Manifestations of Alzheimer's Disease Genetic Risk in the Blood: A Cross-Sectional Multi-Omic Analysis in Healthy Adults Aged 18-90+

Laura M Heath (laura.heath@sagebase.org)
Sage Bionetworks https://orcid.org/0000-0002-0318-2773

John C. Earls
Institute for Systems Biology

Andrew T. Magis
Institute for Systems Biology

Sergey A. Kornilov
Institute for Systems Biology

Jennifer C. Lovejoy
Institute for Systems Biology

Cory C. Funk
Institute for Systems Biology

Noa Rappaport
Institute for Systems Biology

Benjamin A. Logsdon
Sage Bionetworks

Lara M. Mangravite
Sage Bionetworks

Brian W. Kunkle
John P Hussman Institute for Human Genomics

Eden R. Martin
John P Hussman Institute for Human Genomics

Adam C. Naj
University of Pennsylvania

Nilüfer Ertekin-Taner
Mayo Clinic Department of Neurology

Todd E. Golde
University of Florida McKnight Brain Institute

Leroy Hood
Institute for Systems Biology

Nathan D. Price

Institute for Systems Biology

Research article

Keywords: Alzheimer's Disease, Genetics, Genetic Association Study, Deep Phenotyping

Posted Date: November 29th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-1108842/v1

Abstract

Background: Genetics play an important role in late-onset Alzheimer's Disease (AD) etiology and dozens of genetic variants have been implicated in AD risk through large-scale GWAS meta-analyses. However, the precise mechanistic effects of most of these variants have yet to be determined. Deeply phenotyped cohort data can reveal physiological changes associated with genetic risk for AD across an age spectrum that may provide clues to the biology of the disease.

Methods: We utilized over 2000 high-quality quantitative measurements obtained from blood of 2831 cognitively normal adult clients of a consumer-based scientific wellness company, each with CLIA-certified whole-genome sequencing data. Measurements included: clinical laboratory blood tests, targeted chip-based proteomics, and metabolomics. We performed a phenome-wide association study utilizing this diverse blood marker data and 25 known AD genetic variants, adjusting for sex, age, vendor (for clinical labs), and the first four genetic principal components; sex-SNP interactions were also assessed.

Results: We observed statistically significant SNP-analyte associations for five genetic variants after correction for multiple testing (for SNPs in or near *NYAP1*, *ABCA7*, *INPP5D*, and *APOE*), with effects detectable from early adulthood. The *ABCA7* SNP and the *APOE2* and *APOE4* encoding alleles were associated with lipid variability, as seen in previous studies; in addition, six novel proteins were associated with the e2 allele. The most statistically significant finding was between the *NYAP1* variant and PILRA and PILRB protein levels, supporting previous functional genomic studies in the identification of a putative causal variant within the *PILRA* gene. Sex modified the effects of four genetic variants, with multiple interrelated immune-modulating effects associated with the *PICALM* variant. In post-hoc analysis, sex-stratified GWAS results from an independent AD case-control meta-analysis supported sex-specific disease effects of the *PICALM* variant, highlighting the importance of sex as a biological variable.

Conclusions: Known AD genetic variation influenced lipid metabolism and immune response systems in a population of non-AD individuals, with associations observed from early adulthood onward. Further research is needed to determine whether and how these effects are implicated in early-stage biological pathways to AD. These analyses aim to complement ongoing work on the functional interpretation of AD-associated genetic variants.

Background

The rapidly decreasing cost of genomics paired with technological advances in the generation of multi-omic data has resulted in multiple datasets of deeply phenotyped individuals with a variety of health outcomes (1–3). The data collected in these studies have the potential to yield important insights into potential molecular drivers of health observable in the blood periphery. The present study seeks to leverage a unique and relatively large set of multi-omic, deep-phenotyping data to shed light on genetic pathways to late-onset Alzheimer's disease (AD) by assessing differences in ~2000 analytes in the blood

that show association with known genetic risk variants for AD. Coupled with high-dimensional data sets, this approach has the potential to yield clues into gene pleiotropy, disease processes, and possible early-intervention strategies, which are critically important given the essentially untreatable nature of late-stage Alzheimer's disease once significant brain deterioration has occurred.

Genetic variation plays a substantial role in AD risk, with twin studies estimating AD heritability at 58-79% (4). While the emergence of recent large-scale consortia efforts has facilitated well-powered meta-analyses of genome-wide association studies (GWAS) to identify multiple common variants with small effect sizes (5, 6), the research community is still untangling exactly how this genetic variation influences disease risk. Functional genomics studies are beginning to identify likely genetic pathways to disease with the aid of transcriptomic, epigenomic, and endophenotypic data (7–10). So far, genetic and multi-omic studies of AD studies have largely focused on older individuals with either clinically diagnosed AD or milder symptoms of cognitive decline, despite research pointing to highly variable AD pathobiology that occurs on a spectrum, and begins decades before clinical symptoms onset (11).

In this study, we leveraged the results from a large-scale GWAS meta-analysis (5) alongside data from a deeply phenotyped wellness cohort to investigate the physiological periphery effects of genetic risk for AD in individuals without known cognitive impairment, at all ages. We undertook an agnostic approach by adopting a phenome-wide association study (PheWAS) design (12). By examining how genetic variation impacts 2008 analytes in the blood of 2831 individuals, we sought to complement previous functional genomics studies as well as potentially reveal new testable hypotheses for future studies. In addition, we tested for associations between a polygenic risk score (PGRS) for AD and blood analytes to determine if a relative burden of genetic risk might impact observable changes in the blood, and we assessed for effect modification of genetic risk by sex.

Methods

Population

The Institute for Systems Biology, through partnership with their spin-out company Arivale, has access to a wealth of data collected from subscribers in the commercially available (now closed) Arivale Scientific Wellness program (3, 13), from July 2015 to May 2019. In brief, participants in the Arivale program were assigned a health coach upon joining the program, who then utilized data from clinical blood assays and detailed health-history and behavioral questionnaires to personalize health advice and management of health goals. Participants consented to their de-identified data being used for research purposes.

Blood-derived analytes

We identified 2831 individuals with whole genome sequencing (WGS) and at least one class of blood-derived analyte, described as follows. For each participant, fasting clinical blood laboratory tests were measured upon joining the program. Blood samples were collected at either local facilities hosted by LabCorp (North Carolina, USA) or Quest Diagnostics (New Jersey, USA). Whole genome sequencing was

performed on DNA extracted from whole blood with library preparation using the Illumina TruSeq Nano Library prep kit and sequenced using Illumina HiSeq X, PE-150, target 30X coverage at a single Clinical Laboratory Improvement Amendmnets (CLIA)-approved sequencing laboratory. At the baseline blood draw, 2827 of the 2831 individuals with sequenced whole genomes had up to 63 fasting clinical blood lab tests. Clinical blood tests included standard markers for cardiometabolic health (lipid levels), diabetes, inflammation, kidney and liver function, nutrition (vitamins and minerals), and blood cell counts. Frozen plasma samples were also sent to Olink (Olink Bioscience, Sweden) for targeted proteomics assays based on Olink's proximity extension assay technique. Up to 2694 of these participants had quantitative proteomic data on 274 proteins from three Olink panels (Cardiovascular II, Cardiovascular III, and Inflammation panels). An additional 919 proteins (from 10 additional panels available at Olink at the time) were obtained from a subsample of 354 individuals, in which *Apolipoprotein E* (*APOE*) e2/e2 and *APOE* e4/e4 genotypes were overrepresented. Since multiple batches were performed, previously generated pooled control samples were run with each batch and used for batch correction and multiple control samples were included on each plate. Aliquots of frozen plasma samples were sent to Metabolon, Inc. to conduct metabolomics assays using the Metabolon HD4 discovery platform. Up to 1855 of the participants had data from 754 metabolites. Relative concentration values were reported for each metabolite. For all analytes, only analytes with <20% missing were included in analyses.

SNP selection

We selected 25 common and somewhat-rare (>1% allele frequency) single nucleotide polymorphisms (SNPs) that were significantly associated with AD in a large-scale meta-analysis based on updated data from the International Genomics of Alzheimer's Project (IGAP) (5). In addition to these variants, we also included the SNP coding for *APOE* e2 (rs7412). The 25 SNPs were linked to 24 genes (two SNPs in *APOE*), as detailed in **Table S1**.

Polygenic risk score calculation for AD: PGRS for age-associated AD risk was computed using summary statistics from the initial IGAP-driven GWAS meta-analysis (6). Briefly, the set of SNPs included in the PGS was determined as follows. The Benjamini-Hochberg (14) procedure was applied to the p-values for all SNPs tested in the GWAS to account for multiple testing by controlling the false discovery rate (FDR) at a 5% level. This FDR-filtered set of SNPs was then further pruned using linkage disequilibrium (LD): pairs of SNPs in close proximity capturing highly correlated information ($r^2 > 0.2$) were identified, and the SNP with the smaller p-value in the pair was kept; this was repeated until all remaining SNPs were mutually uncorrelated ($r^2 < 0.2$ for all pairs). The PGRS for each individual was then calculated by summing up the published effect size for each selected SNP multiplied by the number of effect alleles the individual carried for that SNP, across all of the selected SNPs. Missing genotypes were mean imputed using the effect allele frequency.

Statistical analysis

Following a phenome-wide association study approach (PheWAS) (12, 15), the primary model for each SNP used linear regression, with genotype (0, 1, or 2, with 0 indicating homozygosity for the major allele and 2 indicating homozygosity for the minor allele) as the predictor, and each continuous quantitative analyte as the dependent variable. Clinical lab and metabolite values were natural log transformed to account for right skewness and outliers, with +1 added to each natural log transformation to prevent zero values. Proteomic quantities were presented as normalized protein expression (NPX), Olink's arbitrary unit, which is in log2 scale. Genetic ancestry was represented by principal components (PCs) 1-4, calculated using previously described methods (16). All SNP models were adjusted for age, sex, genetic ancestry PCs 1-4, and vendor identification for the clinical labs. Secondary models tested effect modification by sex by including a gene x sex interaction term in the models. We accounted for multiple comparisons by applying the Benjamini-Hochberg method (14) at alpha=0.05 on a per-SNP basis and applied to the main effect of genotype in the primary models, while we set B-H alpha=0.1 of the sex-SNP interaction term as the threshold for the gene x sex interaction models. The FDR rate took into account testing for all 2008 possible analytes, with the understanding that this adjustment was highly conservative given a high degree of correlation among multiple groups of analytes, and the fact that some analytes were sampled in only a subset of individuals. Both raw and adjusted p-values are reported.

We also repeated the primary PheWAS approach with participants stratified by self-identified race, due to evidence for variable genetic risk for cognitive outcomes between non-Hispanic white (hereafter referred to as "white") and non-white populations (17, 18). Unfortunately, due to small numbers of individuals in specific non-white racial and ethnic groups, which become vanishingly small when accounting for allele frequency and numbers of available samples (**Table 1**), we were not able to assess genetic risk effects in individual groups with statistical rigor and had to group all non-white participants into one stratum for analysis. The stratified white and non-white group analyses serve as an investigation into whether our primary results reflected the majority-white makeup of the Arivale population. PheWAS was applied as described above, with FDR to account for multiple comparisons.

To visualize genotype-analyte associations across adulthood, we created boxplots of the log-transformed analyte values by genotype, stratified by age group (by decade, from 18-29 to 70 and over). One-way analysis of variance (ANOVA) was used to test whether there was an overall difference between genotypes within each age group. All statistical analyses were performed in R v3.5.1 (https://www.R-project.org/).

In post-hoc exploratory analysis focused on the SNP in the *PICALM* (Phosphatidylinositol Binding Clathrin Assembly Protein) locus (rs3851179), sex-stratified and sex-interaction analyses was performed on 12,324 cases (57.7% female) and 11,453 controls (59.9% female) of European ancestry from the Alzheimer's Disease Genetics Consortium (ADGC) (see Supplementary Table 4 for dataset details). Datasets were imputed to the Haplotype Reference Consortium (HRC) (19) panel using the Michigan Imputation Server

Table 1. Baseline self-reported characteristics of Arivale participants with available whole-genome sequences

Characteristic[a]	N=2831
Age, mean (sd)	47.0 (12.0)
Female, n (%)	1719 (60.7)
Nonwhite[b], n (%) (n=2725)	597 (21.9)
Afro-Caribbean	1 (<0.1)
American Indian or Alaska Native	5 (0.2)
Ashkenazi Jewish	49 (1.8)
Asian	84 (3.1)
Black or African American	64 (2.3)
East Asian	91 (3.3)
Hispanic Latino or Spanish origin	120 (4.4)
Middle Eastern or North African	18 (0.7)
Native Hawaiian or other Pacific Islander	17 (0.6)
Sephardic Jewish	4 (0.1)
South Asian	79 (2.9)
White	2128 (78.1)
Other	65 (2.4)
BMI, mean (sd) (n=2750)	27.9 (6.4)
Obese[c], n (%) (n=2750)	802 (29.2)
Moderate activity ≥ 3x/wk, n (%) (n=2275)	1460 (64.2)
Vigorous activity ≥ 3x/wk, n (%) (n=2271)	697 (30.7)
Ever smoke, n (%) (n=2207)	565 (25.6)
Current meds for cholesterol, n (%) (n=2378)	287 (12.1)
Past and/or current self-report of:	
Migraine, n (%) (n=2229)	558 (25.0)

[a]For categories with missing data, total non-missing N is reported in parentheses

[b]Race/ethnicity categories presented to participants in Arivale questionnaire

[c]Obese defined as BMI≥30

Characteristic[a]	N=2831
High cholesterol, n (%) (n=2301)	558 (24.2)
Depression, n (%) (n=2278)	521 (22.9)
GERD, n (%) (n=2220)	464 (20.9)
Hypertension, n (%) (n=2316)	434 (18.7)
Asthma, n (%) (n=2361)	376 (15.9)
[a]For categories with missing data, total non-missing N is reported in parentheses	
[b]Race/ethnicity categories presented to participants in Arivale questionnaire	
[c]Obese defined as BMI≥30	

(https://imputationserver.sph.umich.edu/index.html#!). Standard pre-imputation quality control was performed on all datasets individually, including exclusion of individuals with low call rate, individuals with a high degree of relatedness, and variants with low call rate (20). Individuals with non-European ancestry according to principal components analysis of ancestry-informative markers were excluded from the further analysis. Detailed descriptions of individual ADGC datasets can be found in Kunkle et al. (5) and **Table S5**. Study-specific logistic regression analyses employed Plink (21) for sex-interaction analysis and SNPTest (22) for sex-stratified analysis. Sex-interaction, which analyzed the sex*variant interaction, and sex-stratified analysis of males and females separately, were performed for two separate models per analysis, one adjusting for age, sex and PCs (model 1) and a second adjusting for age, sex, PCs and *APOE* (model 2). Results were meta-analyzed with METAL using inverse variance-based analysis (23).

Results

Summary of population and study design

Sixty-one percent of Arivale participants were female, 22% were of non-white self-reported ethnicity, and 28% were obese (**Table 1**). The mean age at blood draw was 47 years, with a range of 18 to 89+. In general, individuals who joined Arivale had somewhat higher levels of cardiovascular risk markers compared to the US population, and slightly lower rates of obesity and pre-diabetes (3) (these rates were consistent with rates in the geographies and ethnicities of the population, mostly from the west coast region of the United States).

Phenome-wide association study results

We observed 33 SNP-analyte associations that were statistically significant at FDR-adjusted p-value<0.05, with most of the associations observed for the *APOE* SNPs (rs7412, or the e2-defining allele, and rs429358, or the e4-defining allele). The other SNPs showing significant associations with at least

one clinical chemistry, protein, or metabolite were rs10933431, rs12539172, and rs3752246 (Figure 1, **Table S2**).

NYAP1

The most robust SNP-analyte associations we observed were between rs12539712 in the 3' region of *NYAP1* (Neuronal Tyrosine Phosphorylated Phosphoinositide-3-Kinase Adaptor 1), and two co-regulated proteins, paired immunoglobulin-like type 2 receptors beta and alpha (PILRB and PILRA) (Figure 2). Carriage of the minor allele (AD risk odds ratio (OR)=0.92) was associated with significant reduction in normalized protein expression (NPX) of PILRB and PILRA compared to individuals homozygous for the major allele (FDR-adjusted p-values=$2.2x10^{-33}$ and $2.3x10^{-17}$, respectively), while the overall level of NPX increased with age among all genotypes. The reduction in protein levels appears roughly dose-dependent with the number of minor alleles and was observed in all but the oldest and youngest age groups (likely due to small numbers of the minor allele in these groups (**Table S3A**). These observations led us to previous studies pointing to variation in *PILRA* as the causal gene at this locus, with a missense SNP as a leading candidate (G78R, rs1859788) (24–27). In post-hoc analysis, we repeated the PheWAS with this putative causal SNP (which was in LD with our index SNP rs12539172, R^2=0.77), and the associations became stronger (FDR-adjusted p-value for PILRB=$3.6x10^{-52}$; for PILRA=$1.4x10^{-22}$) (Figure 2**)**, with genotype differences observed in all age groups (**Table S3A**).

APOE4: We observed significant associations between rs429358 (which encodes the e4 allele) and multiple related clinical measures of cholesterol (Figure 3**)**. Differences by genotype were less pronounced in older age groups likely due to statin use (**Table S3B**); exploratory analyses visualizing only individuals who did not report use of statin-lowering medications showed more consistent genotype-dependent differences between rs429358 and the top cholesterol marker, low-density lipoprotein (LDL) particle number (**Figure S1, Table S3B**). The concentration of two proteins in the blood were associated with the e4 allele: PLA2G7 (Platelet Activating Factor Acetylhydrolase) and CD28 (T-Cell-Specific Surface Glycoprotein CD28) (Figure 3). Selected lipid metabolites in the blood were positively associated with e4: two diacylglycerol (DG) metabolites (one of which was measured twice in the Metabolon panel) were higher in e4 carriers compared to non-carriers.

APOE2: we observed significantly lower levels of multiple clinical measures of LDL cholesterol associated with carriage of the e2 allele (Figure 4). As the unadjusted plots show, e2 homozygotes are dramatically different than other genotypes, though it should be noted that few e2 homozygotes were present in the population (n=16) and were within a limited age range (30-59 years). Selected lipid metabolites in the blood were positively associated with e2: a monoglyceride (MG) and four diacylglycerol (DG) metabolites (one of which was a replicate) were higher in e2 carriers compared to non-carriers. We observed six e2-protein associations (Figure 5), such that each of the following proteins were observed at higher levels in e2 carriers: low density lipoprotein receptor (LDLR), heme oxygenase-1 (HMOX-1), SLAM family member 8 (SLAMF8), ring finger protein 31 (RNF31), contactin associated protein 2 (CNTNAP2), and signal recognition particle 14 (SRP14).

ABCA7

The *ABCA7* (ATP Binding Cassette Subfamily A Member 7) variant (rs3752246), which has been associated with increased risk of AD (OR=1.15, **Table S1**), was associated with lower levels of two lactosylceramide (LC) metabolites in the sphingolipid family. These differences were evident starting in the youngest age groups (**Figure S2, Table S3A**). The minor allele of rs3752246 was also associated with higher levels of DEFA1 (Defensin Alpha 1), an antimicrobial peptide.

INPP5D

An intronic SNP in *INPP5D* (Inositol Polyphosphate-5-Phosphatase D) (rs10933431), which was associated with a lowered risk of AD in meta-analyses, was associated with lower levels of the protein IDUA (alpha-L-iduronidase) (**Figure S2**).

Polygenic risk score

No associations were observed between the *APOE*-free PGRS and any analyte after FDR correction for multiple testing, either in primary analyses or in analyses adjusted for e4 status, or among non-e4 individuals only. No effect modification by sex or APOE4 status was observed.

Sex-specific findings

We observed a SNP x sex interaction involving the AD-protective *PICALM* variant, such that the minor allele was associated with higher levels of 30 proteins in men and lower levels of the proteins in women (Figure 6, **Figure S3, Table S4**). These proteins were highly correlated with one another (mean pairwise spearman's rho = 0.49); thus, it is unclear whether the associations are independently biologically meaningful, or whether there is a passenger effect, in which one or a few proteins are driving the sex-differential association with genotype observed in the data. In addition, the *PICALM* variant is associated with a sex-specific effect on five highly correlated long-chain fatty acid (LCFA) metabolites and one polyunsaturated fatty acid (PFA) metabolite (Docosahexaenoic acid) (Figure 6). To investigate further, we conducted a post-hoc analysis examining the impact of this variant on AD risk stratified by sex, in a meta-analysis of clinically diagnosed late-onset AD (18,812 individuals, **Table S5**). While AD risk was reduced in both men and women among carriers of the minor allele, the effect was stronger among men (Table 2 **and Table S6**), which was consistent with the sex-stratified SNP-analyte analyses (data not shown).

Table 2
Results of sex-specific analysis and sex-SNP interaction analysis of PICALM variant 3851179 in the ADGC[a]

Sex-stratified results[b]	Beta	StdError	P-value	MAF
Male model 1	-0.206	0.035	5.62E-09	0.358
Male model 2	-0.176	0.038	4.08E-06	0.359
Female model 1	-0.083	0.029	4.37E-03	0.354
Female model 2	-0.087	0.031	5.60E-03	0.352
Interaction Results[c]	**Interaction beta**	**Std Error**	**P-value**	**MAF**
Model 1	0.116	0.044	8.05E-03	0.354
Model 2	0.372	0.048	7.84E-02	0.354
[a]N=9,135 cases (60% female), 9,677 controls (60% female)				
[b]Model 1: adjusted for age, sex, and PCs; Model 2: adjusted for age, sex, PCs, and APOE genotype.				
[c]Model 1: adjusted for age and PCs; Model 2: adjusted for age, PCs, and APOE.				

Other observed sex-specific effects were more modest (Figure 6). The SNP near *CD2AP* (CD2 Associated Protein) interacted with sex to affect three highly correlated sphingomyelins and three plasmologens, while the SNP in *SPI1* (Transcription Factor PU.1) interacted with sex to affect SPARC related modular calcium binding 2 (SMOC2). Lastly, the missense *ABCA7* SNP interacted with sex to affect levels of Ubiquitin conjugating enzyme E2f (UBE2F).

Stratification by self-identified race/ethnicity

Unfortunately, due to vanishingly small numbers in individual self-identified groups (**Table 1**), we were not able to assess genetic risk effects in individual groups with statistical rigor. As expected, analyses restricted to white individuals recapitulated results of the overall analysis (**Figure S4**). In the nonwhite group overall, we observed effect sizes that were consistent with the overall results and white-only results (**Figure S5**).

Discussion

Our study examines associations between known genetic risk factors for AD and blood markers (clinical labs, proteins, and metabolites). It provides insight into the manifestation of AD-related genetic risk in blood-borne analytes from cognitively normal individuals and demonstrated how AD-related genetic variation manifests in the blood across adulthood. Our results contribute to the growing literature

highlighting a potential causal variant (missense SNP in *PILRA)*, point to potential new mechanisms of protection among *APOE2* carriers, and suggest a role for infectious diseases as AD risk factors, alongside lipid metabolism, immune response, and endocytosis. We also uncovered intriguing differences between men and women in how genetic risk manifests in the blood. These analyses not only add to the existing literature on functional genomics in AD, but also offer up multiple potential new hypotheses to catalyze future studies.

The strongest associations in the study were between the *NYAP1* SNP (rs12539172) and the PILRB/PILRA proteins. This locus was originally identified by rs1476679 near *ZCWCP1* (6). *NYAP1* and *ZCWPW1* are located near *PILRA* and *PILRB* on chromosome 7, within a linkage disequilibrium (LD) block. In previous gene expression studies, the initial index SNP for *ZWCWP1* has been associated with expression of multiple *PILRB* and *PILRA* transcripts in brain (9, 28). *PILRA* and *PILRB* are paired, co-regulated inhibiting/activating receptors, respectively, that are expressed on innate immune cells, recognize certain O-glycosylated proteins, and have an important role in regulating acute inflammatory reactions (29). The R78 substitution in *PILRA* (rs1859788) has been shown to reduce the binding capacity of endogenous ligands and thereby potentially increase microglial activity (27). In addition, while controversial, work from our group and others (30–32) has suggested a potential viral role in AD risk. Notably, the R78 variant has been implicated in HSV-1 (Herpes Simplex Virus type 1) infectivity (27) and differences in HSV-1 antibody titer levels (24). While previous studies have hypothesized that reduced activity of *PILRA* was due to steric conformational changes in the protein leading to reduced binding of key ligands (including HSV-1 glycoprotein B), our results suggest that reduced levels of circulating PILRA protein in R78 carriers could also be a factor in the overall protective effect of this genetic variant.

Statistically significant associations were observed between multiple lipid analytes and the SNPs encoding both *APOE4* (rs429358) and *APOE2* (rs7412). *APOE* normally plays a key role in lipid transport, including shuttling cholesterol to neurons in healthy brains. Notably, *APOE* has a role in beta-amyloid (Aβ) metabolism, and while the exact mechanism is unknown, the e4 variant appears to accelerate neurotoxic Aβ accumulation, aggregation, and deposition in the brain (33). The associations we observed between the e4 variant and increased levels of total cholesterol and LDL cholesterol, along with lower levels of high-density lipoprotein (HDL), were consistent with previous cardiovascular disease cohort studies that included young, middle-aged, and older adults (34–37). The e4 allele was associated with increased NPX of two inflammatory proteins. PLA2G7 is a known cardiovascular risk marker with pro-inflammatory and oxidative activities (38) which has previously been associated with *APOE* genotypes (39) and implicated in AD and cognitive decline (38, 40). To our knowledge, CD28 protein levels have not previously been associated with e4 status, though this relatively weak association may be a downstream result of *APOE* isoform-specific effects on inflammation (41).

Blood cholesterol levels among *APOE2* carriers were also largely consistent with a body of existing data (35); the e2 variant was associated with lower levels of multiple measures of LDL cholesterol. It should be noted that while 5-10% of e2 homozygotes develop type III hyperlipoproteinemia (typically in the presence

of an existing metabolic disorder (42)) resulting in elevated cholesterol levels, all e2 homozygotes in the study had significantly decreased levels of LDL cholesterol compared to other genotypes. In contrast, the e2 variant was associated with higher levels of six lipid metabolites in the diacylglycerol and monoacylglycerol family; interestingly, both the e4 variant and e2 variants were associated with increased levels of the same two lipid metabolites in the diacylglycerol family, despite the opposite effects of these two variants on circulating blood cholesterol. Diacylglycerol is a precursor to triacylglyceride (TG), which is typically higher in *APOE2* carriers (37). The effects of high DGs and TGs remains unclear. DG-rich diets fed to diabetic *APOE*-knockout mice had reduced atherosclerosis and lower plasma cholesterol than mice fed TG-rich or western diets (43, 44); however, non-targeted metabolomics studies have shown elevated levels of DGs and MGs in AD and mild cognitive impairment (MCI) patient brains and blood compared to cognitively intact individuals (45, 46).

We observed six proteins that were significantly upregulated in *APOE2* carriers (Figure 3). The LDLR protein had higher levels of NPX in e2 carriers, particularly in e2 homozygotes (Figure 3a). Though *APOE2* is known to bind poorly to LDLR (~2% of e3 or e4 binding activity) (47), *APOE2* was associated with lower levels of LDL cholesterol across age groups as noted previously, perhaps due to compensatory up-regulation of *LDLR* (37). Greater understanding of the compensatory mechanism leading to upregulated *LDLR* and lower circulating LDL cholesterol is needed. The e2 variant was associated with increased levels of the highly inducible *HMOX-1*, which has antioxidant properties and has been associated with both neuroprotection and neurodegeneration (48). *SLAMF8* may be another link to an antioxidant effect of *APOE2*, as it has been implicated in modulation of reactive oxygen species and inflammation via negative regulation of NOX activity (49). *APOE2* carriers displayed higher levels of RNF31 protein (aka *HOIP*). *HOIP* is the catalytic component of the linear ubiquitin chain assembly complex (LUBAC), which was shown to have a role in the recognition and degradation of misfolded proteins (50). Variation in *CNTNAP2*, a member of the neurexin superfamily of proteins involved in cell-cell interactions in the nervous system, has been associated with neurodevelopmental disorders (51), and has been implicated in AD-related dementia (52). Lastly, *SRP14*, which has a role in targeting secretory proteins to the rough endoplasmic reticulum (ER) membrane, has been identified as one of many tau-associated ER proteins in AD brains (53). To our knowledge, the *APOE2*-protein associations described here are novel and may help point to the mechanisms of protection associated with the e2 variant.

ABCA7 is involved in lipid efflux from cells into lipoprotein particles, plays a role in lipid homeostasis (54), and has also been implicated in Aβ processing and deposition in the brain (55). Our results support *ABCA7*'s lipid-related function by showing lower levels of two LC metabolites among individuals carrying the AD-risk allele of rs3752246. In contrast, we observed higher NPX of DEFA1 protein in carriers of the *ABCA7* variant, which is consistent with previous studies showing higher levels of this protein in cerebral spinal fluid (CSF) and sera of AD patients compared to controls (56, 57), potentially linking *ABCA7* with an inflammatory response pathway to AD. Lastly, lower NPX of IDUA was associated with the *INPP5D* SNP. *INPP5D*, which encodes the lipid phosphatase SHIP1, is a negative regulator of immune signaling and is expressed in microglia (58). To our knowledge, this association has not been previously observed.

Genetic variation likely affects men and women differentially, pointing to mechanisms that contribute to known differences in AD pathology between the sexes (59). The set of proteins that were differentially affected by sex and *PICALM* genotype are primarily implicated in immune processes, cell adhesion, and regulatory processes, with widely overlapping functions (**Figure S6**). Our results highlight an interaction between the AD-risk variant in *PICALM* and multiple proteins implicated in immune response in a sex-specific manner, and support emerging research showing sex differences in the neuroimmune response that impact microglia function (60). We also observed a sex-differential effect of the variant on multiple LCFA metabolites and one PFA metabolite (DHA). A potential link between *PICALM* function, lipids, and AD is feasible: fatty acids, and DHA in particular, have long been known to have a role in maintaining brain health and cognition (61), while *PICALM* expression has been shown to influence cholesterol homeostasis through multiple mechanisms (62). This multi-analyte interaction was supported by results from sex stratified GWAS meta-analyses, which showed differing effect sizes of the variant on men vs. women.

In addition to these sex-specific *PICALM* effects, the SNP near *CD2AP*, a scaffolding protein, interacted with sex to affect three highly correlated sphingomyelins and three plasmologens, while the SNP in *SPI1*, a transcription factor associated with microglial activation (63), interacted with sex to affect SMOC2, a protein involved in microgliosis that has been previously associated with Aβ positivity in CSF (64).

We also examined an AD-specific polygenic risk score. While the PGRS is predictive of disease in case/control studies (65), it was not associated with any blood analytes in the all-ages AD-free Arivale cohort. Combining genetic effects into a single score for AD likely served to dilute any individual genetic effect on the manifestation of genetic risk in the blood. In addition, the relative youth and cognitive health of this cohort should be taken into account. The PGRS may be more likely to detect perturbation in analytes that are markers of systemic inflammation or immune dysfunction in later life and among cohorts experiencing cognitive impairment.

The results presented here are novel and we believe will be of interest to the AD-related functional genomics community, though several limitations should be noted. The study population was not a random sample but was self-selected. The population is largely self-identified non-hispanic white, was mostly located on the west coast, and likely has higher than average socio-economic status (though these data were not captured). Thus, results may not be generalizable to a broader population. At this time, we were not aware of a suitable replication cohort that would contain parallel -omics panels in an all-ages health-heterogeneous cohort. Future studies will be needed to assess generality of the findings to other populations, not only for the sake of replicability of the findings, but due to the relative ancestral homogeneity of this data set. Previous studies have shown genetic heterogeneity between white and non-white individuals, particularly with regard to African Americans and risk of cognitive outcomes among carriers of *APOE* and *ABCA7* variants (17, 18). Given known wide-ranging racial/ethnic disparities in dementia incidence (66), it is imperative that future deep-phenotyping studies are far more inclusive than the study presented here.

Another limitation to the interpretation of results concerns the issue of pleiotropy; we cannot discern pleiotropic, non-AD-related effects from true causal effects that are implicated in AD pathogenesis. However, even if the associations described here are purely the result of pleiotropy and are unrelated to causal mechanisms of AD, the novel associations we described may provide clues to the function of several genes that are highly interesting to the AD community. Related, we only obtained peripheral plasma, and are unable to examine effects in AD-relevant compartments such as brain or CSF. We had high-coverage WGS available and did not interrogate other types of genetic variation such as copy number variants, indels, and short tandem repeats. Lastly, data harmonization with other studies will be a challenge. For instance, most previous metabolomics studies used metabolomics data that lacked complete speciation, and more work is needed with newer technologies that yield high fidelity data to determine the biological effects of specific serum metabolites.

This study also has multiple strengths. While most studies focused on AD-related genetic variation consist of case/control cohorts in older adults, the Arivale data offered an unprecedented look into how genetic variation perturbs physiological pathways in the blood long before disease onset, in health-heterogeneous individuals of all ages. This feature allowed us to observe subtle changes in blood associated with genetic variation, due to the relatively large sample size (2831 individuals with WGS) and the high quality of the blood analytes collected. Our results are from a “real-world” cohort, which promises to be an increasing source of large-scale data in the community going forward, with its accompanying advantages and disadvantages. Some results were previously unobserved and need to be replicated (such as the associations between *APOE2* and multiple proteins), while other results agree with previous findings and serve to reinforce confidence that the results are reasonably representative and not simply spurious.

Conclusions

Due to a unified world-wide effort, dozens of genetic variants have been robustly implicated in the development of AD, though we are still in the early stages of understanding exactly how genetic variation contributes to disease. Our study showed that AD-related genetic variation manifests in the blood, from early adulthood onward, and highlights known targets for prevention in early and mid-life, such as cholesterol monitoring, mitigation of inflammation, and possibly, HSV-1 prevention and/or viral load management. Importantly, as well as yielding new insight into the pathobiology of AD through adulthood, these results may provide a significant number of new drug targets that are highly novel and biologically plausible or may serve as biomarkers if confirmed to have a consistent influence on AD pathophysiology. Lastly, these results highlight the need to assess for sex differences in future studies. Taken together, these results not only illustrate previously unobserved biological phenomenon as a result of AD-associated genetic variation, but also serve as an important resource for the generation of hypotheses for future functional genomics studies and emphasize the potential insight that can be gleaned from deeply phenotyped individuals.

Abbreviations

AD
Alzheimer’s Disease
GWAS
genome-wide association study
PheWAS
phenome-wide association study
PGRS
polygenic risk score
WGS
whole genome sequencing
CLIA
clinical laboratory improvement amendments
APOE
Apolipoprotein E
IGAP
International Genomics of Alzheimer’s Project
FDR
false discovery rate
SNP
single nucleotide polymorphism
LD
linkage disequilibrium
NPX
normalized protein expression
PCs
principal components
ANOVA
analysis of variance
PICALM
Phosphatidylinositol Binding Clathrin Assembly Protein
ADGC
Alzheimer’s Disease Genetics Consortium
HRC
Haplotype Reference Consortium
NYAP1
Neuronal Tyrosine Phosphorylated Phosphoinositide-3-Kinase Adaptor 1
PILRB
paired immunoglobulin-like type 2 receptors beta

PILRA
paired immunoglobulin-like type 2 receptors alpha
OR
odds ratio
LDL
low-denisty lipoprotein
PLA2G7
platelet activating factor acetylhydrolase
CD28
T-Cell-Specific Surface Glycoprotein CD28
DG
diacylglycerol
MG
monoglyceride
LDLR
low density lipoprotein receptor
HMOX1
heme oxygenase-1
SLAMF8
SLAM family member 8
RNF31
ring finger protein 31
CNTNAP2
contactin associated protein 2
SRP14
signal recognition particle 14
ABCA7
ATP Binding Cassette Subfamily A Member 7
LC
lactosylceramide
DEFA1
Defensin Alpha 1
INPP5D
Inositol Polyphosphate-5-Phosphatase D
IDUA
alpha-L-iduronidase
LCFA
long-chain fatty acid
PFA
polyunsaturated fatty acid

CD2AP
CD2 Associated Protein
SPI1
Transcription Factor PU.1
SMOC2
SPARC related modular calcium binding 2
UBE2F
Ubiquitin conjugating enzyme E2f
HSV-1
herpes simplex virus type 1
Aβ
beta-amyloid
HDL
high-density lipoprotein
TG
tracylglyceride
MCI
mild cognitive impairment
CSF
cerebrospinal fluid

Declarations

Ethics approval and consent to participate

Procedures for the Arivale participants were run under the Western Institutional Review Board (WIRB) with Institutional Review Board Study Number 20170658 at the Institute for Systems Biology and 1178906 at Arivale (both in Seattle, WA).

Availability of data and materials

The datasets generated and/or analysed during the current study are not publicly available because the data was generated by a private investment firm under legal terms that mandate researchers to sign a data access agreement permitting the use of these data for non-proift research purposes. Upon reasonable request, researchers can access the Arivale deidentified dataset supporting the findings in this study for research purposes from ISB. Requests should be sent to data-access@isbscience.org. The data are available to qualified researchers on submission and approval of a research plan.

Code used for PheWAS statistical analysis is available through the Sage Bionetworks Github (https://github.com/Sage-Bionetworks/ADsnps_PheWAS.git).

Competing Interests

The authors declare no competing interests.

Funding

This study was supported by the National Institutes of Health, National Institute on Aging grants U01 AG046139 (N.E-T, T.E.G, N.D.P), R01 AG061796 (N.E-T), RF1 AG051504 (N.E-T), R01-AG062634-01 (B.W.K, E.R.M), and U19 AG023122 (N.D.P).

ADGC. The National Institutes of Health, National Institute on Aging (NIH-NIA) supported this work through the following grants: ADGC, U01 AG032984, RC2 AG036528; Samples from the National Cell Repository for Alzheimer's Disease (NCRAD), which receives government support under a cooperative agreement grant (U24 AG21886) awarded by the National Institute on Aging (NIA), were used in this study. We thank contributors who collected samples used in this study, as well as patients and their families, whose help and participation made this work possible; Data for this study were prepared, archived, and distributed by the National Institute on Aging Alzheimer's Disease Data Storage Site (NIAGADS) at the University of Pennsylvania (U24-AG041689-01); NACC, U01 AG016976; NIA LOAD (Columbia University), U24 AG026395, U24 AG026390, R01AG041797; Banner Sun Health Research Institute P30 AG019610; Boston University, P30 AG013846, U01 AG10483, R01 CA129769, R01 MH080295, R01 AG017173, R01 AG025259, R01 AG048927, R01AG33193, R01 AG009029; Columbia University, P50 AG008702, R37 AG015473, R01 AG037212, R01 AG028786; Duke University, P30 AG028377, AG05128; Emory University, AG025688; Group Health Research Institute, UO1 AG006781, UO1 HG004610, UO1 HG006375, U01 HG008657; Indiana University, P30 AG10133, R01 AG009956, RC2 AG036650; Johns Hopkins University, P50 AG005146, R01 AG020688; Massachusetts General Hospital, P50 AG005134; Mayo Clinic, P50 AG016574, R01 AG032990, KL2 RR024151; Mount Sinai School of Medicine, P50 AG005138, P01 AG002219; New York University, P30 AG08051, UL1 RR029893, 5R01AG012101, 5R01AG022374, 5R01AG013616, 1RC2AG036502, 1R01AG035137; North Carolina A&T University, P20 MD000546, R01 AG28786-01A1; Northwestern University, P30 AG013854; Oregon Health & Science University, P30 AG008017, R01 AG026916; Rush University, P30 AG010161, R01 AG019085, R01 AG15819, R01 AG17917, R01 AG030146, R01 AG01101, RC2 AG036650, R01 AG22018; TGen, R01 NS059873; University of Alabama at Birmingham, P50 AG016582; University of Arizona, R01 AG031581; University of California, Davis, P30 AG010129; University of California, Irvine, P50 AG016573; University of California, Los Angeles, P50 AG016570; University of California, San Diego, P50 AG005131; University of California, San Francisco, P50 AG023501, P01 AG019724; University of Kentucky, P30 AG028383, AG05144; University of Michigan, P50 AG008671; University of Pennsylvania, P30 AG010124; University of Pittsburgh, P50 AG005133, AG030653, AG041718, AG07562, AG02365; University of Southern California, P50 AG005142; University of Texas Southwestern, P30 AG012300; University of Miami, R01 AG027944, AG010491, AG027944, AG021547, AG019757; University of Washington, P50 AG005136, R01 AG042437; University of Wisconsin, P50 AG033514; Vanderbilt University, R01 AG019085; and Washington University, P50 AG005681, P01 AG03991, P01 AG026276. The Kathleen Price Bryan Brain Bank at Duke University Medical Center is funded by NINDS grant # NS39764, NIMH MH60451 and by Glaxo Smith Kline. Support was also from the Alzheimer's Association (LAF, IIRG-08-89720; MP-V, IIRG-05-

14147), the US Department of Veterans Affairs Administration, Office of Research and Development, Biomedical Laboratory Research Program, and BrightFocus Foundation (MP-V, A2111048). P.S.G.-H. is supported by Wellcome Trust, Howard Hughes Medical Institute, and the Canadian Institute of Health Research. Genotyping of the TGEN2 cohort was supported by Kronos Science. The TGen series was also funded by NIA grant AG041232 to AJM and MJH, The Banner Alzheimer's Foundation, The Johnnie B. Byrd Sr. Alzheimer's Institute, the Medical Research Council, and the state of Arizona and also includes samples from the following sites: Newcastle Brain Tissue Resource (funding via the Medical Research Council, local NHS trusts and Newcastle University), MRC London Brain Bank for Neurodegenerative Diseases (funding via the Medical Research Council),South West Dementia Brain Bank (funding via numerous sources including the Higher Education Funding Council for England (HEFCE), Alzheimer's Research Trust (ART), BRACE as well as North Bristol NHS Trust Research and Innovation Department and DeNDRoN), The Netherlands Brain Bank (funding via numerous sources including Stichting MS Research, Brain Net Europe, Hersenstichting Nederland Breinbrekend Werk, International Parkinson Fonds, Internationale Stiching Alzheimer Onderzoek), Institut de Neuropatologia, Servei Anatomia Patologica, Universitat de Barcelona. ADNI data collection and sharing was funded by the National Institutes of Health Grant U01 AG024904 and Department of Defense award number W81XWH-12-2-0012. ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California. We thank Drs. D. Stephen Snyder and Marilyn Miller from NIA who are ex-officio ADGC members.

Authors' Contributions

L.Heath and N.D.P. conceived and designed the study; L.Heath, J.C.E., B.W.K., A.C.N, and E.R.M. performed data analyses and figure generation; A.T.M., J.C.L., B.W.K., E.R.M, N.E-T., L.Hood, N.R., and N.D.P. acquired the data; L.Heath, J.C.E., N.R., A.T.M., S.A.K., J.C.L., C.C.F., B.A.L., L.M.M., N.E-T., T.E.G., L.Hood, and N.D.P. made substantial contributions to interpretation of the results. L.Heath and N.D.P. were primary authors of the manuscript. All authors read and approved final manuscript.

References

1. Price ND, Magis AT, Earls JC, Glusman G, Levy R, Lausted C, et al. A wellness study of 108 individuals using personal, dense, dynamic data clouds. Nat Biotechnol. 2017 Jul 17;
2. Schüssler-Fiorenza Rose SM, Contrepois K, Moneghetti KJ, Zhou W, Mishra T, Mataraso S, et al. A longitudinal big data approach for precision health. Nat Med. 2019;25(5):792–804.
3. Zubair N, Conomos MP, Hood L, Omenn GS, Price ND, Spring BJ, et al. Genetic Predisposition Impacts Clinical Changes in a Lifestyle Coaching Program. Sci Rep [Internet]. 2019 May 2 [cited 2019 Jun 18];9. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6497671/
4. Gatz M, Reynolds CA, Fratiglioni L, Johansson B, Mortimer JA, Berg S, et al. Role of genes and environments for explaining Alzheimer disease. Arch Gen Psychiatry. 2006 Feb;63(2):168–74.
5. Kunkle BW, Grenier-Boley B, Sims R, Bis JC, Damotte V, Naj AC, et al. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. Nat Genet. 2019;51(3):414–30.
6. Lambert JC, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nat Genet. 2013 Dec;45(12):1452–8.
7. Karch CM, Goate AM. Alzheimer's disease risk genes and mechanisms of disease pathogenesis. Biol Psychiatry. 2015 Jan 1;77(1):43–51.
8. Pimenova AA, Raj T, Goate AM. Untangling Genetic Risk for Alzheimer's Disease. Biol Psychiatry. 2018 15;83(4):300–10.
9. Allen M, Kachadoorian M, Carrasquillo MM, Karhade A, Manly L, Burgess JD, et al. Late-onset Alzheimer disease risk variants mark brain regulatory loci. Neurol Genet. 2015 Aug;1(2):e15.
10. Ertekin-Taner N. Gene expression endophenotypes: a novel approach for gene discovery in Alzheimer's disease. Mol Neurodegener. 2011 May 14;6:31.
11. Atri A. The Alzheimer's Disease Clinical Spectrum: Diagnosis and Management. Med Clin North Am. 2019 Mar;103(2):263–93.
12. Pendergrass SA, Ritchie MD. Phenome-Wide Association Studies: Leveraging Comprehensive Phenotypic and Genotypic Data for Discovery. Curr Genet Med Rep. 2015 Jun 1;3(2):92–100.
13. Xu X, Conomos MP, Manor O, Rohwer JE, Magis AT, Lovejoy JC. Habitual sleep duration and sleep duration variation are independently associated with body mass index. Int J Obes 2005. 2018;42(4):794–800.
14. Hochberg Y, Benjamini Y. More powerful procedures for multiple significance testing. Stat Med. 1990 Jul;9(7):811–8.
15. Hall MA, Verma A, Brown-Gentry KD, Goodloe R, Boston J, Wilson S, et al. Detection of pleiotropy through a Phenome-wide association study (PheWAS) of epidemiologic data as part of the Environmental Architecture for Genes Linked to Environment (EAGLE) study. PLoS Genet. 2014 Dec;10(12):e1004678.

16. Conomos MP, Laurie CA, Stilp AM, Gogarten SM, McHugh CP, Nelson SC, et al. Genetic Diversity and Association Studies in US Hispanic/Latino Populations: Applications in the Hispanic Community Health Study/Study of Latinos. Am J Hum Genet. 2016 Jan 7;98(1):165–84.
17. N'Songo A, Carrasquillo MM, Wang X, Burgess JD, Nguyen T, Asmann YW, et al. African American exome sequencing identifies potential risk variants at Alzheimer disease loci. Neurol Genet. 2017 Apr;3(2):e141.
18. Berg CN, Sinha N, Gluck MA. The Effects of APOE and ABCA7 on Cognitive Function and Alzheimer's Disease Risk in African Americans: A Focused Mini Review. Front Hum Neurosci. 2019;13:387.
19. McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, et al. A reference panel of 64,976 haplotypes for genotype imputation. Nat Genet. 2016 Oct;48(10):1279–83.
20. Sims R, van der Lee SJ, Naj AC, Bellenguez C, Badarinarayan N, Jakobsdottir J, et al. Rare coding variants in PLCG2, ABI3, and TREM2 implicate microglial-mediated innate immunity in Alzheimer's disease. Nat Genet. 2017 Sep;49(9):1373–84.
21. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience. 2015;4:7.
22. Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. Nat Genet. 2007 Jul;39(7):906–13.
23. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinforma Oxf Engl. 2010 Sep 1;26(17):2190–1.
24. Agostini S, Costa AS, Mancuso R, Guerini FR, Nemni R, Clerici M. The PILRA G78R Variant Correlates with Higher HSV-1-Specific IgG Titers in Alzheimer's Disease. Cell Mol Neurobiol. 2019 Nov;39(8):1217–21.
25. Jansen IE, Savage JE, Watanabe K, Bryois J, Williams DM, Steinberg S, et al. Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. Nat Genet. 2019;51(3):404–13.
26. Patel T, Brookes KJ, Turton J, Chaudhury S, Guetta-Baranes T, Guerreiro R, et al. Whole-exome sequencing of the BDR cohort: evidence to support the role of the PILRA gene in Alzheimer's disease. Neuropathol Appl Neurobiol. 2018;44(5):506–21.
27. Rathore N, Ramani SR, Pantua H, Payandeh J, Bhangale T, Wuster A, et al. Paired Immunoglobulin-like Type 2 Receptor Alpha G78R variant alters ligand binding and confers protection to Alzheimer's disease. PLoS Genet. 2018;14(11):e1007427.
28. Karch CM, Ezerskiy LA, Bertelsen S, Alzheimer's Disease Genetics Consortium (ADGC), Goate AM. Alzheimer's Disease Risk Polymorphisms Regulate Gene Expression in the ZCWPW1 and the CELF1 Loci. PloS One. 2016;11(2):e0148717.
29. Wang J, Shiratori I, Uehori J, Ikawa M, Arase H. Neutrophil infiltration during inflammation is regulated by PILRα via modulation of integrin activation. Nat Immunol. 2013 Jan;14(1):34–40.
30. Eimer WA, Vijaya Kumar DK, Navalpur Shanmugam NK, Rodriguez AS, Mitchell T, Washicosky KJ, et al. Alzheimer's Disease-Associated β-Amyloid Is Rapidly Seeded by Herpesviridae to Protect against

Brain Infection. Neuron. 2018 11;99(1):56-63.e3.

31. Itzhaki RF, Wozniak MA. Herpes Simplex Virus Type 1 in Alzheimer's Disease: The Enemy Within. J Alzheimers Dis. 2008 Jan 1;13(4):393–405.
32. Readhead B, Haure-Mirande J-V, Funk CC, Richards MA, Shannon P, Haroutunian V, et al. Multiscale Analysis of Independent Alzheimer's Cohorts Finds Disruption of Molecular, Genetic, and Clinical Networks by Human Herpesvirus. Neuron. 2018 11;99(1):64-82.e7.
33. Liu C-C, Liu C-C, Kanekiyo T, Xu H, Bu G. Apolipoprotein E and Alzheimer disease: risk, mechanisms and therapy. Nat Rev Neurol. 2013 Feb;9(2):106–18.
34. Karjalainen J-P, Mononen N, Hutri-Kähönen N, Lehtimäki M, Juonala M, Ala-Korpela M, et al. The effect of apolipoprotein E polymorphism on serum metabolome - a population-based 10-year follow-up study. Sci Rep. 2019 Jan 24;9(1):458.
35. Downer B, Estus S, Katsumata Y, Fardo DW. Longitudinal trajectories of cholesterol from midlife through late life according to apolipoprotein E allele status. Int J Environ Res Public Health. 2014 Oct 16;11(10):10663–93.
36. Grönroos P, Raitakari OT, Kähönen M, Hutri-Kähönen N, Marniemi J, Viikari J, et al. Influence of apolipoprotein E polymorphism on serum lipid and lipoprotein changes: a 21-year follow-up study from childhood to adulthood. The Cardiovascular Risk in Young Finns Study. Clin Chem Lab Med. 2007;45(5):592–8.
37. Bennet AM, Di Angelantonio E, Ye Z, Wensley F, Dahlin A, Ahlbom A, et al. Association of apolipoprotein E genotypes with lipid levels and coronary risk. JAMA. 2007 Sep 19;298(11):1300–11.
38. Davidson JE, Lockhart A, Amos L, Stirnadel-Farrant HA, Mooser V, Sollberger M, et al. Plasma lipoprotein-associated phospholipase A2 activity in Alzheimer's disease, amnestic mild cognitive impairment, and cognitively healthy elderly subjects: a cross-sectional study. Alzheimers Res Ther. 2012 Dec 7;4(6):51.
39. Drenos F, Talmud PJ, Casas JP, Smeeth L, Palmen J, Humphries SE, et al. Integrated associations of genotypes with multiple blood biomarkers linked to coronary heart disease risk. Hum Mol Genet. 2009 Jun 15;18(12):2305–16.
40. van Oijen M, van der Meer IM, Hofman A, Witteman JCM, Koudstaal PJ, Breteler MMB. Lipoprotein-associated phospholipase A2 is associated with risk of dementia. Ann Neurol. 2006 Jan;59(1):139–44.
41. Zhang H, Wu L-M, Wu J. Cross-talk between apolipoprotein E and cytokines. Mediators Inflamm. 2011;2011:949072.
42. Mahley RW, Huang Y, Rall SC. Pathogenesis of type III hyperlipoproteinemia (dysbetalipoproteinemia). Questions, quandaries, and paradoxes. J Lipid Res. 1999 Nov;40(11):1933–49.
43. Fujii A, Allen TJ, Nestel PJ. A 1,3-diacylglycerol-rich oil induces less atherosclerosis and lowers plasma cholesterol in diabetic apoE-deficient mice. Atherosclerosis. 2007 Jul;193(1):55–61.

44. Ijiri Y, Naemura A, Yamashita T, Meguro S, Watanabe H, Tokimitsu I, et al. Dietary diacylglycerol extenuates arterial thrombosis in apoE and LDLR deficient mice. Thromb Res. 2006;117(4):411–7.
45. Wood PL, Barnette BL, Kaye JA, Quinn JF, Woltjer RL. Non-targeted lipidomics of CSF and frontal cortex grey and white matter in control, mild cognitive impairment, and Alzheimer's disease subjects. Acta Neuropsychiatr. 2015 Oct;27(5):270–8.
46. Wood PL, Medicherla S, Sheikh N, Terry B, Phillipps A, Kaye JA, et al. Targeted Lipidomics of Fontal Cortex and Plasma Diacylglycerols (DAG) in Mild Cognitive Impairment and Alzheimer's Disease: Validation of DAG Accumulation Early in the Pathophysiology of Alzheimer's Disease. J Alzheimers Dis JAD. 2015;48(2):537–46.
47. Mahley RW. Apolipoprotein E: from cardiovascular disease to neurodegenerative disorders. J Mol Med Berl Ger. 2016;94(7):739–46.
48. Barone E, Di Domenico F, Mancuso C, Butterfield DA. The Janus face of the heme oxygenase/biliverdin reductase system in Alzheimer disease: it's time for reconciliation. Neurobiol Dis. 2014 Feb;62:144–59.
49. Wang G, Abadía-Molina AC, Berger SB, Romero X, O'Keeffe MS, Rojas-Barros DI, et al. Cutting edge: Slamf8 is a negative regulator of Nox2 activity in macrophages. J Immunol Baltim Md 1950. 2012 Jun 15;188(12):5829–32.
50. van Well EM, Bader V, Patra M, Sánchez-Vicente A, Meschede J, Furthmann N, et al. A protein quality control pathway regulated by linear ubiquitination. EMBO J. 2019 May 2;38(9).
51. Rodenas-Cuadrado P, Ho J, Vernes SC. Shining a light on CNTNAP2: complex functions to complex disorders. Eur J Hum Genet EJHG. 2014 Feb;22(2):171–8.
52. van Abel D, Michel O, Veerhuis R, Jacobs M, van Dijk M, Oudejans CBM. Direct downregulation of CNTNAP2 by STOX1A is associated with Alzheimer's disease. J Alzheimers Dis JAD. 2012;31(4):793–800.
53. Meier S, Bell M, Lyons DN, Ingram A, Chen J, Gensel JC, et al. Identification of Novel Tau Interactions with Endoplasmic Reticulum Proteins in Alzheimer's Disease Brain. J Alzheimers Dis JAD. 2015;48(3):687–702.
54. De Roeck A, Van Broeckhoven C, Sleegers K. The role of ABCA7 in Alzheimer's disease: evidence from genomics, transcriptomics and methylomics. Acta Neuropathol (Berl). 2019 Aug;138(2):201–20.
55. Sakae N, Liu C-C, Shinohara M, Frisch-Daiello J, Ma L, Yamazaki Y, et al. ABCA7 Deficiency Accelerates Amyloid-β Generation and Alzheimer's Neuronal Pathology. J Neurosci Off J Soc Neurosci. 2016 Mar 30;36(13):3848–59.
56. Szekeres M, Ivitz E, Datki Z, Kálmán J, Pákáski M, Várhelyi ZP, et al. Relevance of defensin β-2 and α defensins (HNP1-3) in Alzheimer's disease. Psychiatry Res. 2016 30;239:342–5.
57. Watt AD, Perez KA, Ang C-S, O'Donnell P, Rembach A, Pertile KK, et al. Peripheral α-defensins 1 and 2 are elevated in Alzheimer's disease. J Alzheimers Dis JAD. 2015;44(4):1131–43.
58. Efthymiou AG, Goate AM. Late onset Alzheimer's disease genetics implicates microglial pathways in disease risk. Mol Neurodegener. 2017 26;12(1):43.

59. Dumitrescu L, Mayeda ER, Sharman K, Moore AM, Hohman TJ. Sex Differences in the Genetic Architecture of Alzheimer's Disease. Curr Genet Med Rep. 2019 Mar;7(1):13–21.
60. Osborne BF, Turano A, Schwarz JM. Sex Differences in the Neuroimmune System. Curr Opin Behav Sci. 2018 Oct;23:118–23.
61. Fotuhi M, Mohassel P, Yaffe K. Fish consumption, long-chain omega-3 fatty acids and risk of cognitive decline or Alzheimer disease: a complex association. Nat Clin Pract Neurol. 2009 Mar;5(3):140–52.
62. Mercer JL, Argus JP, Crabtree DM, Keenan MM, Wilks MQ, Chi J-TA, et al. Modulation of PICALM Levels Perturbs Cellular Cholesterol Homeostasis. PloS One. 2015;10(6):e0129776.
63. Verheijen J, Sleegers K. Understanding Alzheimer Disease at the Interface between Genetics and Transcriptomics. Trends Genet TIG. 2018;34(6):434–47.
64. Whelan CD, Mattsson N, Nagle MW, Vijayaraghavan S, Hyde C, Janelidze S, et al. Multiplex proteomics identifies novel CSF and plasma biomarkers of early Alzheimer's disease. Acta Neuropathol Commun. 2019 Nov 6;7(1):169.
65. Harrison JR, Mistry S, Muskett N, Escott-Price V. From Polygenic Scores to Precision Medicine in Alzheimer's Disease: A Systematic Review. J Alzheimers Dis JAD. 2020;74(4):1271–83.
66. Mayeda ER, Glymour MM, Quesenberry CP, Whitmer RA. Inequalities in dementia incidence between six racial and ethnic groups over 14 years. Alzheimers Dement J Alzheimers Assoc. 2016 Mar;12(3):216–24.

Figures

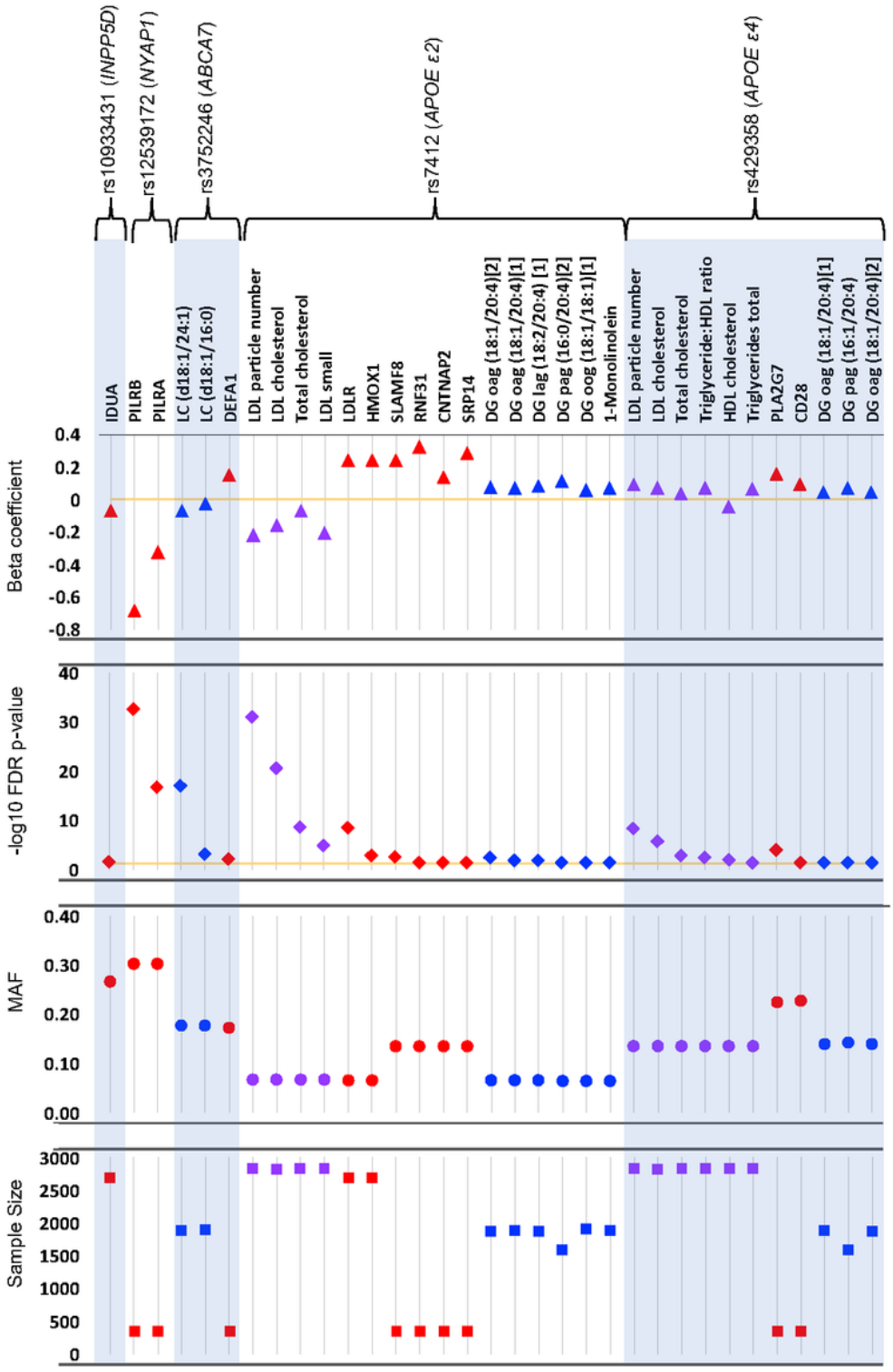

Figure 1

Statistically significant SNP-analyte associations after correcting for multiple testing (threshold FDR-adjusted p-value=0.05), by SNP. Top panel: log-transformed beta-coefficient from the linear regression model adjusted for sex, age, and genetic principal components 1-4; markers above the zero line (orange) indicate analytes that increased in value with the minor allele, while markers below the line indicate markers that decreased in value. Second panel: FDR-adjusted –log10 p-value; orange line at FDR-p=0.05.

Proteins=red, metabolites=blue, clinical chemistries=purple. Metabolite codes: DG=diacylglycerol; LC=lactosylceramide; o=oleoyl; a=arachidonoyl; g=glycerol; l=linoleoyl; p=palmitoyl. Third panel: minor allele frequency (MAF). Bottom panel: Total sample size for each analyte-SNP regression.

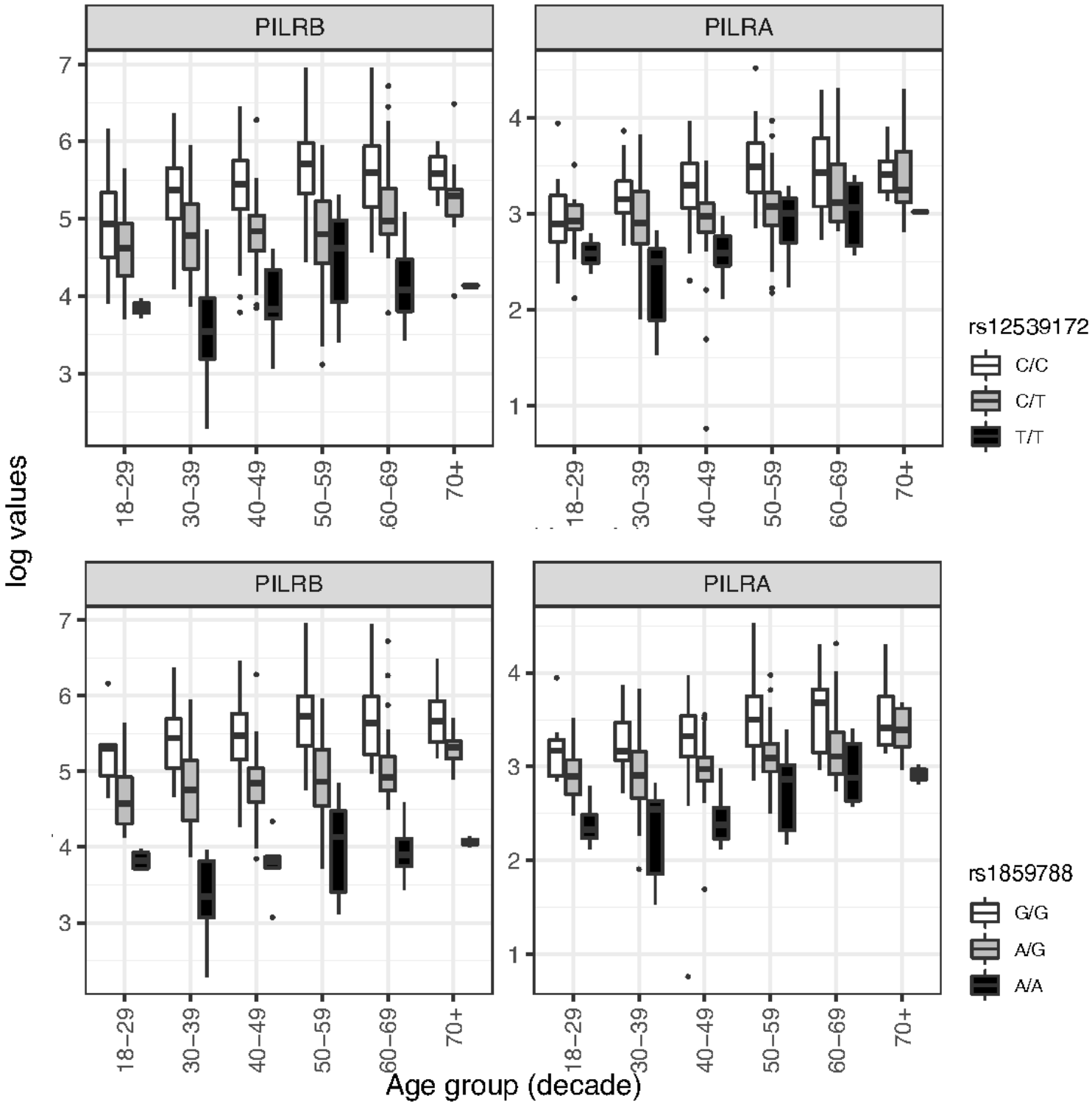

Figure 2

Unadjusted box plots of normalized protein expression (NPX) levels of PILRA and PILRB by genotype and age group. White boxplots=individuals who are homozygous for the major allele, gray

boxplots=heterozygotes, black boxplots=minor allele homozygotes. Box plot midline=median value, lower/upper hinges=25th and 75th percentiles, respectively; lower whisker ends/upper whisker ends no further than 1.5 x interquartile range from the hinge. Data beyond whiskers are outlying points. Top panel: NPX of PILRA and PILRB by rs12539172 (NYAP1) genotype; Bottom panel: NPX of PILRA and PILRB by rs1859788 genotype.

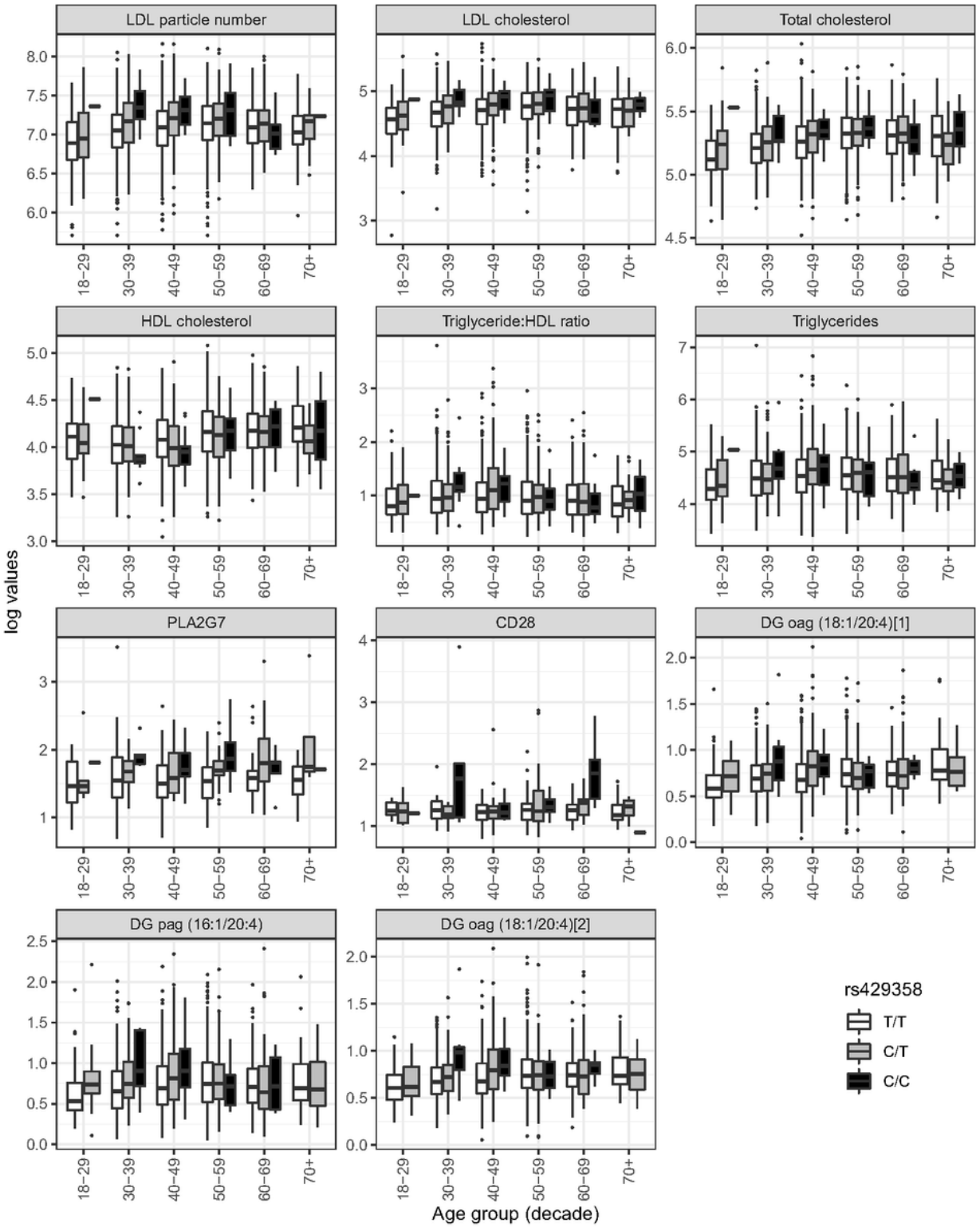

Figure 3

Unadjusted box plots of clinical chemistries and metabolites significantly associated with APOE4 genotype, by age group. White boxplots=individuals who are homozygous for the major allele, gray boxplots=heterozygotes, black boxplots=minor allele homozygotes. Box plot midline=median value, lower/upper hinges=25th and 75th percentiles, respectively; lower whisker ends/upper whisker ends no further than 1.5 x interquartile range from the hinge. Data beyond whiskers are outlying points. Abbreviations/units: LDL particle number (log nmol/L); LDL cholesterol (log mg/dL); Total cholesterol (log mg/dL); HDL cholesterol (log nmol/L); Triglycerides (log nmol/L); T-cell-specific surface glycoprotein CD28 (CD28); oleoyl-arachidonoyl-glycerol (DG oag); palmitoyl-arachidonoyl-glycerol (DG pag); oleoyl-arachidonoyl-glycerol.

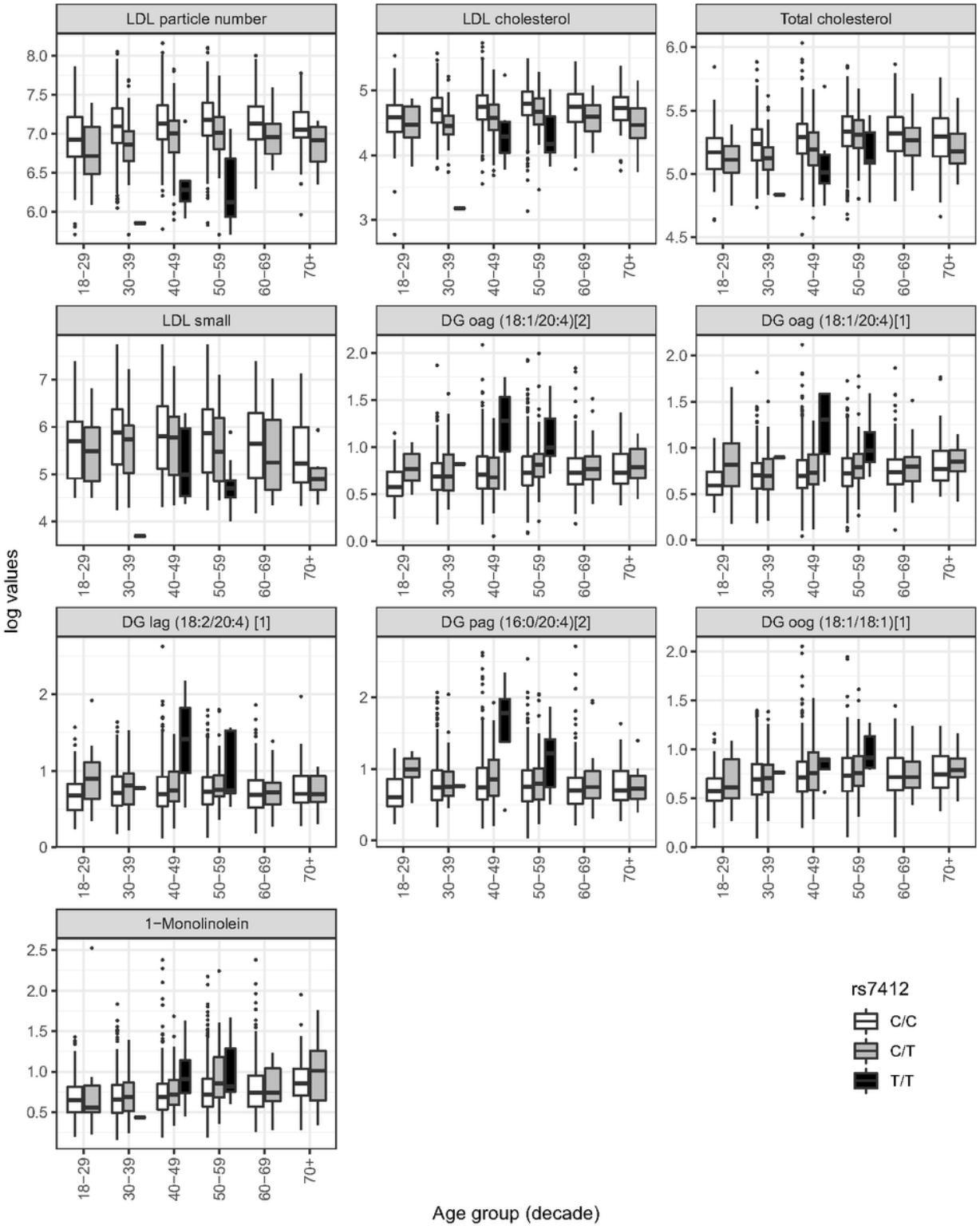

Figure 4

Unadjusted box plots of clinical chemistries and metabolites significantly associated with APOE2 genotype, by age group. White boxplots=individuals who are homozygous for the major allele, gray boxplots=heterozygotes, black boxplots=minor allele homozygotes. Box plot midline=median value, lower/upper hinges=25th and 75th percentiles, respectively; lower whisker ends/upper whisker ends no further than 1.5 x interquartile range from the hinge. Data beyond whiskers are outlying points.

Abbreviations: LDL particle number (log nmol/L); LDL cholesterol (log mg/dL); Total cholesterol (log mg/dL); LDL small (log nmol/L); Diacylglycerol, oleoyl-arachidonoyl-glycerol (DG oag); Diacylglycerol, linoleoyl-arachidonoyl-glycerol (DG lag); Diacylglycerol, palmitoyl-arachidonoyl-glycerol (DG pag); Diacylglycerol, oleoyl-oleoyl-glycerol (DG oog); Monoacelyglycerol, 1-linoleoylglycerol (1-Monolinolein).

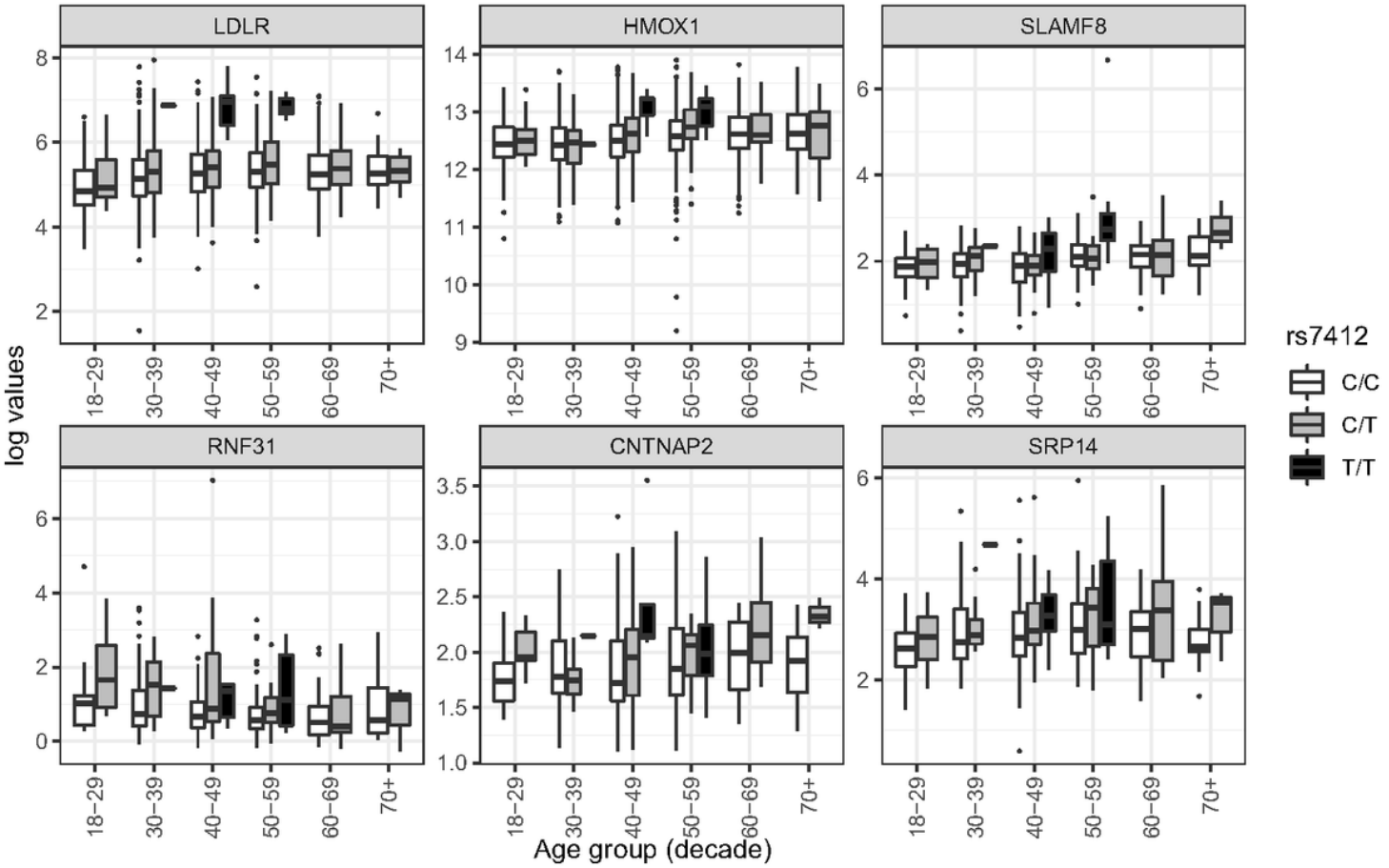

Figure 5

Unadjusted box plots of normalized protein expression levels (NPX) of six proteins significantly associated with APOE2 genotype, by age group. White boxplots=individuals who are homozygous for the major allele, gray boxplots=heterozygotes, black boxplots=minor allele homozygotes. Box plot midline=median value, lower/upper hinges=25th and 75th percentiles, respectively; lower whisker ends/upper whisker ends no further than 1.5 x interquartile range from the hinge. Data beyond whiskers are outlying points. Abbreviations: Low-Density Lipoprotein Receptor (LDLR); heme oxygenase-1 (HMOX1); SLAM family member 8 (SLAMF8); E3 ubiquitin-protein ligase RNF31 (RNF31); Contactin-associated protein-like 2 (CNTNAP2); Signal recognition particle 14 kDa protein (SRP14).

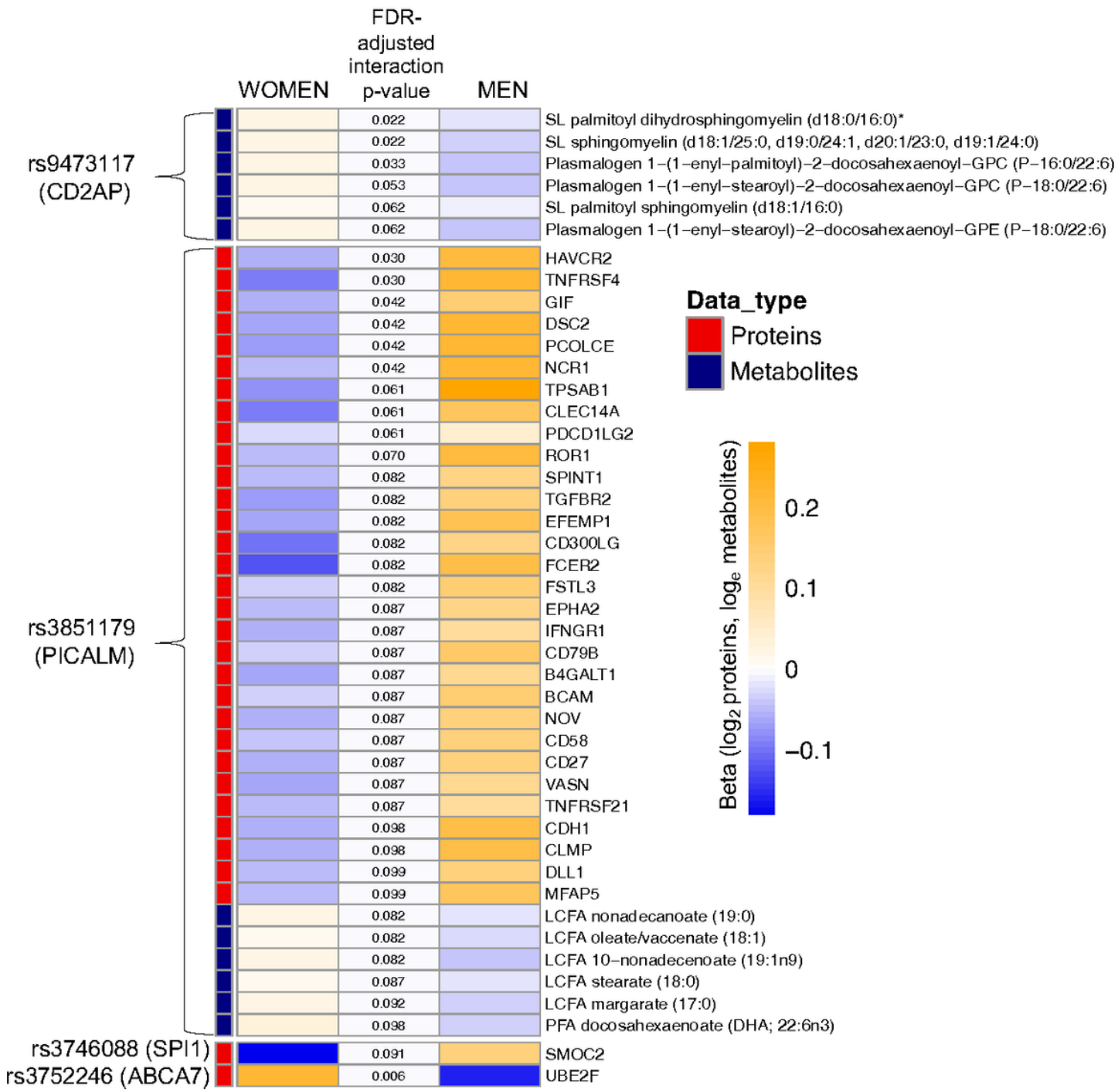

Figure 6

Heatmap of statistically significant genotype x sex interaction terms at FDR-adjusted p-value<0.1. Beta coefficients obtained from sex-stratified analyses, middle-column p-values from interaction term in the full model. SL=sphingolipid; LCFA=long-chain fatty acid; PFA=polyunsaturated fatty acid.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- FigureS1.pdf
- FigureS2.pdf
- FigureS3.pdf
- FigureS4.pdf
- FigureS5.pdf
- FigureS6.pdf
- SupplementaryFigureLegends.docx
- SupplementaryTablesMolNeuro.docx