Manifestations of genetic risk for Alzheimer's Disease in the blood: a cross-sectional multi-omic analysis in healthy adults aged 18-90+

Laura Heath[1,2,*], John C. Earls[1,3], Andrew T. Magis[1], Sergey A. Kornilov[1], Jennifer C. Lovejoy[1], Cory C. Funk[1], Noa Rappaport[1], Benjamin A. Logsdon[2], Lara M. Mangravite[2], Brian W. Kunkle[4,5], Eden R. Martin[4,5], Adam C. Naj[6,7], Nilüfer Ertekin-Taner[8,9], Todd E. Golde[10], Leroy Hood[1,10], Nathan D. Price[1,3,*], Alzheimer's Disease Genetics Consortium

[1]Institute for Systems Biology, Seattle, WA
[2]Sage Bionetworks, Seattle, WA
[3]Onegevity, a division of Thorne HealthTech, New York, NY
[4]John P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL
[5]Dr. John T. Macdonald Foundation Department of Human Genetics, University of Miami Miller School of Medicine, Miami, FL
[6]Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA
[7]Department of Pathology and Laboratory Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA
[8]Mayo Clinic, Department of Neurology, Jacksonville, FL
[9]Mayo Clinic, Department of Neuroscience, Jacksonville, FL
[10]Department of Neuroscience, College of Medicine, Center for Translational Research in Neurodegenerative Disease University of Florida, McKnight Brain Institute, Gainesville, FL
[11]Providence St. Joseph Health, Renton, WA

*Correspondence to: nathan.price@isbscience.org or laura.heath@sagebase.org

Abstract
Deeply phenotyped cohort data can elucidate differences associated with genetic risk for common complex diseases across an age spectrum. Previous work has identified genetic variants associated with Alzheimer's disease (AD) risk from large-scale genome-wide association study meta-analyses. To explore effects of known AD-risk variants, we performed a phenome-wide association study on ~2000 clinical, proteomic, and metabolic blood-based analytes obtained from 2,831 cognitively normal adult clients of a consumer-based scientific wellness company. Results uncovered statistically significant SNP-analyte associations for five genetic variants after correction for multiple testing (for SNPs in or near *NYAP1*, *ABCA7*, *INPP5D*, and *APOE*). These effects were detectable from early adulthood. Sex modified the effects of four genetic variants, with multiple interrelated immune-modulating effects associated with the *PICALM* variant. 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. These analyses support evidence from previous functional genomics studies in the identification of a causal variant within the *PILRA* gene. Taken together, this study highlights clues to the earliest effects of AD genetic risk variants in individuals where disease symptoms have not (yet) arisen.

Introduction

The rapidly decreasing cost of genomics paired with technological advances in the generation of longitudinal 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, differences observed as a function of differential genetic disease risk, and the earliest manifestations of disease transitions. 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 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 effects of genetic risk for AD in individuals without 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.

Results

Summary of population and study design: Data was collected from subscribers in a now-closed scientific wellness program (Arivale, Inc.[3,13]), starting in July 2015 and ending in May 2019. From this population, we identified 2831 participants with whole genome sequencing (WGS) and at least one class of blood-derived analyte (clinical chemistries, proteins, or metabolites) obtained at entry into the program (see Online Methods). 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).

We selected 25 common (>5% minor allele frequency (MAF)) and somewhat rare (>1% MAF) single nucleotide polymorphisms (SNPs) linked to 24 genes that were significantly associated with AD in a recent large-scale meta-analysis[5] (**Supplementary Table 1**). In brief, following a PheWAS approach[14], for each SNP we fitted a linear regression, with genotype as the independent variable and each quantitative log-transformed analyte as the dependent variable, adjusted for age, sex, vendor (in the case of clinical lab values), and four principal genetic components to account for ancestry. We applied the Benjamini-Hochberg false-discovery rate (FDR) correction[15] to account for multiple comparisons. We also assessed for gene by sex interactions. An *APOE*-free polygenic risk score (PGRS) was also incorporated into the PheWAS model. Lastly, since the SNP candidates for this study were derived from meta-analysis of non-

Hispanic white (NHW) cohort populations[5], we stratified PheWAS by self-identified race (white vs. other) in order to determine whether the statistically significant signals we observed were a result of the majority-NHW Arivale cohort demographic makeup.

Phenome-wide association study results: We observed 33 SNP-analyte associations that were statistically significant at FDR-adjusted p-value<0.05, with the majority 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 S1**).

NYAP1: The most robust SNP-analyte associations we observed were between rs12539712 in the 3' region of NYAP1, and two co-regulated proteins, paired immunoglobulin-like type 2 receptors beta and alpha (PILRB and PILRA). 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=2.2×10^{-33} and 2.3×10^{-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 age groups (**Figure 2a**). 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[9,16]. Other recent studies have pointed to variation in *PILRA* as the causal gene at this locus, with a missense SNP as a leading candidate (G78R, rs1859788)[17–20]. 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=3.6×10^{-52}; for PILRA=1.4×10^{-22}) (**Figure 2b).**

APOE4: We observed significant associations between rs429358 (which encodes the e4 allele) and multiple related clinical measures of cholesterol (**Figure S1),** consistent with previous cohort studies that included young, middle-aged, and older adults[21–24]. Differences by genotype were less pronounced in older age groups likely due to statin use; exploratory analyses visualizing only individuals who did not report use of statin-lowering medications showed more consistent genotype-dependent differences in older age groups (**Figure S2**). The concentration of two inflammation-related proteins in the blood were associated with the e4 allele: PLA2G7 and CD28 (**Figure S1**). PLA2G7 (platelet activating factor acetylhydrolase) is a known cardiovascular risk marker with pro-inflammatory and oxidative activities[25] which has previously been associated with APOE genotypes[26] and implicated in AD and cognitive decline[25,27]. 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. These DGs were also elevated in *APOE* e2 carriers (see below).

APOE2: Consistent with previous studies[21–24], we observed significantly lower levels of multiple clinical measures of cholesterol associated with carriage of the e2 allele. As the unadjusted plots show (**Figure S3**), 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. Diacylglycerol is a precursor to triacylglyceride (TG), which is typically higher in APOE2 carriers[24]. 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[28,29];

however, non-targeted metabolomics studies have shown elevated levels of DGs and MGs in AD and MCI patient brains and blood compared to cognitively intact individuals[30,31].

We observed six proteins that were significantly upregulated in APOE2 carriers (**Figure 3**). Though APOE2 is known to bind poorly to LDLR (~2% of e3 or e4 binding activity)[32], APOE2 was associated with lower levels of LDL cholesterol across age groups as noted previously, perhaps due to compensatory up-regulation of LDL receptors (LDLR)[24] (**Figure 3a**). APOE2 was associated with increased levels of the highly inducible HMOX-1, which has antioxidant properties and has been associated with neuroprotection and neurodegeneration[33]. 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[34]. APOE2 carriers displayed higher levels of RNF31 (aka HOIP). HOIP is the catalytic component of the linear ubiquitin chain assembly complex (LUBAC), which was recently shown to have a role in the recognition and degradation of misfolded proteins[35]. 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[36], and has been implicated in AD-related dementia[37]. 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[38].

ABCA7: ABCA7 is involved in lipid efflux from cells into lipoprotein particles, plays a role in lipid homeostasis[39], and has also been implicated in amyloid-β (Aβ) processing and deposition in the brain[40]. Our results support *ABCA7*'s lipid-related function by showing lower levels of two lactosylceramide (LC) metabolites among individuals carrying the AD-risk allele of rs3752246. These differences were evident across all ages but were especially pronounced in the youngest age groups (**Figure S4**). The minor allele of rs3752246 was also associated with higher levels of DEFA1 (neutrophil defensin 1, or alpha-defensin 1), an antimicrobial peptide. This novel gene-protein association is consistent with previous studies showing higher levels of this protein in CSF and sera of AD patients compared to controls[41,42], potentially linking *ABCA7* with an inflammatory response pathway to AD.

INPP5D: An intronic SNP in *INPP5D* (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 S4**). *INPP5D*, which encodes the lipid phosphatase SHIP1, is a negative regulator of immune signaling and is expressed in microglia[43]. To our knowledge, this association has not been previously observed.

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 4, Figure S5**). 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. The set of proteins that are differentially affected by sex and *PICALM* genotype are primarily implicated in immune processes, cell adhesion, and regulatory processes, with widely overlapping functions (**Figure S6**). 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 (DHA) (**Figure 4**). 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[44], while *PICALM* expression has been shown to influence cholesterol homeostasis through multiple mechanisms[45]. 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) (See Online Methods for details). 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 S5**), which was consistent with the sex-stratified SNP-analyte analyses (data not shown).

Other observed sex-specific effects were more modest. 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[46], interacted with sex to affect SMOC2, a protein involved in microgliosis that has been previously associated with Aβ positivity in CSF[47]. Lastly, the missense *ABCA7* SNP interacted with sex to affect levels of Ubiquitin conjugating enzyme E2f (UBE2F).

Stratification by self-identified race/ethnicity: 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[48,49]. 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 S7**). In the nonwhite group overall, we observed effect sizes that were consistent with the overall results and white-only results (**Figure S8**). Despite these overall consistencies, and given known wide-ranging racial/ethnic disparities in dementia incidence[50], it is imperative that future deep-phenotyping studies are far more inclusive than the study presented here.

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 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. PILRA and PILRB are paired 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[51]. The R78 substitution in PILRA (rs1859788) has been shown to reduce the binding capacity of endogenous ligands and thereby potentially increase microglial activity[20]. In addition, while controversial, work from our group and others[52–54] has suggested a potential viral role in AD risk. Notably, the R78 variant has been implicated in HSV-1 infectivity[20] and differences in HSV-1 antibody titer levels[17]. 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.

Strong associations were observed between multiple lipid analytes and the SNPs encoding 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 Aβ metabolism, and while the exact mechanism is unknown, the e4 variant appears to accelerate neurotoxic Aβ accumulation, aggregation, and deposition in the brain[55]. Blood cholesterol levels among APOE4 and APOE2 carriers amongst Arivale participants were largely consistent with a large body of existing data, as this is well studied. It should be noted that while 5-10% of e2 homozygotes develop type III hyperlipoproteinemia (typically in the presence of an existing metabolic disorder[56]) resulting in elevated cholesterol levels, all e2 homozygotes in the study had significantly decreased levels of LDL cholesterol compared to other genotypes. Greater understanding of the compensatory mechanism leading to upregulated *LDLR* and lower circulating LDL cholesterol is needed. In addition, our results indicating a link between APOE2 and several potential protective mechanisms suggested by upregulation of HMOX1, SLAMF8, and RNF31 in e2 carriers, warrants further research.

Genetic variation likely affects men and women differentially, pointing to mechanisms that contribute to known differences in AD pathology between the sexes[57]. Our results highlight a novel 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[58]. 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 effects of individual genetic variants, we also examined an AD-specific polygenic risk score. While the PGRS is predictive of disease in case/control studies[59], 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 NHW, 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. 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. Related, we only obtained peripheral plasma, and are unable to examine effects in AD-relevant compartments such as brain or CSF. We only 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 are in agreement with previous findings and serve to reinforce confidence that the results are reasonably representative and not simply spurious.

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 implicated in development of AD, 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.

Online Methods

Population: The Institute for Systems Biology, through partnership with their spin-out company Arivale, has access to a wealth of data collected longitudinally from subscribers in the commercially available Arivale Scientific Wellness program (as described previously[3,13]), starting in July 2015. 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; coaching generally focused on exercise, nutrition, stress management, and sleep. participants have consented to their de-identified data being used for research purposes.

Blood-derived analytes: for each participant, fasting clinical blood laboratory tests were measured upon joining the program and at regular intervals for those enrolled longitudinally. 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 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 with no more than 20% missing. 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; blood samples were collected and processed by Quest Diagnostics and Labcorp. 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 (https://www.olink.com/products/cvd-ii-panel, Cardiovascular III (https://www.olink.com/products/cvd-iii-panel/), and Inflammation (https://www.olink.com/products/inflammation/). An additional 919 proteins (from 10 additional panels available at Olink at the time) were obtained from a subsample of 354 individuals, in which 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. Olink provides protein measurements as "Normalized Protein Expression" (NPX) values, Olink's arbitrary unit which is in Log2 scale. Up to 1855 of the participants had data from 754 metabolites. Aliquots of frozen plasma samples were sent to Metabolon, Inc. to conduct metabolomics assays using the Metabolon HD4 discovery platform. Relative concentration values were reported for each metabolite.

SNP selection: We selected 25 common and somewhat-rare (>1% allele frequency) SNPs that were significantly associated with Alzheimer's Disease in a recent large-scale meta-analysis based on updated data from the International Genomics of Alzheimer's Project (IGAP)[5]. These variants have been widely studied, and are determined to be within or near genes in one of four pathways known to be associated with AD risk, including cholesterol metabolism, immune response, regulation of endocytosis, and protein ubiquination[5–8,60]. In addition to the variants replicated in Kunkle et al., 2019, we also included the SNP coding for APOE e2 (rs7412), as it is historically understudied due to its rarity and we hypothesized that associations with this SNP may yield valuable clues to the known protective effect of this genotype. 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[15] 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: The study population consists of all participants in the Arivale Wellness program with CLIA-laboratory-generated whole genome sequences. Following a phenome-wide association study approach (PheWAS) [12,14], 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, Olink's arbitrary unit, which is in log2 scale. Genetic ancestry was represented by principal components (PCs) 1-4, calculated using previously described methods [61]. 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 [15] at alpha=0.05 on a per-SNP basis and applied to the main effect of genotype in the primary models, while we set alpha=0.1 as the threshold for the gene x sex interaction models, as interaction terms are typically underpowered compared to main effects and we sought to fully explore sex effects for future hypothesis generation. 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 will be 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 white and nonwhite populations[48,49]. Differences may be due to gene-environment interactions impacting nonwhite populations as a result of sociocultural elements and/or structural inequities due to racism, and/or local ancestry-driven variation near specific loci. As the SNPs selected for this PheWAS were based on a meta-analysis consisting of non-Hispanic white (NHW) individuals, our results may be primarily driven by the fact that the majority of the Arivale population is also NHW, and thus results are not generalizable to other populations (we acknowledge that adjustment for genetic principal components may not be adequate to overcome such strong potential confounding factors). Unfortunately, due to small numbers of individuals in specific non-NHW 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-NHW participants into one stratum for analysis. The stratified NHW and nonwhite group analyses merely serve as an investigation into whether our primary results reflected the

majority-NHW makeup of the Arivale population. PheWAS was applied as described above, with FDR to account for multiple comparisons.

In order 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). 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 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)[62] panel using the Michigan Imputation Server (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[63]. 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]. Study-specific logistic regression analyses employed Plink[64] for sex-interaction analysis and SNPTest[65] 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[66].

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)
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

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.

FIGURE LEGENDS

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 $-\log_{10}$ 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. **a.** NPX of PILRA and PILRB by rs12539172 (NYAP1) genotype; **b.** NPX of PILRA and PILRB by rs1859788 genotype.

Figure 3. 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. **a.** Low-Density Lipoprotein Receptor (LDLR); **b.** heme oxygenase-1 (HMOX1); **c.** SLAM family member 8 (SLAMF8); **d.** E3 ubiquitin-protein ligase RNF31 (RNF31); **e.** Contactin-associated protein-like 2 (CNTNAP2); **f.** Signal recognition particle 14 kDa protein (SRP14).

Figure 4. 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.

Figure 1.

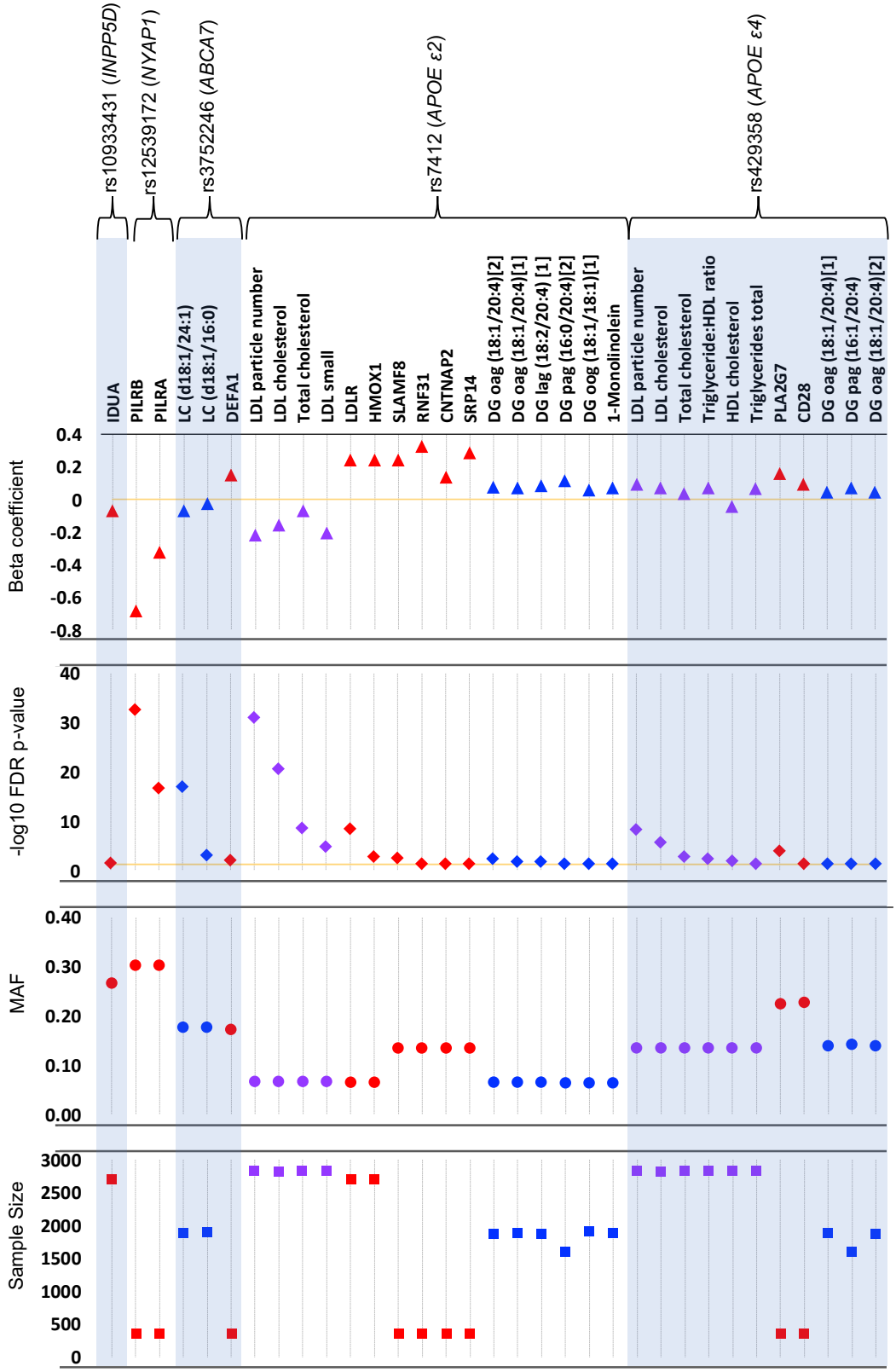
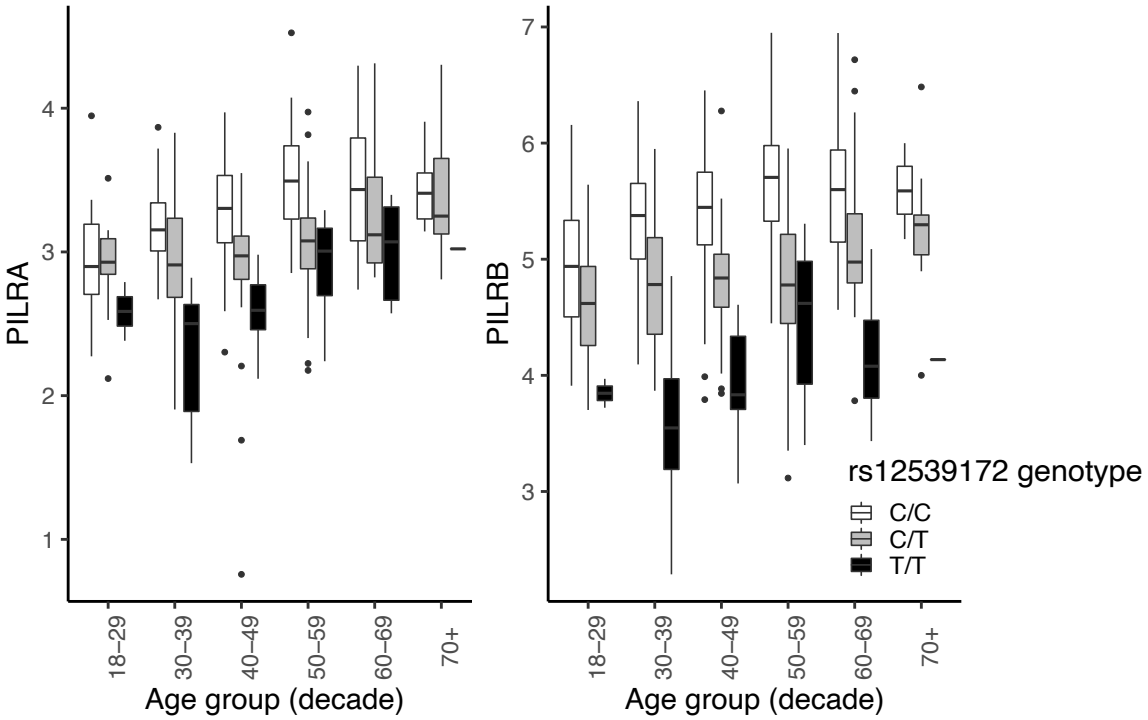

rs10933431 (INPP5D)
rs12539172 (NYAP1)
rs3752246 (ABCA7)
rs7412 (APOE ε2)
rs429358 (APOE ε4)
IDUA
PILRB
PILRA
LC (d18:1/24:1)
LC (d18:1/16:0)
DEFA1
LDL particle number
LDL cholesterol
Total cholesterol
LDL small
LDLR
HMOX1
SLAMF8
RNF31
CNTNAP2
SRP14
DG oag (18:1/20:4)[2]
DG oag (18:1/20:4)[1]
DG lag (18:2/20:4) [1]
DG pag (16:0/20:4)[2]
DG oog (18:1/18:1)[1]
1-Monolinolein
LDL particle number
LDL cholesterol
Total cholesterol
Triglyceride:HDL ratio
HDL cholesterol
Triglycerides total
PLA2G7
CD28
DG oag (18:1/20:4)[1]
DG pag (16:1/20:4)
DG oag (18:1/20:4)[2]
Beta coefficient
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-log10 FDR p-value
40
30
20
10
0
MAF
0.40
0.30
0.20
0.10
0.00
Sample Size
3000
2500
2000
1500
1000
500
0

Figure 2.

a.


PILRA
PILRB
Age group (decade)
18–29
30–39
40–49
50–59
60–69
70+
rs12539172 genotype
C/C
C/T
T/T

b.

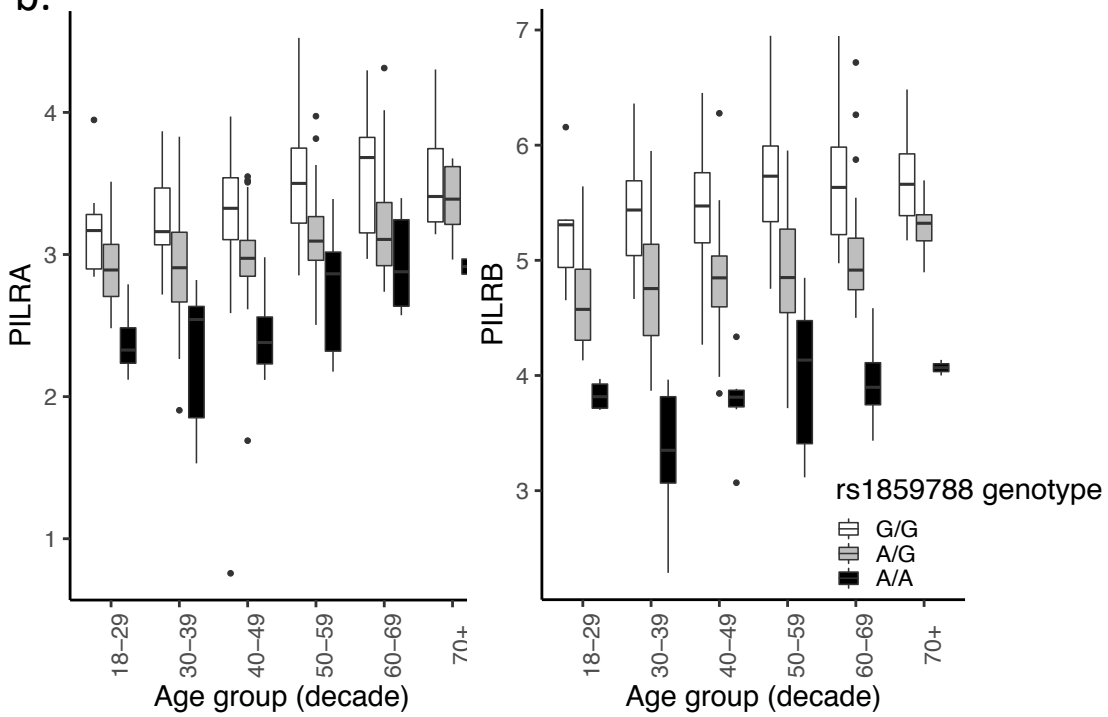

PILRA
PILRB
Age group (decade)
18–29
30–39
40–49
50–59
60–69
70+
rs1859788 genotype
G/G
A/G
A/A

Figure 3.

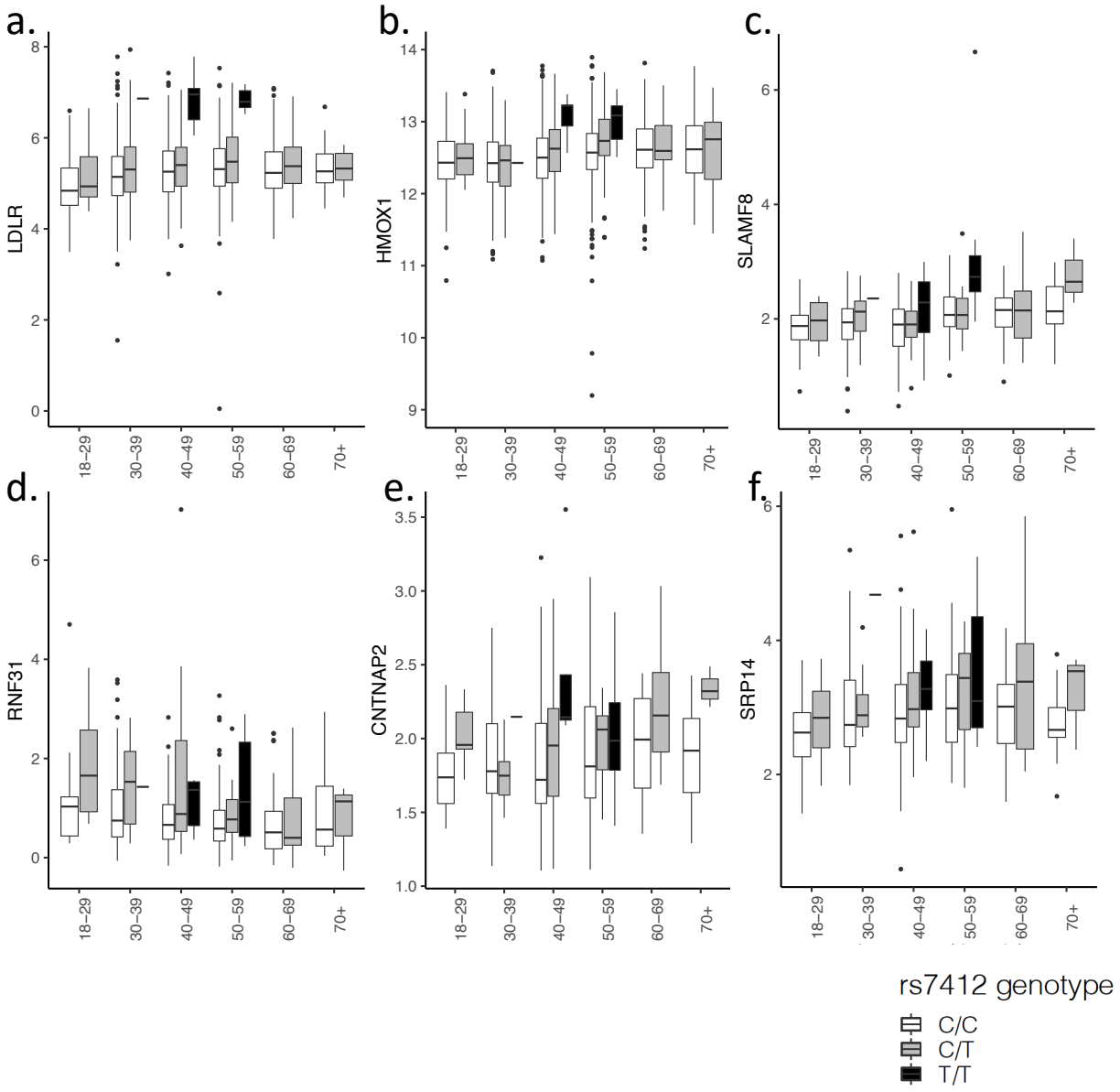
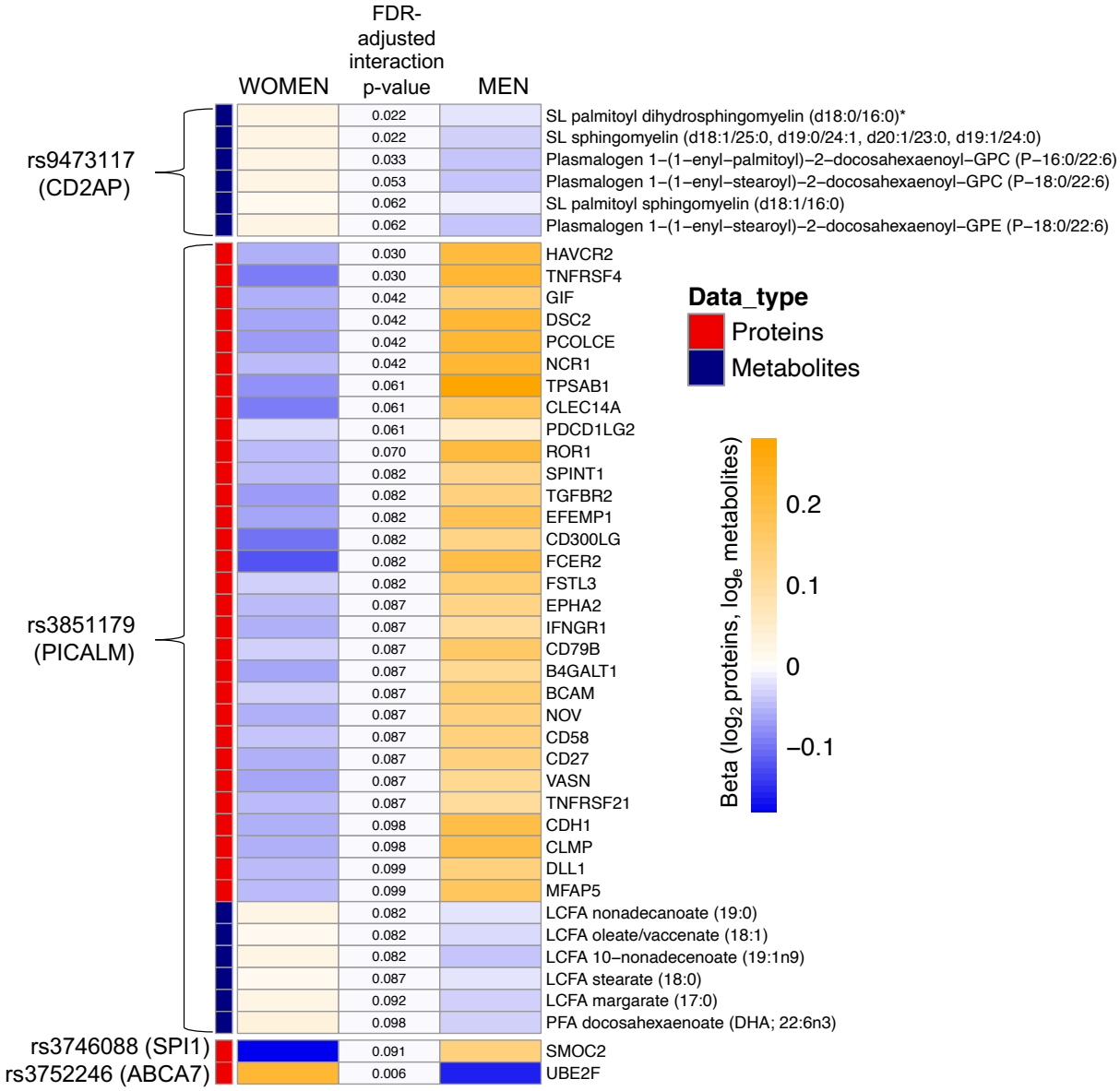

a.
b.
c.
d.
e.
f.
LDLR
HMOX1
SLAMF8
RNF31
CNTNAP2
SRP14
18–29
30–39
40–49
50–59
60–69
70+
rs7412 genotype
C/C
C/T
T/T

Figure 4.


FDR-adjusted interaction p-value
WOMEN
MEN
rs9473117 (CD2AP)
rs3851179 (PICALM)
rs3746088 (SPI1)
rs3752246 (ABCA7)
0.022 SL palmitoyl dihydrosphingomyelin (d18:0/16:0)*
0.022 SL sphingomyelin (d18:1/25:0, d19:0/24:1, d20:1/23:0, d19:1/24:0)
0.033 Plasmalogen 1−(1−enyl−palmitoyl)−2−docosahexaenoyl−GPC (P−16:0/22:6)
0.053 Plasmalogen 1−(1−enyl−stearoyl)−2−docosahexaenoyl−GPC (P−18:0/22:6)
0.062 SL palmitoyl sphingomyelin (d18:1/16:0)
0.062 Plasmalogen 1−(1−enyl−stearoyl)−2−docosahexaenoyl−GPE (P−18:0/22:6)
0.030 HAVCR2
0.030 TNFRSF4
0.042 GIF
0.042 DSC2
0.042 PCOLCE
0.042 NCR1
0.061 TPSAB1
0.061 CLEC14A
0.061 PDCD1LG2
0.070 ROR1
0.082 SPINT1
0.082 TGFBR2
0.082 EFEMP1
0.082 CD300LG
0.082 FCER2
0.082 FSTL3
0.087 EPHA2
0.087 IFNGR1
0.087 CD79B
0.087 B4GALT1
0.087 BCAM
0.087 NOV
0.087 CD58
0.087 CD27
0.087 VASN
0.087 TNFRSF21
0.098 CDH1
0.098 CLMP
0.099 DLL1
0.099 MFAP5
0.082 LCFA nonadecanoate (19:0)
0.082 LCFA oleate/vaccenate (18:1)
0.082 LCFA 10−nonadecenoate (19:1n9)
0.087 LCFA stearate (18:0)
0.092 LCFA margarate (17:0)
0.098 PFA docosahexaenoate (DHA; 22:6n3)
0.091 SMOC2
0.006 UBE2F
Data_type
Proteins
Metabolites
Beta (log2 proteins, loge metabolites)
0.2
0.1
0
−0.1

Competing Interests
The authors declare no competing interests.

Author 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., 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.

Data Availability
Qualified researchers can access the Arivale deidentified dataset supporting the findings in this study for research purposes. 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 Availability
Code used for PheWAS statistical analysis is available through the Sage Bionetworks Github (https://github.com/Sage-Bionetworks/ADsnps_PheWAS.git).

Acknowledgments
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), and R01-AG062634-01 (B.W.K, E.R.M).
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.

REFERENCES

1. Price, N. D. *et al.* A wellness study of 108 individuals using personal, dense, dynamic data clouds. *Nat. Biotechnol.* (2017) doi:10.1038/nbt.3870.
2. Schüssler-Fiorenza Rose, S. M. *et al.* A longitudinal big data approach for precision health. *Nat. Med.* **25**, 792–804 (2019).
3. Zubair, N. *et al.* Genetic Predisposition Impacts Clinical Changes in a Lifestyle Coaching Program. *Sci. Rep.* **9**, (2019).
4. Gatz, M. *et al.* Role of genes and environments for explaining Alzheimer disease. *Arch. Gen. Psychiatry* **63**, 168–174 (2006).
5. Kunkle, B. W. *et al.* Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. *Nat. Genet.* **51**, 414–430 (2019).
6. Lambert, J. C. *et al.* Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat. Genet.* **45**, 1452–1458 (2013).
7. Karch, C. M. & Goate, A. M. Alzheimer's disease risk genes and mechanisms of disease pathogenesis. *Biol. Psychiatry* **77**, 43–51 (2015).
8. Pimenova, A. A., Raj, T. & Goate, A. M. Untangling Genetic Risk for Alzheimer's Disease. *Biol. Psychiatry* **83**, 300–310 (2018).
9. Allen, M. *et al.* Late-onset Alzheimer disease risk variants mark brain regulatory loci. *Neurol. Genet.* **1**, e15 (2015).
10. Ertekin-Taner, N. Gene expression endophenotypes: a novel approach for gene discovery in Alzheimer's disease. *Mol. Neurodegener.* **6**, 31 (2011).
11. Atri, A. The Alzheimer's Disease Clinical Spectrum: Diagnosis and Management. *Med. Clin. North Am.* **103**, 263–293 (2019).
12. Pendergrass, S. A. & Ritchie, M. D. Phenome-Wide Association Studies: Leveraging Comprehensive Phenotypic and Genotypic Data for Discovery. *Curr. Genet. Med. Rep.* **3**, 92–100 (2015).
13. Xu, X. *et al.* Habitual sleep duration and sleep duration variation are independently associated with body mass index. *Int. J. Obes. 2005* **42**, 794–800 (2018).
14. Hall, M. A. *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.* **10**, e1004678 (2014).
15. Hochberg, Y. & Benjamini, Y. More powerful procedures for multiple significance testing. *Stat. Med.* **9**, 811–818 (1990).
16. Karch, C. M., Ezerskiy, L. A., Bertelsen, S., Alzheimer's Disease Genetics Consortium (ADGC) & Goate, A. M. Alzheimer's Disease Risk Polymorphisms Regulate Gene Expression in the ZCWPW1 and the CELF1 Loci. *PloS One* **11**, e0148717 (2016).
17. Agostini, S. *et al.* The PILRA G78R Variant Correlates with Higher HSV-1-Specific IgG Titers in Alzheimer's Disease. *Cell. Mol. Neurobiol.* **39**, 1217–1221 (2019).
18. Jansen, I. E. *et al.* Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nat. Genet.* **51**, 404–413 (2019).
19. Patel, T. *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.* **44**, 506–521 (2018).
20. Rathore, N. *et al.* Paired Immunoglobulin-like Type 2 Receptor Alpha G78R variant alters ligand binding and confers protection to Alzheimer's disease. *PLoS Genet.* **14**, e1007427 (2018).
21. Karjalainen, J.-P. *et al.* The effect of apolipoprotein E polymorphism on serum metabolome - a population-based 10-year follow-up study. *Sci. Rep.* **9**, 458 (2019).
22. Downer, B., Estus, S., Katsumata, Y. & Fardo, D. W. Longitudinal trajectories of cholesterol from midlife through late life according to apolipoprotein E allele status. *Int. J. Environ. Res. Public. Health* **11**, 10663–10693 (2014).

23. Grönroos, P. *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.* **45**, 592–598 (2007).
24. Bennet, A. M. *et al.* Association of apolipoprotein E genotypes with lipid levels and coronary risk. *JAMA* **298**, 1300–1311 (2007).
25. Davidson, J. E. *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.* **4**, 51 (2012).
26. Drenos, F. *et al.* Integrated associations of genotypes with multiple blood biomarkers linked to coronary heart disease risk. *Hum. Mol. Genet.* **18**, 2305–2316 (2009).
27. van Oijen, M. *et al.* Lipoprotein-associated phospholipase A2 is associated with risk of dementia. *Ann. Neurol.* **59**, 139–144 (2006).
28. Fujii, A., Allen, T. J. & Nestel, P. J. A 1,3-diacylglycerol-rich oil induces less atherosclerosis and lowers plasma cholesterol in diabetic apoE-deficient mice. *Atherosclerosis* **193**, 55–61 (2007).
29. Ijiri, Y. *et al.* Dietary diacylglycerol extenuates arterial thrombosis in apoE and LDLR deficient mice. *Thromb. Res.* **117**, 411–417 (2006).
30. Wood, P. L., Barnette, B. L., Kaye, J. A., Quinn, J. F. & Woltjer, R. L. Non-targeted lipidomics of CSF and frontal cortex grey and white matter in control, mild cognitive impairment, and Alzheimer's disease subjects. *Acta Neuropsychiatr.* **27**, 270–278 (2015).
31. Wood, P. L. *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* **48**, 537–546 (2015).
32. Mahley, R. W. Apolipoprotein E: from cardiovascular disease to neurodegenerative disorders. *J. Mol. Med. Berl. Ger.* **94**, 739–746 (2016).
33. Barone, E., Di Domenico, F., Mancuso, C. & Butterfield, D. A. The Janus face of the heme oxygenase/biliverdin reductase system in Alzheimer disease: it's time for reconciliation. *Neurobiol. Dis.* **62**, 144–159 (2014).
34. Wang, G. *et al.* Cutting edge: Slamf8 is a negative regulator of Nox2 activity in macrophages. *J. Immunol. Baltim. Md 1950* **188**, 5829–5832 (2012).
35. van Well, E. M. *et al.* A protein quality control pathway regulated by linear ubiquitination. *EMBO J.* **38**, (2019).
36. Rodenas-Cuadrado, P., Ho, J. & Vernes, S. C. Shining a light on CNTNAP2: complex functions to complex disorders. *Eur. J. Hum. Genet. EJHG* **22**, 171–178 (2014).
37. van Abel, D. *et al.* Direct downregulation of CNTNAP2 by STOX1A is associated with Alzheimer's disease. *J. Alzheimers Dis. JAD* **31**, 793–800 (2012).
38. Meier, S. *et al.* Identification of Novel Tau Interactions with Endoplasmic Reticulum Proteins in Alzheimer's Disease Brain. *J. Alzheimers Dis. JAD* **48**, 687–702 (2015).
39. 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.)* **138**, 201–220 (2019).
40. Sakae, N. *et al.* ABCA7 Deficiency Accelerates Amyloid-β Generation and Alzheimer's Neuronal Pathology. *J. Neurosci. Off. J. Soc. Neurosci.* **36**, 3848–3859 (2016).
41. Szekeres, M. *et al.* Relevance of defensin β-2 and α defensins (HNP1-3) in Alzheimer's disease. *Psychiatry Res.* **239**, 342–345 (2016).
42. Watt, A. D. *et al.* Peripheral α-defensins 1 and 2 are elevated in Alzheimer's disease. *J. Alzheimers Dis. JAD* **44**, 1131–1143 (2015).
43. Efthymiou, A. G. & Goate, A. M. Late onset Alzheimer's disease genetics implicates microglial pathways in disease risk. *Mol. Neurodegener.* **12**, 43 (2017).

44. 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.* **5**, 140–152 (2009).
45. Mercer, J. L. *et al.* Modulation of PICALM Levels Perturbs Cellular Cholesterol Homeostasis. *PloS One* **10**, e0129776 (2015).
46. Verheijen, J. & Sleegers, K. Understanding Alzheimer Disease at the Interface between Genetics and Transcriptomics. *Trends Genet. TIG* **34**, 434–447 (2018).
47. Whelan, C. D. *et al.* Multiplex proteomics identifies novel CSF and plasma biomarkers of early Alzheimer's disease. *Acta Neuropathol. Commun.* **7**, 169 (2019).
48. N'Songo, A. *et al.* African American exome sequencing identifies potential risk variants at Alzheimer disease loci. *Neurol. Genet.* **3**, e141 (2017).
49. Berg, C. N., Sinha, N. & Gluck, M. A. The Effects of APOE and ABCA7 on Cognitive Function and Alzheimer's Disease Risk in African Americans: A Focused Mini Review. *Front. Hum. Neurosci.* **13**, 387 (2019).
50. Mayeda, E. R., Glymour, M. M., Quesenberry, C. P. & Whitmer, R. A. Inequalities in dementia incidence between six racial and ethnic groups over 14 years. *Alzheimers Dement. J. Alzheimers Assoc.* **12**, 216–224 (2016).
51. 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.* **14**, 34–40 (2013).
52. Eimer, W. A. *et al.* Alzheimer's Disease-Associated β-Amyloid Is Rapidly Seeded by Herpesviridae to Protect against Brain Infection. *Neuron* **99**, 56-63.e3 (2018).
53. Itzhaki, R. F. & Wozniak, M. A. Herpes Simplex Virus Type 1 in Alzheimer's Disease: The Enemy Within. *J. Alzheimers Dis.* **13**, 393–405 (2008).
54. Readhead, B. *et al.* Multiscale Analysis of Independent Alzheimer's Cohorts Finds Disruption of Molecular, Genetic, and Clinical Networks by Human Herpesvirus. *Neuron* **99**, 64-82.e7 (2018).
55. Liu, C.-C., Liu, C.-C., Kanekiyo, T., Xu, H. & Bu, G. Apolipoprotein E and Alzheimer disease: risk, mechanisms and therapy. *Nat. Rev. Neurol.* **9**, 106–118 (2013).
56. Mahley, R. W., Huang, Y. & Rall, S. C. Pathogenesis of type III hyperlipoproteinemia (dysbetalipoproteinemia). Questions, quandaries, and paradoxes. *J. Lipid Res.* **40**, 1933–1949 (1999).
57. Dumitrescu, L., Mayeda, E. R., Sharman, K., Moore, A. M. & Hohman, T. J. Sex Differences in the Genetic Architecture of Alzheimer's Disease. *Curr. Genet. Med. Rep.* **7**, 13–21 (2019).
58. Osborne, B. F., Turano, A. & Schwarz, J. M. Sex Differences in the Neuroimmune System. *Curr. Opin. Behav. Sci.* **23**, 118–123 (2018).
59. Harrison, J. R., Mistry, S., Muskett, N. & Escott-Price, V. From Polygenic Scores to Precision Medicine in Alzheimer's Disease: A Systematic Review. *J. Alzheimers Dis. JAD* **74**, 1271–1283 (2020).
60. International Genomics of Alzheimer's Disease Consortium (IGAP). Convergent genetic and expression data implicate immunity in Alzheimer's disease. *Alzheimers Dement. J. Alzheimers Assoc.* **11**, 658–671 (2015).
61. Conomos, M. P. *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.* **98**, 165–184 (2016).
62. McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283 (2016).
63. Sims, R. *et al.* Rare coding variants in PLCG2, ABI3, and TREM2 implicate microglial-mediated innate immunity in Alzheimer's disease. *Nat. Genet.* **49**, 1373–1384 (2017).
64. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* **4**, 7 (2015).

65. 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.* **39**, 906–913 (2007).
66. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinforma. Oxf. Engl.* **26**, 2190–2191 (2010).