the protein prediction models and the
189 reliability of the imputation-based PWAS results.
190 We next used the protein prediction models to perform PWAS for TC, TG, HDL, LDL.
191 The quantile-quantile plot of the PWAS p-values for each lipid is shown in Figure 1(c). Overall,
192 we observed that 23, 17, 17, and 16 proteins were significantly associated (FDR $\leq$ 0.05) with TC,
193 TG, HDL, and LDL, respectively, and 42 proteins were significantly associated with at least one
194 lipid (Table 1, Figure 1(d)). Among these proteins, apolipoprotein E (APOE), haptoglobin (HP),
195 and interleukin 1 receptor antagonist (IL1RN) have been identified for their associations with
196 lipids in previous studies (Schubert et al., 2022).
197
198 ***Comparison of MESA-trained PWAS and MESA-trained TWAS***

To compare lipid PWAS with lipid TWAS from the same study samples, we also conducted TWAS using GLGC summary data with the predictive models trained on the MESA PBMC gene expression data. For each lipid trait, we compared the signed log p-value of the genes in PWAS and TWAS and computed the Spearman correlation coefficient (Myers & Sirois, 2006) (Figure 2), where the sign reflects the direction of association. The PWAS and TWAS signed log p-values were modestly positively correlated, where the correlation coefficient ranged from 0.083 to 0.144 and the correlation p-value were all below 0.05. For TC/TG/HDL/LDL, among the 23/17/17/16 genes whose proteins are associated with the lipid (Figure 1(d), Table 1), 10/2/4/5 genes have both protein and gene expression associated with the lipid. Out of these 10/2/4/5 genes, 6/2/2/3 genes' protein-lipid association direction and gene expression-lipid association direction are concordant. In particular, APOE was significantly and positively associated with LDL in PWAS but significantly and negatively associated with LDL in TWAS; leukocyte immunoglobulin-like receptor B2 (LILRB2) and Fc gamma receptor IIb (FCGR2B) were significantly negatively associated with two lipids in PWAS and positively associated with the same lipids in TWAS.

To better understand the opposing PWAS and TWAS effects in some of the genes, we used APOE and LDL as an example and compared the LDL GWAS summary statistics with the cis-SNPs' weights in the protein and gene expression prediction models. Figure 3 (top panel) shows the signed log p-values of the association between LDL and the cis-SNPs of APOE in GLGC. Effect alleles were chosen so that all the GWAS effect sizes for LDL were positive. Among SNPs with very significant GWAS p-values, effect allele C in SNP rs7412 corresponds to the Apoε2 allele of APOE (H. Wu et al., 2020; Zhen et al., 2017). This SNP is related to the stability of the APOE isoforms (Clément-Collin et al., 2006) and is a risk factor for coronary

heart disease (Tejedor et al., 2014). Another SNP with a very strong GWAS effect is rs4420638, whose effect allele G may elevate TC, TG, and HDL (Huang et al., 2015). As indicated by the colors, the sets of predictive cis-SNPs for protein and gene expression have little overlap with each other, with only one SNP (rs1114832) having a nonzero weight in both predictive models.

Figure 3 (middle panel) shows the weights of the cis-SNPs in the prediction model of APOE protein. The effects of most cis-SNPs on APOE protein had the same direction as their effects on LDL, with only four exceptions below the $y = 0$ line. In particular, the effects of rs7412 for LDL and APOE protein were both strong and of the same sign, dominating all the other cis-SNPs. Thus, the resulting association between APOE protein and LDL was positive, as indicated by the positive weighted average of the predictive weights (dashed line). On the other hand, compared to the PWAS results, the directions of the effects of the predictive cis-SNPs on APOE gene expression were approximately equally split between positive and negative, as shown in Figure 3 (bottom panel). Nevertheless, the negative weights outweighed the positive weights, with the greatest contribution from rs4420638 and rs112776896, which has a strong positive association with LDL, but strong negative association with APOE gene expression. Thus the resulting association between LDL and APOE gene expression was negative, as indicated by the negative weighted average of the gene expression predictive weights (dashed line). Overall, due to the small proportion of overlapping nonzero predictive weights and their different directions of effects (Figure S2), APOE protein and gene expression have opposite directions of association with LDL. In addition, similar patterns were observed for LDL with other genes, such as FCGR2B, LILRB2, major histocompatibility complex class I polypeptide-related sequence B (MICB) (Figures S3-S8), as well as for the other lipids (Figures S9-S16, S18-S25, S27-S34).

Comparison of MESA-trained PWAS and GTEx-trained TWAS

The TWAS results obtained from MESA only used gene expression measurements in PBMCs. Since the gene expression levels in some tissues, such as liver, may be more relevant to lipid levels compared to those in other tissues, we extended our TWAS analysis using gene expression data from 49 GTEx tissues. Results of MESA-trained PWAS, MESA-trained TWAS, and GTEx-trained TWAS are compared in Figures 4(a), S17 (a), S26 (a), S35 (a). Overall, for all lipids, the significance and direction of association for PWAS and TWAS are heterogeneous across individual genes. For some genes, the predicted protein and gene expression levels had very consistent directions of association with LDL. For example, for major histocompatibility complex class I polypeptide-related sequence A (MICA), LDL was positively associated with both protein and gene expression in MESA and with gene expression in 43 out of 49 tissues in GTEx. Other examples with similar patterns were observed for MICA with TC and HDL, copine 1 (CPNE1) with TC, and cathepsin B (CTSB) with TG. On the other hand, for some other genes, the protein and gene expression had mixed directions of association. For instance, LDL was positively associated with HP protein levels, but had approximately equal numbers of positive and negative associations with gene expression levels across tissues. Similar inconsistent patterns were observed for HP with TC, APOE with TC and LDL, and apolipoprotein B (APOB) with TC, TG, and HDL.

We next evaluated the correlation patterns of PWAS and TWAS effects when aggregated across all the genes and how this correlation varied across tissues. Figure 4(b) shows the Spearman correlations for each tissue between the signed log p-values for MESA-trained PWAS and GTEx-trained TWAS for LDL. Out of the 49 tissues in GTEx, the PWAS-TWAS correlation

was positive in 47 of them (binomial test p-value: 2.2×10^{-12}). For TC, TG, and HDL, the PWAS-TWAS correlations were positive in 41, 43, 40 tissues, respectively (Figures S17(b), S26(b), and 35(b)). These findings indicate that although the relation between the proteins' and the tissue-specific gene expressions' effects on lipids can be mixed on a single gene, the aggregated correlations between TWAS and PWAS results for lipids across all genes were mostly positive, even if the gene expression predictive models and the protein predictive models were trained using different datasets (i.e. MESA and GTEx).

Discussion

In this work, we conducted PWAS for blood lipids and identified 42 proteins significantly associated with at least one of TC, TG, HDL, and LDL. Several of these proteins, such as tyrosine kinase 2 (TYK2) (Grunert et al., 2011; Qi et al., 2019), MICA and MICB (Bilotta et al. 2019; Yamamoto et al. 2001), IL1RN (Schubert et al., 2022), HP (Braeckman et al., 1999; Schubert et al., 2022), APOE and APOB (Abd El-Aziz & Mohamed, 2016; Schubert et al., 2022; The Emerging Risk Factors Collaboration*, 2009; Weisgraber, 1994), have been previously identified for their association with blood lipids and related diseases. In particular, we found APOE and APOB to be significantly associated with all four lipid traits. Other proteins, such as lymphotoxin alpha (LTA), C-C motif chemokine ligand 17 (CCCL17), and LILRB2, are novel proteins that have not been previously identified for their associations with blood lipids.

Moreover, we conducted TWAS for blood lipids in different tissues and compared the results with the PWAS results. We found that PWAS and TWAS effects for lipids were heterogeneous across tissues and genes, and demonstrated that one cause of this discrepancy is the limited proportion of overlapping SNPs with nonzero predictive weights and their different

291 directions of effect. Nevertheless, when we computed the correlation between the PWAS and
292 TWAS signed log p-values for all the genes in every tissue, the correlation coefficients across
293 various tissues were almost all positive. These results demonstrate that for a single gene, its gene
294 expression's association with lipids may differ from its protein's association with lipids, but when
295 the results for all the genes are aggregated, the lipid TWAS and lipid PWAS results are more
296 consistent.

297 One limitation of our analyses is that not all confounders of omic or lipid levels might
298 have been accounted for. Blood lipids in GWAS can come from a variety of sources, and there
299 could be factors that are correlated with omic levels but not included in the study. Similarly, for
300 training the omic prediction models, although we computed the surrogate values to adjust for
301 unobserved factors that are relevant to the analysis, there could still be factors that are not
302 reflected by the surrogate values and other covariates in the model, such as those related to the
303 collection, processing, and storage of blood or plasma as well as machine artifacts. Furthermore,
304 the set of covariates included in the GWAS might not be the same as those that are adjusted for
305 in the omic prediction models. These potential issues with the covariates and unobserved factors
306 may cause suboptimal accuracy or efficiency in the imputation-based PWAS and TWAS results.

307 A limitation of our tissue-specific GTEx-based TWAS for lipids is the high number of
308 missing gene-tissue pairs, due to their absence in the GTEx data. Imputation methods can be
309 applied to these gene-tissue pairs, so that the missing signed p-values of the tissue-specific gene
310 expression-lipid associations could be imputed, which could provide more insight into the
311 connection between the lipid PWAS and lipid TWAS.

312 Another limitation of our analyses is that for training the omic prediction models,
313 samples from all ancestry groups were used in order to gain power, but in GLGC, most samples

314 are European. This discrepancy in study populations could cause inaccuracy in the analysis
315 (Abdellaoui et al., 2019; Price et al., 2010; Zhang et al., 2020). A multi-ethnic omic dataset with
316 a larger sampler size than MESA will facilitate the training of ancestry-specific, high-power
317 prediction models, and lipid GWAS with more diverse samples will make imputation–based
318 lipid PWAS and lipid TWAS findings more applicable to individuals from non-European
319 populations (Bhattacharya et al., 2020; Keys et al., 2020).
320

321 **Acknowledgements**

322 This research was mainly supported by NIH grant R01HL142023. Whole genome sequencing
323 (WGS) for the Trans-Omics in Precision Medicine (TOPMed) program was supported by the
324 National Heart, Lung and Blood Institute (NHLBI). WGS for "NHLBI TOPMed: Multi-Ethnic
325 Study of Atherosclerosis (MESA)" (phs001416.v1.p1) was performed at the Broad Institute of
326 MIT and Harvard (3U54HG003067-13S1). Centralized read mapping and genotype calling,
327 along with variant quality metrics and filtering were provided by the TOPMed Informatics
328 Research Center (3R01HL-117626-02S1). Phenotype harmonization, data management, sample-
329 identity QC, and general study coordination, were provided by the TOPMed Data Coordinating
330 Center (3R01HL-120393-02S1), and TOPMed MESA Multi-Omics
331 (HHSN2682015000031/HSN26800004). The MESA projects are conducted and supported by
332 the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA
333 investigators. Support for the Multi-Ethnic Study of Atherosclerosis (MESA) projects are
334 conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in
335 collaboration with MESA investigators. Support for MESA is provided by contracts
336 75N92020D00001, HHSN268201500003I, N01-HC-95159, 75N92020D00005, N01-HC-95160,

337 75N92020D00002, N01-HC-95161, 75N92020D00003, N01-HC-95162, 75N92020D00006,
338 N01-HC-95163, 75N92020D00004, N01-HC-95164, 75N92020D00007, N01-HC-95165, N01-
339 HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, UL1-TR-000040, UL1-TR-
340 001079, UL1-TR-001420, UL1TR001881, DK063491, and R01HL105756. The authors thank
341 the other investigators, the staff, and the participants of the MESA study for their valuable
342 contributions. A full list of participating MESA investigators and institutes can be found at
343 http://www.mesa-nhlbi.org. SL is supported by by the Brain Pool Plus (BP+, Brain Pool+)
344 Program through the National Research Foundation of Korea (NRF) funded by the Ministry of
345 Science and ICT (2020H1D3A2A03100666). GMP is supported by NIH grants R01HL142711
346 and R01HL127564. The authors also thank Hae Kyung Im and Alvaro Barbeira for their help
347 with using SPrediXcan.

348

349 **References**

350 Abd El-Aziz, T. A., & Mohamed, R. H. (2016). LDLR, ApoB and ApoE genes polymorphisms
351 and classical risk factors in premature coronary artery disease. *Gene*, *590*(2), 263–269.
352 https://doi.org/10.1016/j.gene.2016.05.032
353 Abdellaoui, A., Hugh-Jones, D., Yengo, L., Kemper, K. E., Nivard, M. G., Veul, L., Holtz, Y.,
354 Zietsch, B. P., Frayling, T. M., Wray, N. R., Yang, J., Verweij, K. J. H., & Visscher, P.
355 M. (2019). Genetic correlates of social stratification in Great Britain. *Nature Human*
356 *Behaviour*, *3*(12), Article 12. https://doi.org/10.1038/s41562-019-0757-5
357 Ahmad, F. B., & Anderson, R. N. (2021). The Leading Causes of Death in the US for 2020.
358 *JAMA*, *325*(18), 1829–1830. https://doi.org/10.1001/jama.2021.5469
359 Arneson, D., Shu, L., Tsai, B., Barrere-Cain, R., Sun, C., & Yang, X. (2017). Multidimensional

Integrative Genomics Approaches to Dissecting Cardiovascular Disease. *Frontiers in Cardiovascular Medicine*, *4*, 8. https://doi.org/10.3389/fcvm.2017.00008

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., Stahl, E. A., Huckins, L. M., Nicolae, D. L., Cox, N. J., & Im, H. K. (2018). Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *Nature Communications*, *9*(1), 1825. https://doi.org/10.1038/s41467-018-03621-1

Bhattacharya, A., García-Closas, M., Olshan, A. F., Perou, C. M., Troester, M. A., & Love, M. I. (2020). A framework for transcriptome-wide association studies in breast cancer in diverse study populations. *Genome Biology*, *21*(1), 42. https://doi.org/10.1186/s13059-020-1942-6

Bild, D. E., Bluemke, D. A., Burke, G. L., Detrano, R., Diez Roux, A. V., Folsom, A. R., Greenland, P., JacobsJr., D. R., Kronmal, R., Liu, K., Nelson, J. C., O'Leary, D., Saad, M. F., Shea, S., Szklo, M., & Tracy, R. P. (2002). Multi-Ethnic Study of Atherosclerosis: Objectives and Design. *American Journal of Epidemiology*, *156*(9), 871–881. https://doi.org/10.1093/aje/kwf113

Bilotta, M. T., Abruzzese, M. P., Molfetta, R., Scarno, G., Fionda, C., Zingoni, A., Soriani, A., Garofalo, T., Petrucci, M. T., Ricciardi, M. R., Paolini, R., Santoni, A., & Cippitelli, M. (2019). Activation of liver X receptor up-regulates the expression of the NKG2D ligands MICA and MICB in multiple myeloma through different molecular mechanisms. *The FASEB Journal*, *33*(8), 9489–9504. https://doi.org/10.1096/fj.201900319R

Braeckman, L., De Bacquer, D., Delanghe, J., Claeys, L., & De Backer, G. (1999). Associations between haptoglobin polymorphism, lipids, lipoproteins and inflammatory variables.

Atherosclerosis, *143*(2), 383–388. https://doi.org/10.1016/S0021-9150(98)00330-X

Brandes, N., Linial, N., & Linial, M. (2020). PWAS: Proteome-wide association study—linking genes and phenotypes by functional variation in proteins. *Genome Biology*, *21*(1), 173. https://doi.org/10.1186/s13059-020-02089-x

Brown, K. M., Diez-Roux, A. V., Smith, J. A., Needham, B. L., Mukherjee, B., Ware, E. B., Liu, Y., Cole, S. W., Seeman, T. E., & Kardia, S. L. R. (2019). Expression of socially sensitive genes: The multi-ethnic study of atherosclerosis. *PLOS ONE*, *14*(4), e0214061. https://doi.org/10.1371/journal.pone.0214061

Burke, G., Lima, J., Wong, N. D., & Narula, J. (2016). The Multiethnic Study of Atherosclerosis. *Global Heart*, *11*(3), 267–268. https://doi.org/10.1016/j.gheart.2016.09.001

Burkhardt, R., Toh, S.-A., Lagor, W. R., Birkeland, A., Levin, M., Li, X., Robblee, M., Fedorov, V. D., Yamamoto, M., Satoh, T., Akira, S., Kathiresan, S., Breslow, J. L., & Rader, D. J. (2010). Trib1 is a lipid- and myocardial infarction–associated gene that regulates hepatic lipogenesis and VLDL production in mice. *The Journal of Clinical Investigation*, *120*(12), 4410–4414. https://doi.org/10.1172/JCI44213

Cao, C., Ding, B., Li, Q., Kwok, D., Wu, J., & Long, Q. (2021). Power analysis of transcriptome-wide association study: Implications for practical protocol choice. *PLOS Genetics*, *17*(2), e1009405. https://doi.org/10.1371/journal.pgen.1009405

Chen, C., Yang, B., Zeng, Z., Yang, H., Liu, C., Ren, J., & Huang, L. (2013). Genetic dissection of blood lipid traits by integrating genome-wide association study and gene expression profiling in a porcine model. *BMC Genomics*, *14*(1), 848. https://doi.org/10.1186/1471-2164-14-848

Clément-Collin, V., Barbier, A., Dergunov, A. D., Visvikis, A., Siest, G., Desmadril, M.,

406 Takahashi, M., & Aggerbeck, L. P. (2006). The structure of human apolipoprotein E2, E3
407 and E4 in solution. 2. Multidomain organization correlates with the stability of apoE
408 structure. *Biophysical Chemistry*, *119*(2), 170–185.
409 https://doi.org/10.1016/j.bpc.2005.07.009
410 de Vries, P. S., Brown, M. R., Bentley, A. R., Sung, Y. J., Winkler, T. W., Ntalla, I., Schwander,
411 K., Kraja, A. T., Guo, X., Franceschini, N., Cheng, C.-Y., Sim, X., Vojinovic, D.,
412 Huffman, J. E., Musani, S. K., Li, C., Feitosa, M. F., Richard, M. A., Noordam, R., …
413 Morrison, A. C. (2019). Multiancestry Genome-Wide Association Study of Lipid Levels
414 Incorporating Gene-Alcohol Interactions. *American Journal of Epidemiology*, *188*(6),
415 1033–1054. https://doi.org/10.1093/aje/kwz005
416 Feng, H., Mancuso, N., Pasaniuc, B., & Kraft, P. (2021). Multitrait transcriptome-wide
417 association study (TWAS) tests. *Genetic Epidemiology*, *45*(6), 563–576.
418 https://doi.org/10.1002/gepi.22391
419 Ferreira, J. A., & Zwinderman, A. H. (2006). On the Benjamini–Hochberg method. *The Annals*
420 *of Statistics*, *34*(4), 1827–1849. https://doi.org/10.1214/009053606000000425
421 Gamazon, E. R., Wheeler, H. E., Shah, K. P., Mozaffari, S. V., Aquino-Michaels, K., Carroll, R.
422 J., Eyler, A. E., Denny, J. C., Nicolae, D. L., Cox, N. J., & Im, H. K. (2015). A gene-
423 based association method for mapping traits using reference transcriptome data. *Nature*
424 *Genetics*, *47*(9), 1091–1098. https://doi.org/10.1038/ng.3367
425 Gold, L., Ayers, D., Bertino, J., Bock, C., Bock, A., Brody, E. N., Carter, J., Dalby, A. B., Eaton,
426 B. E., Fitzwater, T., Flather, D., Forbes, A., Foreman, T., Fowler, C., Gawande, B., Goss,
427 M., Gunn, M., Gupta, S., Halladay, D., … Zichi, D. (2010). Aptamer-Based Multiplexed
428 Proteomic Technology for Biomarker Discovery. *PLOS ONE*, *5*(12), e15004.

https://doi.org/10.1371/journal.pone.0015004

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., Locke, A. E., Marouli, E., Hwang, M. Y., Han, S., Narita, A., Choudhury, A., Bentley, A. R., Ekoru, K., Verma, A., … Willer, C. J. (2021). The power of genetic diversity in genome-wide association studies of lipids. *Nature*, *600*(7890), Article 7890. https://doi.org/10.1038/s41586-021-04064-3

Grunert, T., Leitner, N. R., Marchetti-Deschmann, M., Miller, I., Wallner, B., Radwan, M., Vogl, C., Kolbe, T., Kratky, D., Gemeiner, M., Allmaier, G., Müller, M., & Strobl, B. (2011). A comparative proteome analysis links tyrosine kinase 2 (Tyk2) to the regulation of cellular glucose and lipid metabolism in response to poly(I:C). *Journal of Proteomics*, *74*(12), 2866–2880. https://doi.org/10.1016/j.jprot.2011.07.006

Gusev, A., Ko, A., Shi, H., Bhatia, G., Chung, W., Penninx, B. W. J. H., Jansen, R., de Geus, E. J. C., Boomsma, D. I., Wright, F. A., Sullivan, P. F., Nikkola, E., Alvarez, M., Civelek, M., Lusis, A. J., Lehtimäki, T., Raitoharju, E., Kähönen, M., Seppälä, I., … Pasaniuc, B. (2016). Integrative approaches for large-scale transcriptome-wide association studies. *Nature Genetics*, *48*(3), 245–252. https://doi.org/10.1038/ng.3506

Gusev, A., Mancuso, N., Won, H., Kousi, M., Finucane, H. K., Reshef, Y., Song, L., Safi, A., McCarroll, S., Neale, B. M., Ophoff, R. A., O'Donovan, M. C., Crawford, G. E., Geschwind, D. H., Katsanis, N., Sullivan, P. F., Pasaniuc, B., & Price, A. L. (2018). Transcriptome-wide association study of schizophrenia and chromatin activity yields mechanistic disease insights. *Nature Genetics*, *50*(4), 538–548. https://doi.org/10.1038/s41588-018-0092-1

Hasin, Y., Seldin, M., & Lusis, A. (2017). Multi-omics approaches to disease. *Genome Biology*,

18(1), 83. https://doi.org/10.1186/s13059-017-1215-1

Hoffmann, T. J., Theusch, E., Haldar, T., Ranatunga, D. K., Jorgenson, E., Medina, M. W., Kvale, M. N., Kwok, P.-Y., Schaefer, C., Krauss, R. M., Iribarren, C., & Risch, N. (2018). A large electronic-health-record-based genome-wide study of serum lipids. *Nature Genetics*, *50*(3), Article 3. https://doi.org/10.1038/s41588-018-0064-5

Hu, Y., Li, M., Lu, Q., Weng, H., Wang, J., Zekavat, S. M., Yu, Z., Li, B., Gu, J., Muchnik, S., Shi, Y., Kunkle, B. W., Mukherjee, S., Natarajan, P., Naj, A., Kuzma, A., Zhao, Y., Crane, P. K., Lu, H., & Zhao, H. (2019). A statistical framework for cross-tissue transcriptome-wide association analysis. *Nature Genetics*, *51*(3), 568–576. https://doi.org/10.1038/s41588-019-0345-7

Huang, Y., Ye, H. D., Gao, X., Nie, S., Hong, Q. X., Ji, H. H., Sun, J., Zhou, S. J., Fei, B., Li, K. Q., Zhao, J. K., Wang, Z. P., Xu, M. Q., & Duan, S. W. (2015). Significant interaction of APOE rs4420638 polymorphism with HDL-C and APOA-I levels in coronary heart disease in Han Chinese men. *Genetics and Molecular Research*, *14*(4), 13414–13424. https://doi.org/10.4238/2015.October.28.3

Kannel, W. B., Dawber, T. R., Kagan, A., Revotskie, N., & Stokes, J. (1961). Factors of Risk in the Development of Coronary Heart Disease—Six-Year Follow-up Experience. *Annals of Internal Medicine*, *55*(1), 33–50. https://doi.org/10.7326/0003-4819-55-1-33

Kathiresan, S., Manning, A. K., Demissie, S., D'Agostino, R. B., Surti, A., Guiducci, C., Gianniny, L., Burtt, N. P., Melander, O., Orho-Melander, M., Arnett, D. K., Peloso, G. M., Ordovas, J. M., & Cupples, L. A. (2007). A genome-wide association study for blood lipid phenotypes in the Framingham Heart Study. *BMC Medical Genetics*, *8*(1), S17. https://doi.org/10.1186/1471-2350-8-S1-S17

475 Keys, K. L., Mak, A. C. Y., White, M. J., Eckalbar, W. L., Dahl, A. W., Mefford, J.,
476 Mikhaylova, A. V., Contreras, M. G., Elhawary, J. R., Eng, C., Hu, D., Huntsman, S.,
477 Oh, S. S., Salazar, S., Lenoir, M. A., Ye, J. C., Thornton, T. A., Zaitlen, N., Burchard, E.
478 G., & Gignoux, C. R. (2020). On the cross-population generalizability of gene expression
479 prediction models. *PLOS Genetics*, *16*(8), e1008927.
480 https://doi.org/10.1371/journal.pgen.1008927
481 Kowalski, M. H., Qian, H., Hou, Z., Rosen, J. D., Tapia, A. L., Shan, Y., Jain, D., Argos, M.,
482 Arnett, D. K., Avery, C., Barnes, K. C., Becker, L. C., Bien, S. A., Bis, J. C., Blangero,
483 J., Boerwinkle, E., Bowden, D. W., Buyske, S., Cai, J., … Li, Y. (2019). Use of >100,000
484 NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium whole genome
485 sequences improves imputation quality and detection of rare variant associations in
486 admixed African and Hispanic/Latino populations. *PLOS Genetics*, *15*(12), e1008500.
487 https://doi.org/10.1371/journal.pgen.1008500
488 Kozlitina, J., Smagris, E., Stender, S., Nordestgaard, B. G., Zhou, H. H., Tybjærg-Hansen, A.,
489 Vogt, T. F., Hobbs, H. H., & Cohen, J. C. (2014). Exome-wide association study
490 identifies a TM6SF2 variant that confers susceptibility to nonalcoholic fatty liver disease.
491 *Nature Genetics*, *46*(4), 352–356. https://doi.org/10.1038/ng.2901
492 Lee, S., Sun, W., Wright, F. A., & Zou, F. (2017). An improved and explicit surrogate variable
493 analysis procedure by coefficient adjustment. *Biometrika*, *104*(2), 303–316.
494 https://doi.org/10.1093/biomet/asx018
495 Leon-Mimila, P., Wang, J., & Huertas-Vazquez, A. (2019). Relevance of Multi-Omics Studies in
496 Cardiovascular Diseases. *Frontiers in Cardiovascular Medicine*, *6*, 91.
497 https://doi.org/10.3389/fcvm.2019.00091

Liu, J., Li, X., & Luo, X.-J. (2021). Proteome-wide Association Study Provides Insights Into the Genetic Component of Protein Abundance in Psychiatric Disorders. *Biological Psychiatry*, *90*(11), 781–789. https://doi.org/10.1016/j.biopsych.2021.06.022

Liu, Y., Ding, J., Reynolds, L. M., Lohman, K., Register, T. C., De La Fuente, A., Howard, T. D., Hawkins, G. A., Cui, W., Morris, J., Smith, S. G., Barr, R. G., Kaufman, J. D., Burke, G. L., Post, W., Shea, S., Mccall, C. E., Siscovick, D., Jacobs, D. R., Jr, … Hoeschele, I. (2013). Methylomics of gene expression in human monocytes. *Human Molecular Genetics*, *22*(24), 5065–5074. https://doi.org/10.1093/hmg/ddt356

Lonsdale, J., Thomas, J., Salvatore, M., Phillips, R., Lo, E., Shad, S., Hasz, R., Walters, G., Garcia, F., Young, N., Foster, B., Moser, M., Karasik, E., Gillard, B., Ramsey, K., Sullivan, S., Bridge, J., Magazine, H., Syron, J., … Moore, H. F. (2013). The Genotype-Tissue Expression (GTEx) project. *Nature Genetics*, *45*(6), 580–585. https://doi.org/10.1038/ng.2653

Manichaikul, A., Mychaleckyj, J. C., Rich, S. S., Daly, K., Sale, M., & Chen, W.-M. (2010). Robust relationship inference in genome-wide association studies. *Bioinformatics*, *26*(22), 2867–2873. https://doi.org/10.1093/bioinformatics/btq559

Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A., Hunter, D. J., McCarthy, M. I., Ramos, E. M., Cardon, L. R., Chakravarti, A., Cho, J. H., Guttmacher, A. E., Kong, A., Kruglyak, L., Mardis, E., Rotimi, C. N., Slatkin, M., Valle, D., Whittemore, A. S., … Visscher, P. M. (2009). Finding the missing heritability of complex diseases. *Nature*, *461*(7265), 747–753. https://doi.org/10.1038/nature08494

Musunuru, K., Strong, A., Frank-Kamenetsky, M., Lee, N. E., Ahfeldt, T., Sachs, K. V., Li, X., Li, H., Kuperwasser, N., Ruda, V. M., Pirruccello, J. P., Muchmore, B., Prokunina-

521 Olsson, L., Hall, J. L., Schadt, E. E., Morales, C. R., Lund-Katz, S., Phillips, M. C.,
522 Wong, J., … Rader, D. J. (2010). From noncoding variant to phenotype via SORT1 at the
523 1p13 cholesterol locus. *Nature*, *466*(7307), 714–719. https://doi.org/10.1038/nature09266
524 Myers, L., & Sirois, M. J. (2006). Spearman Correlation Coefficients, Differences between. In
525 *Encyclopedia of Statistical Sciences*. American Cancer Society.
526 https://doi.org/10.1002/0471667196.ess5050.pub2
527 Pilia, G., Chen, W.-M., Scuteri, A., Orrú, M., Albai, G., Dei, M., Lai, S., Usala, G., Lai, M., Loi,
528 P., Mameli, C., Vacca, L., Deiana, M., Olla, N., Masala, M., Cao, A., Najjar, S. S.,
529 Terracciano, A., Nedorezov, T., … Schlessinger, D. (2006). Heritability of
530 Cardiovascular and Personality Traits in 6,148 Sardinians. *PLOS Genetics*, *2*(8), e132.
531 https://doi.org/10.1371/journal.pgen.0020132
532 Price, A. L., Zaitlen, N. A., Reich, D., & Patterson, N. (2010). New approaches to population
533 stratification in genome-wide association studies. *Nature Reviews Genetics*, *11*(7), 459–
534 463. https://doi.org/10.1038/nrg2813
535 Qi, W., Zhou, L., Zhao, T., Ding, S., Xu, Q., Han, X., Zhao, Y., Song, X., Zhao, T., Zhang, X., &
536 Ye, L. (2019). Effect of the TYK-2/STAT-3 pathway on lipid accumulation induced by
537 mono-2-ethylhexyl phthalate. *Molecular and Cellular Endocrinology*, *484*, 52–58.
538 https://doi.org/10.1016/j.mce.2019.01.012
539 Raffield, L. M., Dang, H., Pratte, K. A., Jacobson, S., Gillenwater, L. A., Ampleford, E.,
540 Barjaktarevic, I., Basta, P., Clish, C. B., Comellas, A. P., Cornell, E., Curtis, J. L.,
541 Doerschuk, C., Durda, P., Emson, C., Freeman, C. M., Guo, X., Hastie, A. T., Hawkins,
542 G. A., … Consortium, N. T.-O. for P. M. (TOPMed). (2020). Comparison of Proteomic
543 Assessment Methods in Multiple Cohort Studies. *PROTEOMICS*, *20*(12), 1900278.

https://doi.org/10.1002/pmic.201900278

Ramazzotti, D., Lal, A., Wang, B., Batzoglou, S., & Sidow, A. (2018). Multi-omic tumor data reveal diversity of molecular mechanisms that correlate with survival. *Nature Communications*, *9*(1), 4453. https://doi.org/10.1038/s41467-018-06921-8

Roger, V. L., Go, A. S., Lloyd-Jones, D. M., Adams, R. J., Berry, J. D., Brown, T. M., Carnethon, M. R., Dai, S., de Simone, G., Ford, E. S., Fox, C. S., Fullerton, H. J., Gillespie, C., Greenlund, K. J., Hailpern, S. M., Heit, J. A., Ho, P. M., Howard, V. J., Kissela, B. M., … American Heart Association Statistics Committee and Stroke Statistics Subcommittee. (2011). Heart disease and stroke statistics--2011 update: A report from the American Heart Association. *Circulation*, *123*(4), e18–e209. https://doi.org/10.1161/CIR.0b013e3182009701

Schubert, R., Geoffroy, E., Gregga, I., Mulford, A. J., Aguet, F., Ardlie, K., Gerszten, R., Clish, C., Berg, D. V. D., Taylor, K. D., Durda, P., Johnson, W. C., Cornell, E., Guo, X., Liu, Y., Tracy, R., Conomos, M., Blackwell, T., Papanicolaou, G., … Wheeler, H. E. (2022). Protein prediction for trait mapping in diverse populations. *PLOS ONE*, *17*(2), e0264341. https://doi.org/10.1371/journal.pone.0264341

Taliun, D., Harris, D. N., Kessler, M. D., Carlson, J., Szpiech, Z. A., Torres, R., Taliun, S. A. G., Corvelo, A., Gogarten, S. M., Kang, H. M., Pitsillides, A. N., LeFaive, J., Lee, S., Tian, X., Browning, B. L., Das, S., Emde, A.-K., Clarke, W. E., Loesch, D. P., … Abecasis, G. R. (2021). Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature*, *590*(7845), Article 7845. https://doi.org/10.1038/s41586-021-03205-y

Tejedor, M. T., Garcia-Sobreviela, M. P., Ledesma, M., & Arbones-Mainar, J. M. (2014). The Apolipoprotein E Polymorphism rs7412 Associates with Body Fatness Independently of

Plasma Lipids in Middle Aged Men. *PLoS ONE*, *9*(9), e108605. https://doi.org/10.1371/journal.pone.0108605

The Emerging Risk Factors Collaboration*. (2009). Major Lipids, Apolipoproteins, and Risk of Vascular Disease. *JAMA*, *302*(18), 1993–2000. https://doi.org/10.1001/jama.2009.1619

Veturi, Y., Lucas, A., Bradford, Y., Hui, D., Dudek, S., Theusch, E., Verma, A., Miller, J. E., Kullo, I., Hakonarson, H., Sleiman, P., Schaid, D., Stein, C. M., Edwards, D. R. V., Feng, Q., Wei, W.-Q., Medina, M. W., Krauss, R. M., Hoffmann, T. J., … Ritchie, M. D. (2021). A unified framework identifies new links between plasma lipids and diseases from electronic medical records across large-scale cohorts. *Nature Genetics*, *53*(7), Article 7. https://doi.org/10.1038/s41588-021-00879-y

Visscher, P. M., Brown, M. A., McCarthy, M. I., & Yang, J. (2012). Five Years of GWAS Discovery. *The American Journal of Human Genetics*, *90*(1), 7–24. https://doi.org/10.1016/j.ajhg.2011.11.029

Wainberg, M., Sinnott-Armstrong, N., Mancuso, N., Barbeira, A. N., Knowles, D. A., Golan, D., Ermel, R., Ruusalepp, A., Quertermous, T., Hao, K., Björkegren, J. L. M., Im, H. K., Pasaniuc, B., Rivas, M. A., & Kundaje, A. (2019). Opportunities and challenges for transcriptome-wide association studies. *Nature Genetics*, *51*(4), 592–599. https://doi.org/10.1038/s41588-019-0385-z

Wang, S.-B., Feng, J.-Y., Ren, W.-L., Huang, B., Zhou, L., Wen, Y.-J., Zhang, J., Dunwell, J. M., Xu, S., & Zhang, Y.-M. (2016). Improving power and accuracy of genome-wide association studies via a multi-locus mixed linear model methodology. *Scientific Reports*, *6*(1), 19444. https://doi.org/10.1038/srep19444

Weisgraber, K. H. (1994). Apolipoprotein E: Structure-Function Relationships. In C. B.

590 Anfinsen, J. T. Edsall, F. M. Richards, & D. S. Eisenberg (Eds.), *Advances in Protein*
591 *Chemistry* (Vol. 45, pp. 249–302). Academic Press. https://doi.org/10.1016/S0065-
592 3233(08)60642-7
593 Willer, C. J., & Mohlke, K. L. (2012). Finding genes and variants for lipid levels after genome-
594 wide association analysis. *Current Opinion in Lipidology*, *23*(2), 98–103.
595 https://doi.org/10.1097/MOL.0b013e328350fad2
596 Willer, C. J., Sanna, S., Jackson, A. U., Scuteri, A., Bonnycastle, L. L., Clarke, R., Heath, S. C.,
597 Timpson, N. J., Najjar, S. S., Stringham, H. M., Strait, J., Duren, W. L., Maschio, A.,
598 Busonero, F., Mulas, A., Albai, G., Swift, A. J., Morken, M. A., Narisu, N., … Abecasis,
599 G. R. (2008). Newly identified loci that influence lipid concentrations and risk of
600 coronary artery disease. *Nature Genetics*, *40*(2), 161–169. https://doi.org/10.1038/ng.76
601 Willer, C. J., Schmidt, E. M., Sengupta, S., Peloso, G. M., Gustafsson, S., Kanoni, S., Ganna, A.,
602 Chen, J., Buchkovich, M. L., Mora, S., Beckmann, J. S., Bragg-Gresham, J. L., Chang,
603 H.-Y., Demirkan, A., Den Hertog, H. M., Do, R., Donnelly, L. A., Ehret, G. B., Esko, T.,
604 … Global Lipids Genetics Consortium. (2013). Discovery and refinement of loci
605 associated with lipid levels. *Nature Genetics*, *45*(11), 1274–1283.
606 https://doi.org/10.1038/ng.2797
607 Wingo, A. P., Liu, Y., Gerasimov, E. S., Gockley, J., Logsdon, B. A., Duong, D. M., Dammer,
608 E. B., Robins, C., Beach, T. G., Reiman, E. M., Epstein, M. P., De Jager, P. L., Lah, J. J.,
609 Bennett, D. A., Seyfried, N. T., Levey, A. I., & Wingo, T. S. (2021). Integrating human
610 brain proteomes with genome-wide association data implicates new proteins in
611 Alzheimer's disease pathogenesis. *Nature Genetics*, *53*(2), 143–146.
612 https://doi.org/10.1038/s41588-020-00773-z

Wingo, T. S., Gerasimov, E. S., Liu, Y., Duong, D. M., Vattathil, S. M., Lori, A., Gockley, J., Breen, M. S., Maihofer, A. X., Nievergelt, C. M., Koenen, K. C., Levey, D. F., Gelernter, J., Stein, M. B., Ressler, K. J., Bennett, D. A., Levey, A. I., Seyfried, N. T., & Wingo, A. P. (2022). Integrating human brain proteomes with genome-wide association data implicates novel proteins in post-traumatic stress disorder. *Molecular Psychiatry*, *27*(7), Article 7. https://doi.org/10.1038/s41380-022-01544-4

Wingo, T. S., Liu, Y., Gerasimov, E. S., Gockley, J., Logsdon, B. A., Duong, D. M., Dammer, E. B., Lori, A., Kim, P. J., Ressler, K. J., Beach, T. G., Reiman, E. M., Epstein, M. P., De Jager, P. L., Lah, J. J., Bennett, D. A., Seyfried, N. T., Levey, A. I., & Wingo, A. P. (2021). Brain proteome-wide association study implicates novel proteins in depression pathogenesis. *Nature Neuroscience*, *24*(6), 810–817. https://doi.org/10.1038/s41593-021-00832-6

Wu, B.-S., Chen, S.-F., Huang, S.-Y., Ou, Y.-N., Deng, Y.-T., Chen, S.-D., Dong, Q., & Yu, J.-T. (2022). Identifying causal genes for stroke via integrating the proteome and transcriptome from brain and blood. *Journal of Translational Medicine*, *20*(1), 181. https://doi.org/10.1186/s12967-022-03377-9

Wu, H., Huang, Q., Yu, Z., Wu, H., & Zhong, Z. (2020). The SNPs rs429358 and rs7412 of APOE gene are association with cerebral infarction but not SNPs rs2306283 and rs4149056 of SLCO1B1 gene in southern Chinese Hakka population. *Lipids in Health and Disease*, *19*(1), 202. https://doi.org/10.1186/s12944-020-01379-4

Xiao, H., Bartoszek, K., & Lio', P. (2018). Multi–omic analysis of signalling factors in inflammatory comorbidities. *BMC Bioinformatics*, *19*(15), 439. https://doi.org/10.1186/s12859-018-2413-x

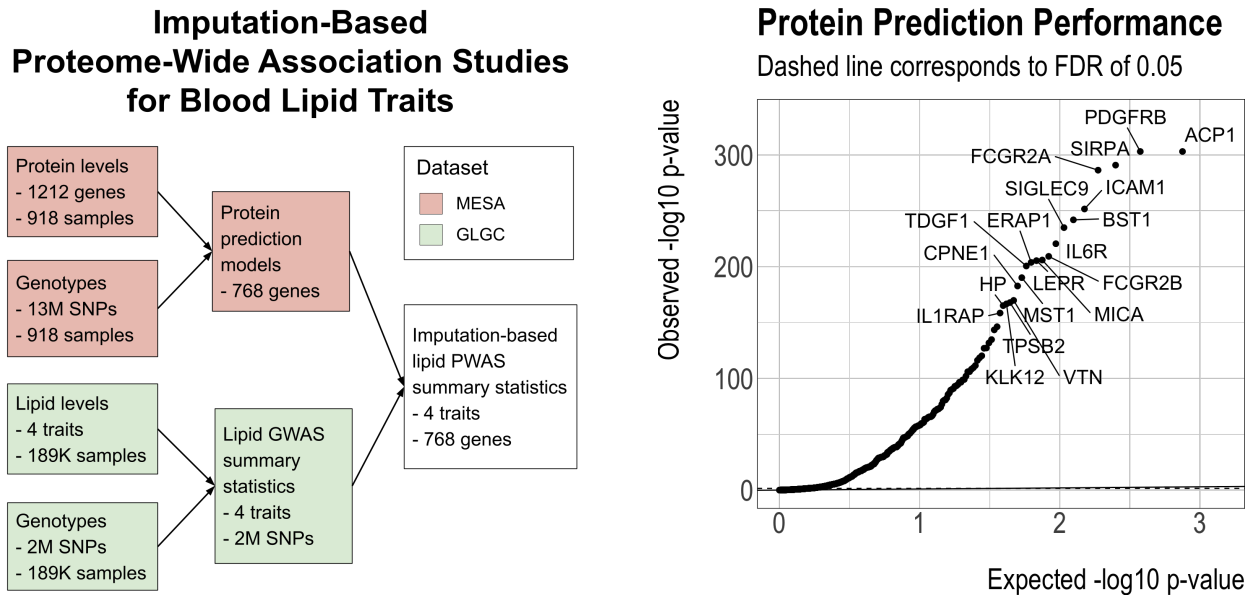
636 Yamamoto, K., Fujiyama, Y., Andoh, A., Bamba, T., & Okabe, H. (2001). Oxidative stress
637 increases MICA and MICB gene expression in the human colon carcinoma cell line
638 (CaCo-2). *Biochimica et Biophysica Acta (BBA) - General Subjects*, *1526*(1), 10–12.
639 https://doi.org/10.1016/S0304-4165(01)00099-X
640 Yang, T., Wu, C., Wei, P., & Pan, W. (2020). Integrating DNA sequencing and transcriptomic
641 data for association analyses of low-frequency variants and lipid traits. *Human Molecular*
642 *Genetics*, *29*(3), 515–526. https://doi.org/10.1093/hmg/ddz314
643 Zhang, D., Dey, R., & Lee, S. (2020). Fast and robust ancestry prediction using principal
644 component analysis. *Bioinformatics*, *36*(11), 3439–3446.
645 https://doi.org/10.1093/bioinformatics/btaa152
646 Zhao, B., Shan, Y., Yang, Y., Yu, Z., Li, T., Wang, X., Luo, T., Zhu, Z., Sullivan, P., Zhao, H.,
647 Li, Y., & Zhu, H. (2021). Transcriptome-wide association analysis of brain structures
648 yields insights into pleiotropy with complex neuropsychiatric traits. *Nature*
649 *Communications*, *12*(1), 2878. https://doi.org/10.1038/s41467-021-23130-y
650 Zhen, J., Huang, X., Van Halm-Lutterodt, N., Dong, S., Ma, W., Xiao, R., & Yuan, L. (2017).
651 ApoE rs429358 and rs7412 Polymorphism and Gender Differences of Serum Lipid
652 Profile and Cognition in Aging Chinese Population. *Frontiers in Aging Neuroscience*, *9*.
653 https://www.frontiersin.org/article/10.3389/fnagi.2017.00248
654 Zhu, H., & Zhou, X. (2020). Transcriptome-wide association studies: A view from Mendelian
655 randomization. *Quantitative Biology*. https://doi.org/10.1007/s40484-020-0207-4
656 Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of*
657 *the Royal Statistical Society: Series B (Statistical Methodology)*, *67*(2), 301–320.
658 https://doi.org/10.1111/j.1467-9868.2005.00503.x

Tables and Figures

Table 1: PWAS results for proteins that are significantly (FDR ≤ 0.05) associated with at least one blood lipid trait. Next to the PWAS (P) summary statistics of every protein, the TWAS (T) summary statistics of the same gene are also displayed. Inside each cell is the $\log_{10}$ p-value, followed by the direction of association in parentheses.

	TC		TG		HDL		LDL	
Gene	P	T	P	T	P	T	P	T
APOE	406(+)	16(-)	14(-)	.	37(-)	4(+)	850(+)	19(-)
TYK2	43(-)	.	.	.	.	.	53(-)	.
HP	45(-)	3(-)	4(-)	.	.	.	47(-)	3(-)
LTA	.	.	13(-)	.	.	.	.	.
MICB	12(+)	3(-)	9(+)	.	.	.	5(+)	.
CCL17	.	.	.	.	10(+)	9(+)	.	.
LILRB2	4(-)	4(+)	.	.	10(-)	8(+)	.	.
RBM39	8(-)	.	.	.	.	.	5(-)	.
PCSK7	4(-)	3(-)	8(-)	4(-)	.	.	.	.
FN1	7(+)	.	.	.	.	.	8(+)	.
RSPO3	.	.	6(+)	.	8(-)	.	.	.
PDPK1	.	.	7(-)	.	4(+)	.	.	.
MICA	6(-)	6(-)	.	.	4(-)	3(-)	4(-)	4(-)
IL1RN	6(+)	.	.	.	.	.	3(+)	.
MMP9	.	.	5(+)	.	4(-)	.	.	.
FCGR2A	5(-)	6(-)	.	.	.	.	5(-)	6(-)
SERPINA1	4(-)	.	.	.	.	.	5(-)	.
ICAM5	5(-)	.	.	.	.	.	4(-)	.
EPHB6	.	.	.	.	4(-)	.	.	.
CTSB	.	.	4(+)	6(+)	.	.	.	.
HAVCR2	4(-)	.	.	.	.	.	3(-)	.
MET	.	.	4(+)	.	.	.	.	.
FCGR2B	3(-)	5(+)	.	.	.	.	4(-)	4(+)
ICAM3	4(+)	.	.	.	.	.	.	.
CPNE1	4(+)	6(+)	.	.	.	.	.	.
COLEC11	4(+)	.	.	.	.	.	.	.
AIF1	.	.	.	.	4(-)	.	.	.
HSPA1A	.	.	4(-)	.	.	.	.	.
TYRO3	.	.	3(+)	.	3(-)	.	.	.
MMP1	3(-)	3(-)	.	.	.	.	3(-)	.
SHBG	.	.	.	.	3(+)	.	.	.
VWF	.	.	.	.	.	.	3(+)	.
AGRP	.	.	.	.	3(+)	.	.	.
TKT	.	.	.	.	3(+)	.	.	.
CSF3	4(-)	.	.	.	8(-)	.	.	.
NAPA	.	.	.	.	3(-)	.	.	.
APOB	16(-)	.	10(-)	.	9(+)	.	20(-)	.
F2	.	.	5(-)	.	13(+)	.	.	.
HGFAC	6(-)	.	6(-)	.	.	.	.	.
MDK	5(+)	.	.	.	.	.	.	.
BCAM	.	.	3(-)	.	.	.	.	.
CFC1	.	.	4(-)	.	.	.	.	.

Figure 1: Imputation-based proteome-wide association studies (PWAS) for lipids. Panels (b) and (c): the solid line is the identity line, while the dashed line represents the false discovery rate (FDR) threshold of 0.05.



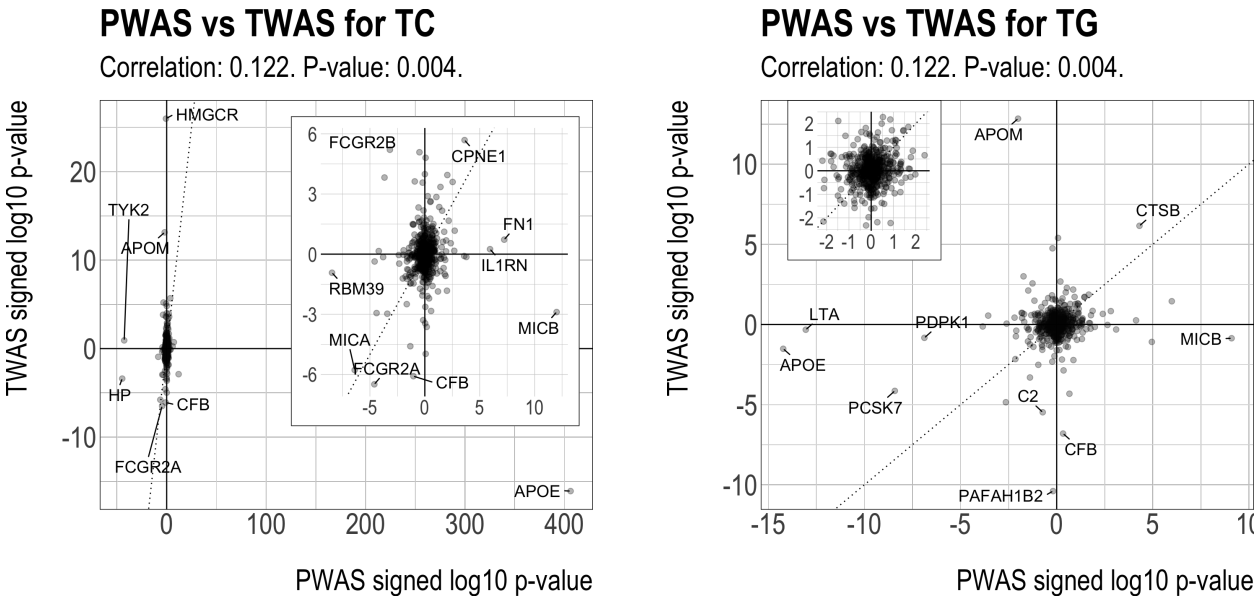
(a)

(b)

(c)

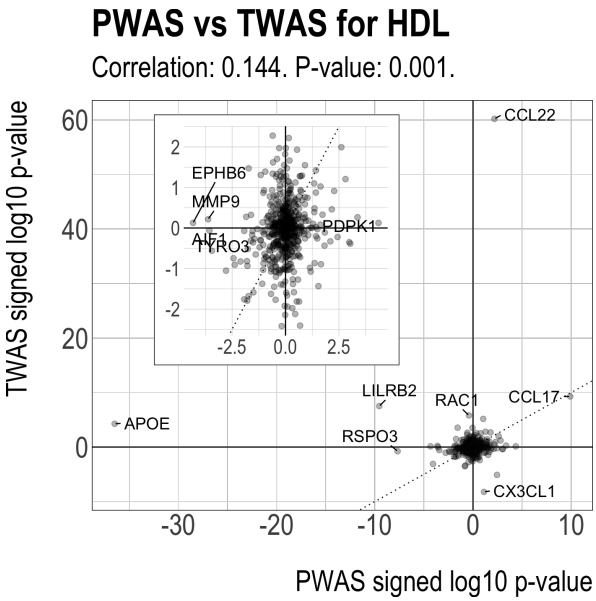
(d)

Figure 2: Comparison of PWAS and TWAS results for lipids. The subplot inside each panel shows zoomed results.

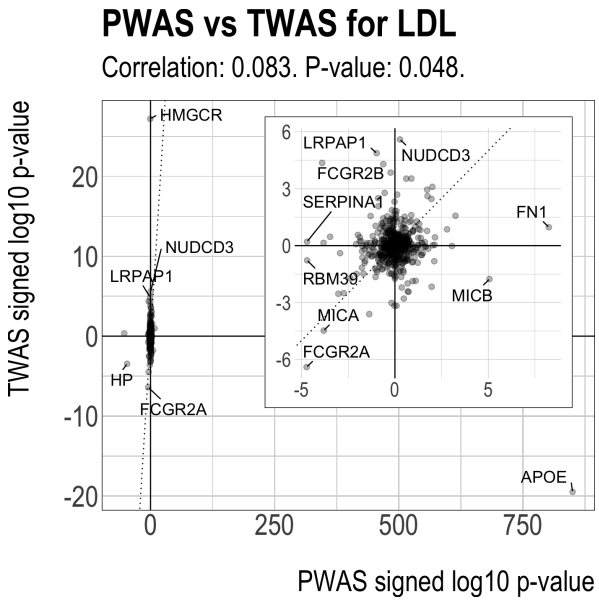


(a)

(b)



(c)



(d)

Figure 3: GWAS for LDL and prediction models for APOE's protein and gene expression levels. The reference and alternative alleles for GWAS and the predictive models have been aligned and reordered so that all the SNPs have positive GWAS effects. In the center and bottom panels, the size of the circles indicates the SNP's GWAS z-score. The z-scores are used to compute the weighted average of the model weights (dashed line), which has the same sign as and is proportional to the predicted effect of protein or gene expression on the GWAS outcome.

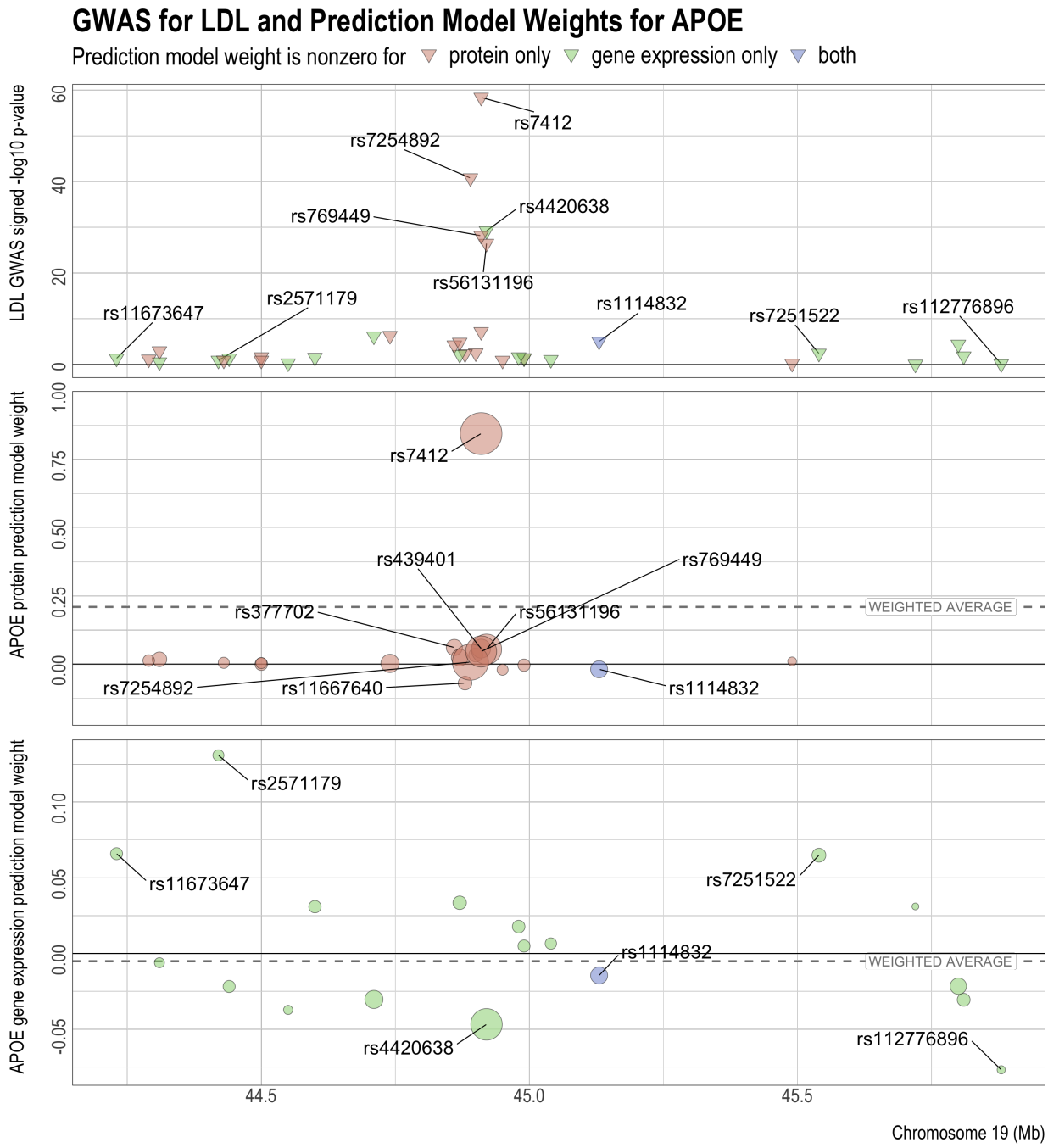
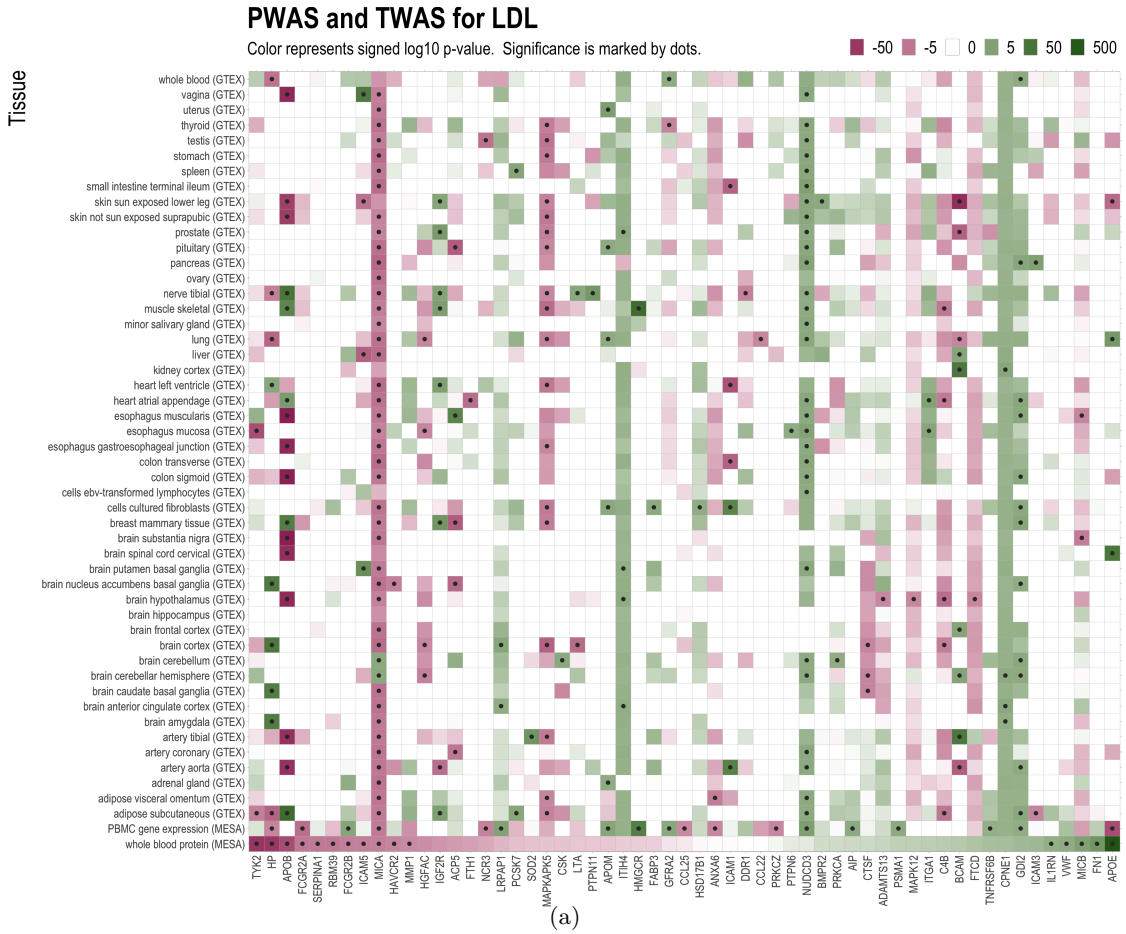
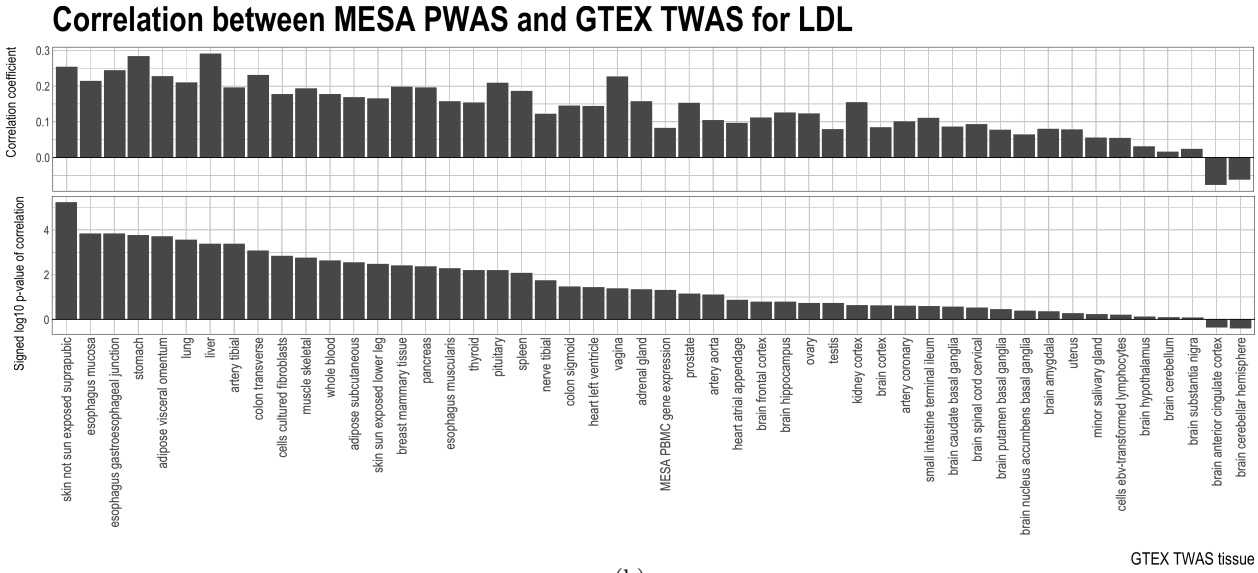


Figure 4: Comparison of MESA PBMC PWAS, MESA PBMC TWAS, and GTEx tissue-specific TWAS results for LDL. Panel (a): signed log p-value and significance of association. Missing values are shown in white. Significance of association is deteremined by the false discovery rate (FDR) threshold of 0.05. Only genes with at least one significant association with LDL are displayed. Panel (b): correlation between signed log p-values of MESA PBMC PWAS and signed log p-values of each GTEx tissue-specific TWAS (i.e. the correlation between the bottom row and every other row of the grid in Panel (a)).



(a)



(b)