Received: 10 July 2025 | Revised: 20 November 2025 | Accepted: 30 December 2025

DOI: 10.1002/alz.71164

Alzheimer's & Dementia®
THE JOURNAL OF THE ALZHEIMER'S ASSOCIATION

RESEARCH ARTICLE

Analyses of plasma multi-omic data across ancestries identify novel pathways implicated in Alzheimer's disease

Chengran Yang[1,2] | **Jigyasha Timsina**[1,2] | **Menghan Liu**[1,2] | **John Budde**[1,2] | **Pat Kohlfeld**[1,2] | **William Brock**[1,2] | **John C. Morris**[3,4] | **David M. Holtzman**[3,4,5] | **Yun Ju Sung**[1,2,6] | **Carlos Cruchaga**[1,2,3,4,5]

[1]Department of Psychiatry, Washington University School of Medicine, St. Louis, Missouri, USA

[2]NeuroGenomics and Informatics Center, Washington University School of Medicine, St. Louis, Missouri, USA

[3]Department of Neurology, Washington University School of Medicine, St. Louis, Missouri, USA

[4]Charles F. and Joanne Knight Alzheimer Disease Research Center, Washington University School of Medicine, St. Louis, Missouri, USA

[5]Hope Center for Neurological Disorders, Washington University School of Medicine, St. Louis, Missouri, USA

[6]Division of Biostatistics, Washington University School of Medicine, St. Louis, Missouri, USA

Correspondence
Carlos Cruchaga, Department of Psychiatry, Washington University School of Medicine, 425 South Euclid Avenue, BJC Institute of Health, Box 8134, St. Louis, MO 63110, USA.
Email: cruchagac@wustl.edu

Funding information
National Institutes of Health, Grant/Award Numbers: R01AG044546, RF1AG053303, RF1AG058501, U01AG058922, RF1AG074007, P30AG066444, P01AG03991, P01AG026276; Chan Zuckerberg Initiative; Alzheimer's Association Zenith Fellows Award, Grant/Award Number: ZEN-22-848604; National Institute on Aging

Abstract

INTRODUCTION: Few genetic studies on Alzheimer's disease (AD) have incorporated multiple ancestries and omic datasets to pinpoint actionable AD risk effectors for each ancestry.

METHODS: Here, we first performed genetic colocalization between molecular phenotypes (proteomics and metabolomics) from two ancestral groups (European [EUR] and African [AFR]) and the two largest EUR AD genome-wide association studies. We next performed pathway enrichment analyses to identify biological mechanisms.

RESULTS: We found 21 proteins and one metabolite colocalized with AD risk that were shared between the EUR and AFR ancestry groups. We also identified 25% AFR and 60% EUR proteins; 50% AFR and 10% EUR metabolites were unique. The pathway enrichment analyses nominated interleukin-1 production and lipid pathway were shared underlying proteomic and metabolomic findings, respectively.

DISCUSSION: Our findings indicate that these four plasma datasets may pinpoint different effectors of AD risk in diverse populations; findings from AFR participants require validation with AFR-based genome-wide association study data.

KEYWORDS
Alzheimer's disease, genetic ancestry, genetic variants, metabolomics, omics, plasma, proteomics

Highlights

- For proteomics, 61% of findings for European (EUR) ancestry and 72% for African (AFR) ancestry were not previously reported.
- For metabolomics, 83% of findings for EUR ancestry and 50% AFR ancestry were not previously reported.
- Both convergent and divergent pathways were identified in EUR- and AFR-ancestry stratified analyses in either proteomics or metabolomics findings.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.
© 2026 The Author(s). *Alzheimer's & Dementia* published by Wiley Periodicals LLC on behalf of Alzheimer's Association.

1 | BACKGROUND

To our knowledge, few genetic studies on Alzheimer's disease (AD) have included more than one molecular layer from more than one genetic ancestry when prioritizing functional molecular traits underlying AD risk loci. For example, the latest two AD genome-wide association studies (GWASs)[1,2] mostly included participants of European (EUR) ancestry. Recently, using cerebrospinal fluid (CSF) tissue, we have identified novel proteins[3] and metabolites[4] that are part of the causal pathways of AD, but focused only on the proteomic and metabolomic datasets from participants of EUR ancestry. It remains unclear whether these findings apply to other ancestries. Moreover, as we demonstrated in a smaller scale study,[5] tissues can exhibit specific effects on protein measures; thus, plasma may nominate different effector proteins than CSF.

Previous studies have found that proteins and metabolites implicated in AD are enriched in pathways underlying AD pathogenesis, such as microglial efferocytosis,[6] lysosomal dysfunction,[3] or the lipid super-pathway (phosphatidyl-choline sub-pathway).[4] These pathways, in addition to individual genes, may represent another druggable tier for AD treatment. However, because multi-ancestry GWASs cannot adequately distinguish ancestry-stratified biological contexts, precision medicine for AD has not yet been fully realized.

There are several approaches to link the genetic architecture of molecular traits, such as mRNA, protein, and metabolite levels, with diseases, which allows us to identify potential effect genes and proteins, and metabolites implicated in disease. Those methods include genetic colocalization[7–13] or transcriptome/proteome/metabolome-wide association studies (TWAS for mRNA levels, PWAS for protein levels, MWAS for metabolite levels)[9,14,15] approaches. Colocalization methods proved to be useful in identifying the key effectors underlying a complex region of a disease. Moreover, given no pleiotropy assumption in the colocalization framework, it can be used to nominate more biologically meaningful molecular features than other methods, such as Mendelian randomization.[16] Additionally, colocalization analysis is also used as a complementary approach after TWAS analyses to reduce false positives. However, as these TWAS approaches only consider the *cis* region of a gene, this limits their application to accounting for the *trans* region of the same gene.

Here, in this study (Figure 1), we used four xQTL (quantitative trait locus) datasets that include two omic layers (proteomics and metabolomics) in two ancestral groups[17] (EUR and African [AFR], Table S1–S4) to identify key proteins and metabolites associated with AD risk by integrating summary statistics–level data from the two recent EUR ancestry–based AD GWASs[1,2] to date (Table S5). By applying genetic colocalization approaches that assume single-variant (coloc.abf[7]) and multi-variant (coloc.susie[10]) models, we nominated AD effector molecular phenotypes in an ancestry-stratified manner. Finally, we explored aggregated pathways and conducted cross-tissue comparisons of these features.

RESEARCH IN CONTEXT

1. **Systematic review**: The authors reviewed the literature using PubMed. Few genetic studies on Alzheimer's disease (AD) have incorporated multiple ancestries and omic datasets to pinpoint actionable AD risk genes for non-European ancestry. Mixed multi-ancestry studies alone are insufficient to distinguish ancestry-specific biological contexts, and thus, the promise of precision medicine for AD remains unfulfilled.
2. **Interpretation**: Our findings indicated that these four plasma-based proteomic and metabolomic datasets may pinpoint different effectors of AD risk in European and African ancestral populations.
3. **Future directions**: (a) Expand the throughput for metabolome profiling and further integrate with proteomics underlying AD risk loci. (b) Examine the potential mechanisms of the apolipoprotein E locus linked with trans-associated proteins. (c) Examine the potential mechanisms of the *ADAM10* locus associated with metabolites from the glycero–phosphatidyl–ethanolamine sub-pathway. (d) Additional studies using AD genome-wide association studies from African ancestry would be needed to validate these findings.

2 | METHODS

2.1 | Ethics declarations

This project was approved by the ethics committee of the Washington University School of Medicine in St. Louis. We certified that the study was performed in accordance with the ethical standards as laid down in the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards.

2.2 | Cohort information

All participants were recruited at the Knight Alzheimer Disease Research Center (Knight ADRC). In total, 3170 participants from all genetic ancestries were selected for both proteomics and metabolomics profiling. Here, we followed the Trans-Omics for Precision Medicine (TOPMed) recommendations[18] to define ancestries based on genetic information (see section 2.3.3).

The plasma proteomic and metabolomic datasets were generated from participants, which included 1254 AD patients, 1720 healthy controls, 34 frontotemporal dementia patients, and 162 individuals with an unclassified neurodegenerative disease. The cohort was a subset

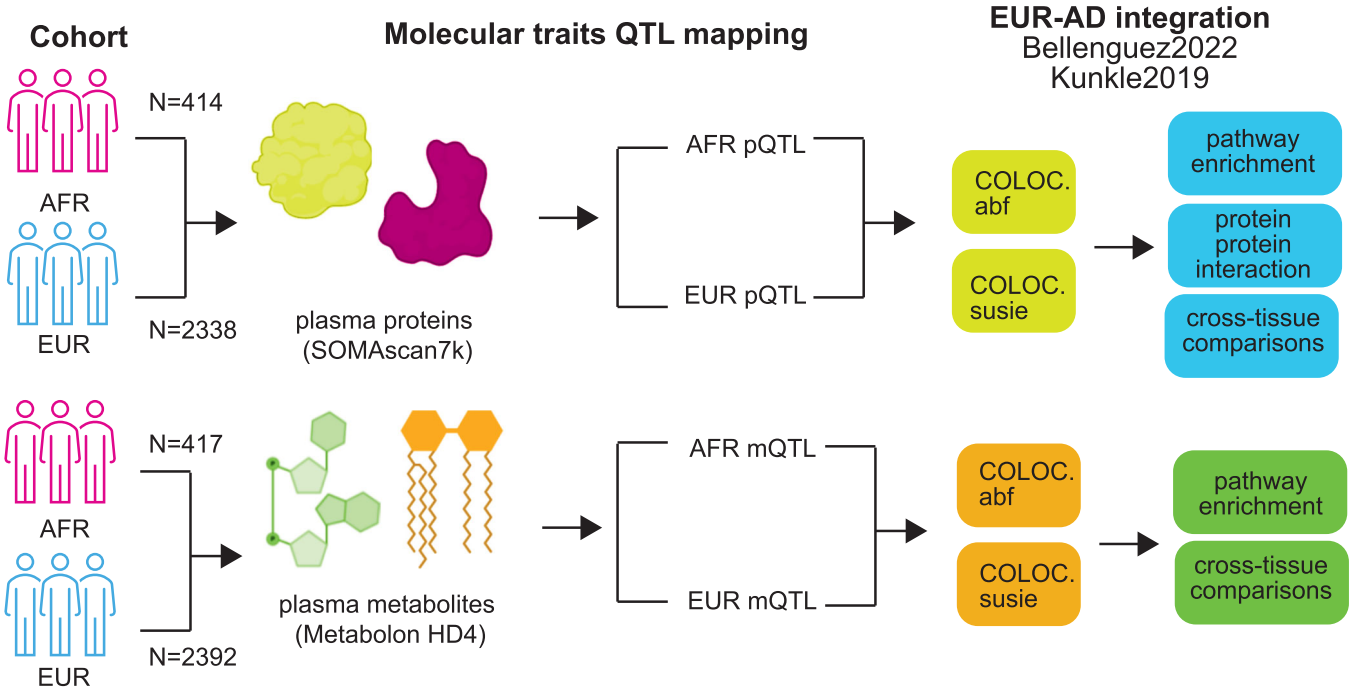


FIGURE 1 Schematic of study design to identify ancestry-aware AD risk effector proteins and metabolites by integrating EUR and AFR genetic ancestry stratified pQTL and mQTL datasets (left) using pairwise-trait genetic colocalization methods (middle). Biological functions (right) were annotated with pathway enrichment analyses, cross-tissue comparisons, and protein–protein interaction network (specific to proteomic findings). AD, Alzheimer's disease; AFR, African; EUR, European; mQTL, metabolite quantitative trait locus; pQTL, protein quantitative trait locus.

of the participants recruited from the Knight ADRC,[19] which includes community-dwelling adults > 27 years old via prospective studies of memory and aging since 1979. All participants recruited from the Knight ADRC are required to participate in core study procedures, including annual longitudinal clinical assessments, neuropsychological testing, neuroimaging, and biofluid biomarker studies. If such a participant was genotyped already in our cohort, we prioritized them for multi-omics profiling. The corresponding genotype was a priori to choosing participants for profiling proteomics and metabolomics, specifically in this study. Plasma samples were collected in the morning after an overnight fast, immediately centrifuged, and stored at −80°C.

2.3 | Data quality control (proteomics, metabolomics, genotype data quality control)

2.3.1 | Proteomics data quality control

In brief, 3132 participants and 6907 aptamers passed proteomics quality control (QC). Seven thousand five hundred eighty-four aptamers were measured before proteomics QC using the SomaScan 7k platform.[20] Plasma proteomics data from all genetic ancestries were QCed[17] with seven steps. Step (1)—Limit of detection, scale factor difference, and coefficient of variation: we kept proteins/aptamers with ≥ 85% limit of detection, ≥ 0.5 scale factor difference, ≥0.15 coefficient of variation. Step (2)—Interquartile range (IQR)-based outlier expression level detections: we log10-transformed the values and detection the outlier per the first and third quantile for larger than 1.5-fold of IQR. Step (3)—Remove analytes and samples with < 65% call rate. Step (4)—Recalculate call rate for analytes and remove analytes with call rate < 85%. Step (5)—Recalculate missing rate for subjects and remove subjects with < 85% call rate cut-off. Step (6)—Back transform into raw values from log10-scale. Step (7)—Remove non-human and analytes without protein targets and output the final matrix.

2.3.2 | Metabolomics data QC

Briefly, 3169 participants and 1508 metabolites passed metabolomics QC. One thousand seven hundred eighteen metabolites were measured before QC with the Metabolon HD4 platform.[21] Plasma metabolomics data from all genetic ancestries were QCed[17] with 11 steps. Step (1)—Normalize volume. Step (2)—Sample missingness: we kept samples < 50% missingness. Step (3)—Metabolite missingness: we kept metabolites from non-xenobiotics groups < 80% missingness. Step (4)—Perform Fisher exact test and differential expression check to recover metabolites if their missingness was associated with disease status. Step (5)—Perform minimum value imputation on non-xenobiotic group of metabolites. Step (6)—Remove non-informative metabolites after log10-transformation: non-informative metabolites were defined as IQR = 0 and variance < 0.001. Step (7)—Perform IQR-based outlier detection: remove outlier values based on if they exceed the first and third quantile for > 1.5 x IQR. Step (8)—Remove metabolites with < 50 data points. Step (9)—Remove sample outliers based

on metabolite principal component analysis (PCA): > 5 standard deviation of the PC1 and PC2. Step (10)—Batch effects of metabolomics data: Post QC, two different batches were identified. The batch effect was found by sample drawing year, such that samples drawn from 1998 to 2006 clustered together, whereas those from other years were clustered into separate groups. But after adjusting the metabolite PC1 and PC2, the batch effects can be corrected. Based on the model test, it was decided that PC1 and PC2 would be used as covariates in downstream analyses. Step (11)—Output the final matrix and back transform metabolite levels. Notably, for step 5, missing values were imputed. However, as mentioned earlier, xenobiotics are indeed expected to be missing, and imputing their values could skew the results. So, imputation is performed only for the non-xenobiotic group of metabolites.

2.3.3 | Genotype QC, imputation, and ancestral group stratification

At the pre-imputation stage, the directly genotyped variants were kept, agreeing to three criteria: (1) genotyping success rate ≥ 98% per variant or per individual; (2) minor allele frequency (MAF) ≥ 0.01; and (3) Hardy–Weinberg equilibrium (HWE; $p \geq 1\times10^{-6}$). Imputation was performed on the TOPMed[22] imputation server using the hg38 Version R2 reference panel. The TOPMed Imputation Reference panel contains information from 97,256 deeply sequenced human genomes. Imputed genotypes with an imputation quality of R2 ≥ 0.3 were kept. At the post-imputation stage, the genotyped and imputed variants remained on the two criteria: (1) genotyping missing rate ≤90% per variant; (2) MAF ≥ 0.0005. Multiple genotyping arrays were included for this cohort, including CoreEx, GSA_v1, GSA_v2, GSA_v3, Human1M.Duov3, NeuroX2, OmniEx, quad660. For each genotype array, we performed pre-imputation, imputation, and post-imputation separately. We merged into one dataset before performing the QTL analyses. Thus, the final number of genetic variants located on autosomal chromosomes was 10,448,203. In total, 3081 out of 3170 participants had corresponding genotype data.

Ancestral group stratification was performed using Plink1.9[23] pca function (after linkage disequilibrium [LD] pruning); 2598 participants were classified as EUR and 433 as AFR per PCA with the reference by the 1000 Genome Project.[24] We defined the genetic ancestry per genotype PCA anchored with participants from the 1000 Genome Project within the boundary with the mean ± 3 times the standard deviation of each PC. Moreover, we noted that > 400 AFR participants we defined as AFR were within three standard deviations of the AFR participants in the 1000 Genome Project, we cannot omit the admixed ancestral groups within these participants. Relatedness was performed using plink1.9[23] genome function on the IBD. Unrelated participants were defined as PI_HAT < 0.25; 2395 EUR and 418 AFR participants were kept as unrelated participants.

2.4 | Integration of proteomics and metabolomics with genotype data

The four sets of plasma protein QTL (pQTL; EUR and AFR) and metabolite QTL (mQTL; EUR and AFR) were generated with an additive linear regression model from the plink2[23] glm function. Details for deriving the ancestry-stratified xQTL sets were documented in Yang *et al.*[17] Briefly, we identified 2400 proteins with 2848 significant pQTLs from the EUR set and 881 proteins with 954 significant pQTLs from the AFR set. We identified 490 significant mQTLs associated with 403 metabolites from the EUR set and 65 significant mQTLs associated with 60 metabolites from the AFR set.

2.5 | AD risk GWAS

We used the latest population-scale AD risk GWAS of EUR ancestry by Bellenguez *et al.*[2] The study contained 111,326 clinically diagnosed/"proxy" AD cases and 677,663 controls. The proxy-AD is defined per questionnaire data in which participants are asked whether their parents had dementia. The summary statistics file of the Stage 1 dataset was downloaded from the GWAS catalog.

We used another relatively recent population-scale AD risk GWAS of EUR ancestry by Kunkle *et al.*[1] The study contained 35,274 clinically diagnosed AD cases and 59,163 controls. The summary statistics file was downloaded after requesting both Stage 1 and Stage 2 data.

In Table S5, we summarized the AD risk locus information by Bellenguez *et al.*[2] If the locus was old (defined as the locus was identified in earlier vs. the most recent study in 2022), we kept the results using the study by Kunkle *et al.*;[1] if the locus was new (defined as the locus was not identified in earlier vs. the most recent study in 2022), we kept the results using the study by Bellenguez *et al.* The findings related to the apolipoprotein E (*APOE*) locus as the most significant AD risk locus were kept, considering that both studies gave variants nearby that were not complete in either study.

2.6 | Colocalization of molecular traits and AD

To rigorously test the variant sharing between proteins/metabolites and AD, we performed pairwise colocalization analysis using both coloc.abf function from the R package coloc v3.1[7] and coloc.susie function from the R package coloc v5.1[10] with a wrapper for the susie_rss function from the susieR[25,26] package. We next set the window size to ± 1Mb, centering on IV per trait–AD pair. We used default priors, with p1 as 1×10^{-4}, p2 as 1×10^{-4}, and p12 as 1×10^{-5}. Evidence for colocalization was assessed using the posterior probability (PP) for hypothesis 4 (indicating there is an association for both protein and disease, and they are driven by the same causal variant[s]). We

used PP.H4_final > 80% as a threshold to suggest that two traits were highly colocalized. Under the assumption of only a single causal variant, we used the PP.H4 from coloc.abf output of the trait–disease pair. Under the assumption that multiple causal variants exist, we used the maximum PP.H4 of multiple credible sets from coloc.susie output.

Geni.plot R package[27] was used to draw the customized stacked regional plot given a single AD locus.

2.7 | Pathway and cell type enrichment of proteins in *trans* association with the AD loci

To test the biological pathway enrichment of proteins pinpointed via colocalization in EUR and AFR stratified analyses, given a *trans* association, we used the enrichGO function from the ClusterProfiler package.[28] The background list contained all SomaScan 7k proteins that passed QC. The top three ranked pathways being enriched were selected based on *p* value. False discovery rate (FDR)–based *q* value < 0.05 was used as a significance threshold in both EUR- and AFR-stratified analyses.

To test the cell type enrichment of the proteins pinpointed via colocalization in EUR- and AFR-stratified analyses, given a *trans* association, we used the gene expression information curated from two recent single-nuclei RNA-seq datasets[29,30] containing > 700 participants. We computed the cell type proportion of each gene using the pseudo-bulk strategy across all seven major cell types. To define the protein as cell type specific, we required its corresponding gene expression to have at least a half-fold higher in one cell type compared to any other cell type. To test for the enrichment, we calculated the odds ratio via Fisher exact test for each cell type within each locus. The top-ranked cell type being enriched was mentioned based on the *p* value. FDR *q* value < 0.1 was used as a significance threshold in both EUR- and AFR-stratified analyses.

2.8 | Identification of *APOE* ε4 conditional pQTLs and colocalization with AD

An additive linear regression model was used from plink2[23] glm function for each protein showed high colocalization with AD and *trans*-associated a pQTL near the *APOE* locus. Protein abundances were log-10 transformed first, and *z* scale normalized next, as consistent with the pre-conditional pQTL identification per Yang *et al*.[17] Covariates were the *APOE* ε4 variant rs429358, along with the original age, sex, genotyping array types, genotype PC 1–10, and proteomics PC 1–2 per Yang *et al*. For the *trans*-pQTL analysis, we used the same study-wide significance as pre-conditional pQTL identification for EUR was 3.40×10^{-11} and for AFR was 1.49×10^{-10}.

Colocalization analyses were performed using the post-conditional pQTL summary statistics and AD risk following the same thresholds of PP.H4_final > 80% to suggest that two traits were highly colocalized.

2.9 | Pathway enrichment of proteins and metabolites implicated in AD

To test the pathway enrichment of proteins pinpointed via colocalization in EUR- and AFR-stratified analyses, we used enrichGO and enrichDO functions from the ClusterProfiler package.[28] The background list contained all SomaScan 7k proteins that passed QC. The significance threshold of the pathway being enriched was determined by FDR-based *q* value: EUR proteins with *q* value < 0.1 and AFR proteins with *q* value < 0.2 (if using *q* value < 0.1 for AFR proteins, only the top 2 Disease Ontology terms [syphilis and frontotemporal dementia] remained significant). Treemap plots were drawn using the function from the rrvgo R package[31] when clustering different clusters of terms.

To test the pathway enrichment of metabolites pinpointed via colocalization in EUR- and AFR-stratified analyses, we used the Fisher exact test testing against Metabolon's officially annotated sub-pathway information.[32] The outcome was tested based on the odds ratio of the metabolite highly colocalized (sub-pathway A alone over all other pathways) over the metabolite passed QC (sub-pathway A alone over all other pathways). FDR-based *q* value < 0.1 was used as a significance threshold in both EUR- and AFR-stratified analyses. PathBank[33] was used for visualizing the metabolites in the corresponding pathways.

2.10 | Protein–protein interaction prediction of proteins implicated in AD

The protein–protein interaction (PPI) networks were obtained using the STRING database.[34] The network plots were visualized with rbioapi R package.[35] The combined score, as well as gene neighborhood score, gene fusion score, phylogenetic profile score, co-expression score, experimental score, database score, and textmining score, were included. The confidence score was used to determine how likely STRING concludes an interaction to be true. A default of medium confidence, ≥ 0.4, was chosen.

2.11 | Cross-tissue comparisons of proteins and metabolites implicated in AD

To compare the protein and metabolite findings in this study, we queried the output from two previous (also from our lab) studies of CSF pQTL[3] and mQTL[4] on AD, separately, against the plasma EUR and AFR protein/metabolite lists.

3 | RESULTS

3.1 | Genetic colocalization of proteins and AD

To pinpoint which proteins are key effectors of AD risk, we conducted genetic colocalization of AD risk GWAS loci with the plasma pQTLs,

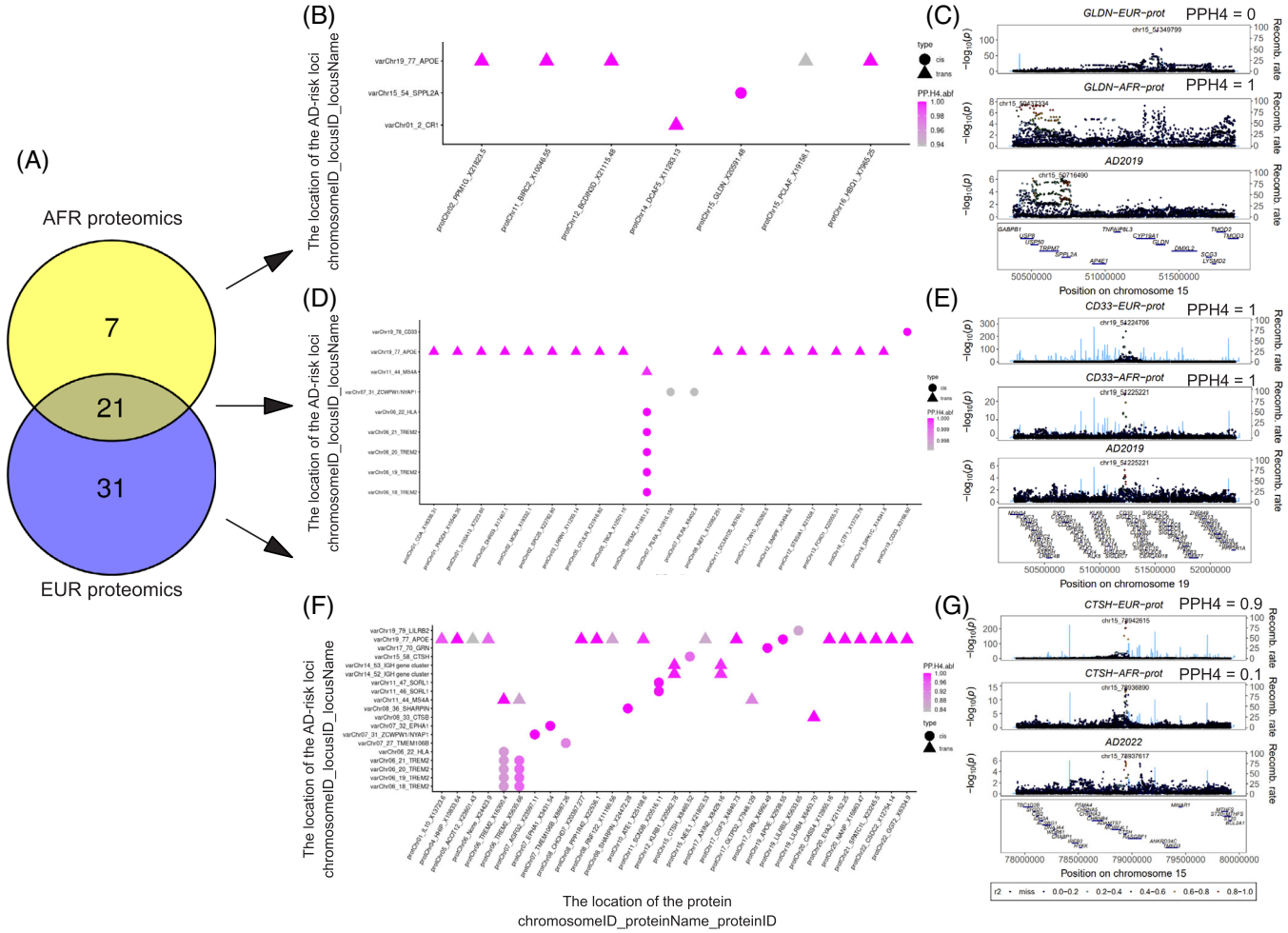


FIGURE 2 Genetic colocalization results between proteins and AD. A, Venn diagram of EUR and AFR proteomics findings. B, Dotplot of seven proteins specific to AFR set with *X* axis as the protein chromosome location and *Y* axis as the genetic variant chromosome location within the AD risk locus ID. The shape represents the type of pQTL (used for centering the colocalization) as *cis* a circle, and *trans* as a triangle; the dot color represents the PP.H4 value, the darker the higher. C, Stacked regional plot of protein GLDN as one of the AFR-specific findings from the EUR set, AFR set, and the AD risk GWAS by Kunkle *et al.*,[1] given that it was an old locus. The *X* axis represents the genetic variant chromosome location within the AD risk locus ID, and the *Y* axis as the negative log10 *p* value for each variant association. The color code corresponds to the linkage disequilibrium r^2 value relative to the index variants. D, Dotplot of 21 proteins shared between the AFR and EUR set. The PP.H4 value for each protein locus is based on a larger value from the EUR and AFR sets (but both passed > 0.8 threshold). E, Stacked regional plot of protein CD33 as one of the EUR and AFR shared findings from the EUR set, AFR set, and the AD risk GWAS by Kunkle *et al.*,[1] given that it was an old locus. F, Dotplot of 31 proteins specific to the EUR set. Note: PP13G is labeled None as aptamer ID X24423.9 per SomaScan7k metadata. G, Stacked regional plot of protein CTSH as one of the EUR-specific findings from the EUR set, AFR set, and the AD risk GWAS by Bellenguez *et al.*[2] given it was a new locus. AD, Alzheimer's disease; AFR, African; CTSH, cathepsin H; EUR, European; GLDN, gliomedin; GWAS, genome-wide association study; pQTL, protein quantitative trait locus.

focusing on all proteins with at least one study-wide signal, either *cis* or *trans* (Figure 2; Tables S6 and S7). To achieve this, we first leveraged two ancestry-stratified pQTL atlases derived from 2338 EUR and 414 AFR participants separately from the Knight ADRC cohort at Washington University.[17] In this previous study, we identified 2848 pQTLs from the EUR participants and 954 pQTLs from the AFR participants. Colocalization analyses were next performed between proteins from both ancestries and the AD risk GWAS by Bellenguez *et al.*,[2] and Kunkle *et al.*,[1] as Bellenguez *et al.* only shared their Stage 1 summary statistics, and we found several disease-associated loci were reducing the reported negative log10-base *p* value by 50% given both Stage 1 and Stage 2 participants (Figure S1), including variants near the locus of *APOE*, *MS4A*, *ACE*, and *CD33*. Thus, for all 84 indexed loci defined by Bellenguez *et al.*, if the locus was previoulsy reported as old (Table S5), we used the study by Kunkle *et al.*, instead. In theory, analyses using AFR ancestry omics datasets will be ideal, integrating with larger AFR-matched GWAS equivalent to the sample size of EUR GWAS (see Discussion for details).

Using this approach, we identified 21 proteins that highly colocalized in both EUR and AFR ancestry groups (genetic colocalization's posterior probability of two traits sharing the same genetic signal [PP.H4] > 0.8; Figure 2A; Tables S6 and S7). These 21 ancestry-shared proteins (Figure 2A,D) included NEFL, LRRN1, TBCA, CTF1, PHGDH, ST8SIA1, FOXO1, SPC25, DCUN1D5, TREM2, DHRS9,

CDA, CD33, SNRPF, ZW10, OTULIN, DIPK1C, MOB4, S100A13, and PILRA (two aptamers as X10816.150 and X6402.8). Specifically, the colocalized association of CD33 (Figure 2E), TREM2, and PILRA (two aptamers) with AD risk was driven by *cis* pQTLs and the other 17 proteins by *trans* pQTL at a known locus, the *APOE* region.

We also identified seven ancestry-specific proteins that colocalized with AD in AFR (Figure 2A,B and Table S6) and 31 proteins in EUR (Figure 2A,F and Table S7). The seven AFR-specific proteins associated with AD risk included BIRC2, DCAF5, BCDIN3D, HBQ1, GLDN, PPM1G, and PCLAF. Only GLDN (Figure 2C) was driven by a *cis* signal near the AD locus *SPPL2A*. For the other six proteins associated with *trans* loci: BIRC2, BCDIN3D, HBQ1, PPM1G, and PCLAF were driven by variants in *APOE*, and DCAF5 by variants in *CR1*. These proteomic findings observed in participants of AFR ancestry warrant validation with AFR-matched GWAS.

Among the 31 unique proteins highly colocalized in the EUR-stratified analysis, 11 (CTSH [Figure 2G], apoE, AGFG2, EPHA1, SCN3B, TREM2 [two aptamers X5635.66 and X16300.4], TMEM106B, SHARPIN, GRN, and LILRB2) were driven by *cis* signals, and 21 by *trans* signals, mainly due to *APOE* (locus ID as 77, in association with 16 proteins, in Table S6). TREM2 colocalized with AD via a *cis* signal, but also with *trans* signals in the *MS4A* region, as reported previously in CSF.[36,37]

To determine which proteins had been previously reported to be associated with AD, we reviewed the literature on plasma protein effects on AD. In a recent study by Yao *et al*.,[38] the authors used a new method of Mendelian randomization (termed MR-SPI after selecting valid pQTL IVs and then performing robust post-selection inference), identifying seven proteins significantly associated with AD. The proteins were measured using the Olink platform (912 proteins included) from the UK Biobank Pharma Proteomics Project[39] and used Jansen *et al*.,[40] for the AD GWAS. The proteins identified by Yao *et al*., included CD33, CD55, EPHA1, PILRA, PILRB, RET, and TREM2. Of those, our findings replicated four proteins (EPHA1, TREM2, PILRA, and CD33) in the EUR-stratified analyses and three proteins (TREM2, CD33, and PILRA) in the AFR-stratified analyses.

In total, our study identified 52 novel proteins compared to the study by Yao *et al*. Of these, 45 were found in the EUR findings and 23 in the AFR findings. The differences may be due to four reasons: (1) analytical methods were different: we used colocalization, while Yao *et al*. used Mendelian randomization. This major difference suggested the variants used in the analyses were different, as Mendelian randomization had a strict assumption of no pleiotropy, while colocalization has no such assumption. (2) The AD GWAS were two different studies: we used Bellenguez *et al*.,[2] and Kunkle *et al*.,[1] while Yao *et al*. used Jansen *et al*.[40] (3) The proteomic platform was different (we used SomaScan, Yao *et al*. used Olink), and the coverage was different (we used 7k proteins, Yao *et al*., used < 1k). (4) The power given the sample size of the UKB-PPP preprint version is much larger than our study (20-fold).

3.2 | Comparisons of reported AD effector genes identified proteins as alternative effectors

In the AD-GWAS by Bellenguez *et al*.,[2] and Kunkle *et al*.,[1] the authors reported a total of 84 index variants (75 independent loci) containing ≈ 600 potential functional genes, each supported by at least one piece of evidence from colocalization and transcriptome-wide association studies (TWASs) with multi-tissue expression QTL (eQTLs), multi-tissue splicing QTL (sQTLs), and brain pQTLs.

Our analyses support some of the effector genes nominated by Bellenguez *et al*., and Kunkle *et al*. For the AFR-stratified analyses (Table S6), we nominated four functional genes (*PILRA, TREM2 [X11851.21], CD33, OTULIN*) that align with those reported by Bellenguez *et al*.,[2] and Kunkle *et al*.[1] In our EUR proteomic analysis (Table S7), we nominated 13 functional genes (*APOE, EPHA1, TREM2 [X11851.21, X5635.66, and X16300.4], SHARPIN, CD33, GRN, PILRA, OTULIN, CTSH, TMEM106B, LILRB2, CASS4, AGFG2*) that overlap with those previously reported.

To investigate which proteins were nominated as alternative effector genes (rather than the nearest gene), we searched all functional genes for each AD risk locus from these two GWASs. From the EUR findings, we identified three alternative proteins located near the known loci in *cis*: For the *NYAP1* locus, two proteins, PILRA (two aptamers) and AGFG2, colocalized in our analyses, suggesting that these could be the effector genes. From the AFR findings, we identified three alternative proteins located near the known loci in *cis*: the *NYAP1* locus had the PILRA protein (two aptamers). We also nominated 36 novel proteins from EUR and 23 from AFR not reported by the original GWAS papers (Tables S6 and S7). Most of these findings were in *trans* associations, driven by ten unique pleiotropic loci, seven in EUR (Figures S2A,B and S3) and three in AFR (Figures S2C,D and S4). These loci that colocalized with AD risk in EUR (Figures S2A and S3) included *CTSB* (locusID as 33 with one protein: LIRB4), *MS4A* (locusID as 44 with one protein: GLTD2), two loci near *SORL1* (locusID as 46 and 47 each with one protein: SCN3B), two loci near *IGH* (locusID as 52 and 53 each with two proteins: KLRB1 and AXIN2), and *APOE* (locusID as 77 with 31 proteins, Table S7). There were three genetic loci associated with 23 proteins from the AFR findings (Figures S2C and S4): *CR1* (locusID as 2 with one protein: DCAF5), *SPPL2A* (locusID as 54 with one protein: GLDN), and *APOE* (locusID as 77 with 21 proteins).

These newly reported associated proteins were enriched in various biological pathways. For the EUR-stratified findings (Figure S2B), the protein associated with *CTSB* (locusID 33) was found to be enriched in the T cell tolerance induction pathway; *MS4A* (locusID 44) was enriched in processes involving lipid transfer or transport; *SORL1* (locusID 46 and 47) was enriched in positive regulation of heart rate; and near *IGH* (locusID 52 and 53) were both enriched in heart field specification. For the AFR proteins (Figure S2D), *CR1* (locusID 2) was involved in the negative regulation of fatty acid metabolic or lipid biosynthesis processes; *SPPL2A* (locusID 54) was overrepresented in pathways of microvillus organization and neuron maturation. These proteins were also enriched in diverse cell types (Figures S3 and S4). The EUR protein associated with *CTSB*, LIRB4, was found to be

enriched in microglia or immune cells (FDR = 0.09); the EUR protein *SORL1* was enriched in excitatory neurons (FDR = 0.088). Both EUR and AFR proteins associated with *APOE* were enriched in vascular cell types (FDR = 3.4×10^{-8} for EUR and 0.06 for AFR).

3.3 | *APOE ε4* variant rs429358 conditional analyses

To determine which *APOE*-driven *trans*-pQTLs remain after accounting for *APOE ε4* variant directly, we performed conditional analyses of *APOE ε4* variant rs429358 on the 33 EUR proteins and 22 AFR proteins reported in this study that showed high colocalization results with AD risk (Figure S5). Of the 33 EUR proteins, 26 pQTLs remained study-wide significant; of the 22 AFR proteins, six pQTLs remained study-wide significant. These still-significant proteins may have direct effects independent of the *APOE ε4* variant. Additionally, we performed colocalization analyses using the post-conditional pQTL summary statistics with the AD risk GWAS and found that 14 EUR and 4 AFR proteins remained highly colocalized with AD.

3.4 | Genetic colocalization of metabolites and AD

In parallel, we examined which metabolite QTLs also colocalize with AD risk loci in an ancestry-stratified manner. We considered all metabolites with at least one mQTL (Figure 3A,B; Tables S8 and S9). Metabolite levels were measured in 2392 EUR and 417 AFR participants from the same cohort as proteomics. In the same previous study[17] that included ancestry-stratified pQTLs mentioned above, we identified a total of 65 AFR and 490 EUR mQTLs.

Overall, among the metabolites that colocalized with AD risk (PP.H4 > 0.8), we found one shared metabolite, 1-palmitoyl-2-arachidonoyl-glycero-phosphatidyl-ethanolamine (GPE; 16:0/20:4), between the EUR and AFR ancestry groups (Figure 3C). For the ancestry-specific findings, we nominated 11 metabolites from the EUR-specific analyses (Figure 3A,B) and 1 metabolite for AFR (Figure 3A,B). Moreover, for the 11 EUR metabolite findings, all 9 GPEs were associated with the locus colocalized with *ADAM10* (AD-locusID as 55). An unknown metabolite, X-24309, was associated with two other AD loci near *PTK2B* and *CLU*. For the AFR metabolite finding (6-oxopiperidine-2-carboxylate), the only locus colocalized with AD was *SHARPIN* (locusID as 36). The nine EUR known metabolites associated with AD (Table S9) were annotated as part of the phosphatidyl–ethanolamine (PE biochemical sub-pathway, which is under the lipid super-pathway (Figure 3B). The metabolite identified in the AFR-specific analyses (Figure 3B) was 6-oxopiperidine-2-carboxylate (Figure 3D), classified under the lysine metabolism biochemical sub-pathway, belonging to the amino acid super-pathway. Additional studies using AD GWASs from AFR ancestry would be needed to validate these findings.

To determine whether our metabolic findings had been previously reported, we compared our results to a recent study by Chen *et al.*,[41] who measured 13 metabolites using an in-house mass spectrometry platform. None of the 12 metabolites identified in our EUR and AFR sets were were measured by Chen *et al*. For the metabolites from the lipid super-pathway, we used an additional lipidomic profiling study by Liu *et al.*[42] From our EUR set, 5 of 10 lipids were measured in Liu *et al*. Two lipids, 1-palmitoyl-2-arachidonoyl-GPE (16:0/20:4) and 1-palmitoyl-2-oleoyl-GPE (16:0/18:1), were significantly different between AD and controls. From the AFR set, only lipid 1-palmitoyl-2-arachidonoyl-GPE (16:0/20:4) was significantly different between AD and controls. Additionally, we found no common genetic loci between the AD-colocalized metabolites and proteins in both the EUR and AFR sets (Figures 2 and 3). No AD risk loci were shared between metabolites and proteins. This negative observation may be attributed to the smaller coverage of the metabolomics dataset (≈ 1.5k metabolites) compared to the proteomics dataset (≈ 7k proteins).

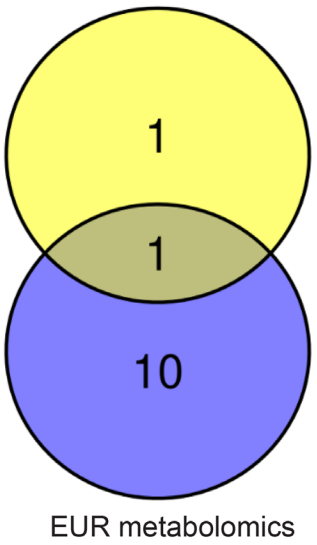
3.5 | Biological insights of proteins implicated in AD

To elucidate the potential mechanisms by which the identified proteins contribute to AD risk, we conducted two in silico analyses: pathway enrichment analysis, PPI analysis. For the pathway enrichment analysis of the EUR-stratified proteins (Figures 4A,B, and S6A; Table S10), the most significant (q value < 0.1) pathways included the immune system, with three clustered Gene Ontology (GO) terms as negative regulation of interleukin-1 production (four proteins: TREM2, CD33, IL10, LILRB4), negative regulation of cell activation (seven proteins: TREM2, CD33, IL10, LILRB4, GRN, LILRB2, apoE), neuroinflammatory response (six proteins: OTULIN, TREM2, IL10, SHARPIN, GRN, apoE), along with disease enrichment terms such as abdominal obesity-metabolic syndrome, frontotemporal dementia, and syphilis.

In contrast, for the AFR proteins (Figures 4C,D, and S6B; Table S11), the most significant terms were from the disease enrichment analysis (q value < 0.1): syphilis and frontotemporal dementia, both with the same two proteins, NEFL and TREM2, as input. Instead, all the biological pathway terms were significant with a q value < 0.2, clustered into three categories: attachment of spindle microtubules to kinetochore (two proteins: SPC25, ZW10), negative regulation of interleukin-1 beta production (two proteins: TREM2, CD33), and regulation of small molecule metabolic process (four proteins: FOXO1, DCAF5, TREM2, CDA). The PPI analysis, in the EUR-stratified findings (Figure 4E; Table S12), revealed that apoE interacted with its receptors (LILRB2, LILRB4) and the immune axis (TREM2, CD33, IL10). TREM2 was found on the cell membrane or released into the extracellular space. apoE was released into the extracellular space or found in the endosome. CD33, LILRB2, and LILRB4 were exclusively located on the cell membrane. apoE and TREM2 interacted through co-expression and experimental/biochemical data. apoE and LILRB4/LILRB2 interacted through co-expression. apoE and CD33 interacted through phylogenetics.

The PPI analyses based on the AFR-stratified findings (Figure 4F; Table S13) indicate that TREM2 was interacting with CD33, NEFL, and

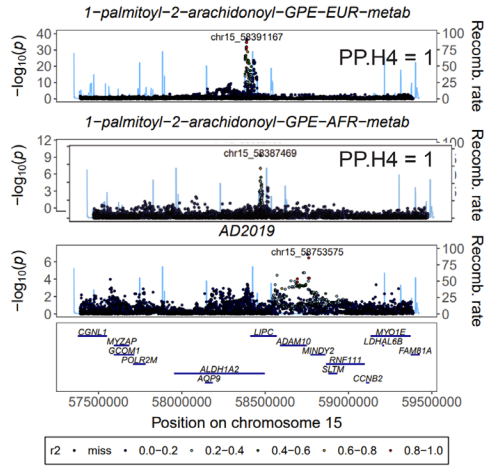
(A)



(B)

pop	CHEMICAL_NAME	SUB_PATHWAY	varCHR_locusID	indep.locusName	PP.H4.abf
AFR	6-oxopiperidine-2-carboxylate	Lysine Metabolism	varChr08_36	SHARPIN	1.00
AFR	1-palmitoyl-2-arachidonoyl-GPE (16:0/20:4)*	Phosphatidylethanolamine (PE)	varChr15_55	ADAM10	1.00
EUR	1-palmitoyl-2-linoleoyl-GPE (16:0/18:2)	Phosphatidylethanolamine (PE)	varChr15_55	ADAM10	1.00
EUR	1-stearoyl-2-linoleoyl-GPE (18:0/18:2)*	Phosphatidylethanolamine (PE)	varChr15_55	ADAM10	1.00
EUR	1-stearoyl-2-arachidonoyl-GPE (18:0/20:4)	Phosphatidylethanolamine (PE)	varChr15_55	ADAM10	1.00
EUR	1-palmitoyl-2-arachidonoyl-GPE (16:0/20:4)*	Phosphatidylethanolamine (PE)	varChr15_55	ADAM10	1.00
EUR	1-palmitoyl-2-docosahexaenoyl-GPE (16:0/22:6)*	Phosphatidylethanolamine (PE)	varChr15_55	ADAM10	1.00
EUR	1-stearoyl-2-docosahexaenoyl-GPE (18:0/22:6)*	Phosphatidylethanolamine (PE)	varChr15_55	ADAM10	1.00
EUR	1-oleoyl-2-linoleoyl-GPE (18:1/18:2)*	Phosphatidylethanolamine (PE)	varChr15_55	ADAM10	1.00
EUR	1-stearoyl-2-dihomo-linolenoyl-GPE (18:0/20:3n3 or 6)*	Phosphatidylethanolamine (PE)	varChr15_55	ADAM10	1.00
EUR	1-oleoyl-2-arachidonoyl-GPE (18:1/20:4)*	Phosphatidylethanolamine (PE)	varChr15_55	ADAM10	1.00
EUR	1-palmitoyl-2-oleoyl-GPE (16:0/18:1)	Phosphatidylethanolamine (PE)	varChr15_55	ADAM10	1.00
EUR	X-24309		varChr08_34	PTK2B	1.00
EUR	X-24309		varChr08_35	CLU	1.00

(C)



(D)

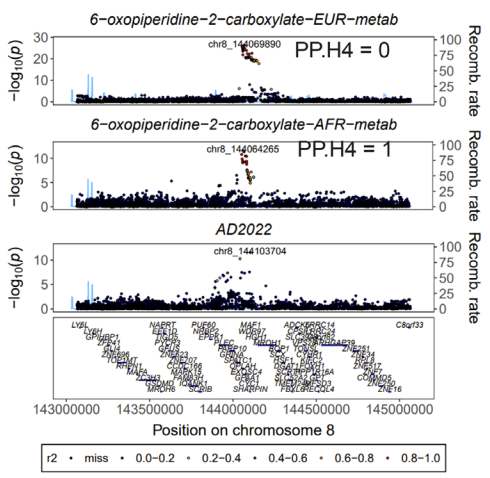


FIGURE 3 Genetic colocalization results between metabolites and AD. A, Venn diagram of EUR and AFR metabolomics findings. B, Table summarizing highly colocalized metabolites from the EUR and AFR metabolomics integration with AD risk. *Note*: X-24309 was annotated with no pathway information as an unknown category per Metabolon. C, Stacked regional plot of metabolite 1-palmitoyl-2-arachidonoyl-GPE (16:0/20:4) as the only EUR–AFR shared finding from the EUR set, AFR set, and the AD risk GWAS by Kunkle *et al*.,[1] given that it was an old locus. D, Stacked regional plot of metabolite 6-oxopiperidine-2-carboxylate as the only AFR-specific finding from the EUR set, AFR set, and the AD risk GWAS by Bellenguez *et al*.,[2] given that it was a new locus. AD, Alzheimer's disease; AFR, African; EUR, European; GWAS, genome-wide association study.

PILRA. These interactions centered on TREM2 through co-expression. Moreover, SPC25 and PCLAF interacted with each other via co-expression as well.

3.6 | Biological insights of metabolites implicated in AD

To gain further biological insights for metabolites implicated in AD risk, we performed the pathway enrichment analyses on the highly colocalized metabolite–AD pairs. For the pathway enrichment analysis of EUR metabolites (Figure 5A; Table S14), we found enrichment in the lipid super pathway. The metabolites underlying this pathway included 1-palmitoyl-2-linoleoyl-GPE (16:0/18:2), 1-stearoyl-2-linoleoyl-GPE (18:0/18:2), 1-stearoyl-2-arachidonoyl-GPE (18:0/20:4), 1-palmitoyl-2-arachidonoyl-GPE (16:0/20:4), 1-palmitoyl-2-docosahexaenoyl-GPE (16:0/22:6), 1-stearoyl-2-docosahexaenoyl-GPE (18:0/22:6), 1-oleoyl-2-linoleoyl-GPE (18:1/18:2), 1-stearoyl-2-dihomo-linolenoyl-GPE (18:0/20:3n3 or 6), 1-oleoyl-2-arachidonoyl-GPE (18:1/20:4), 1-palmitoyl-2-oleoyl-GPE (16:0/18:1). PE as the second most abundant glycerophospholipid played essential roles in mammalian cells (Figure 5B): acting as a precursor for phosphatidyl-

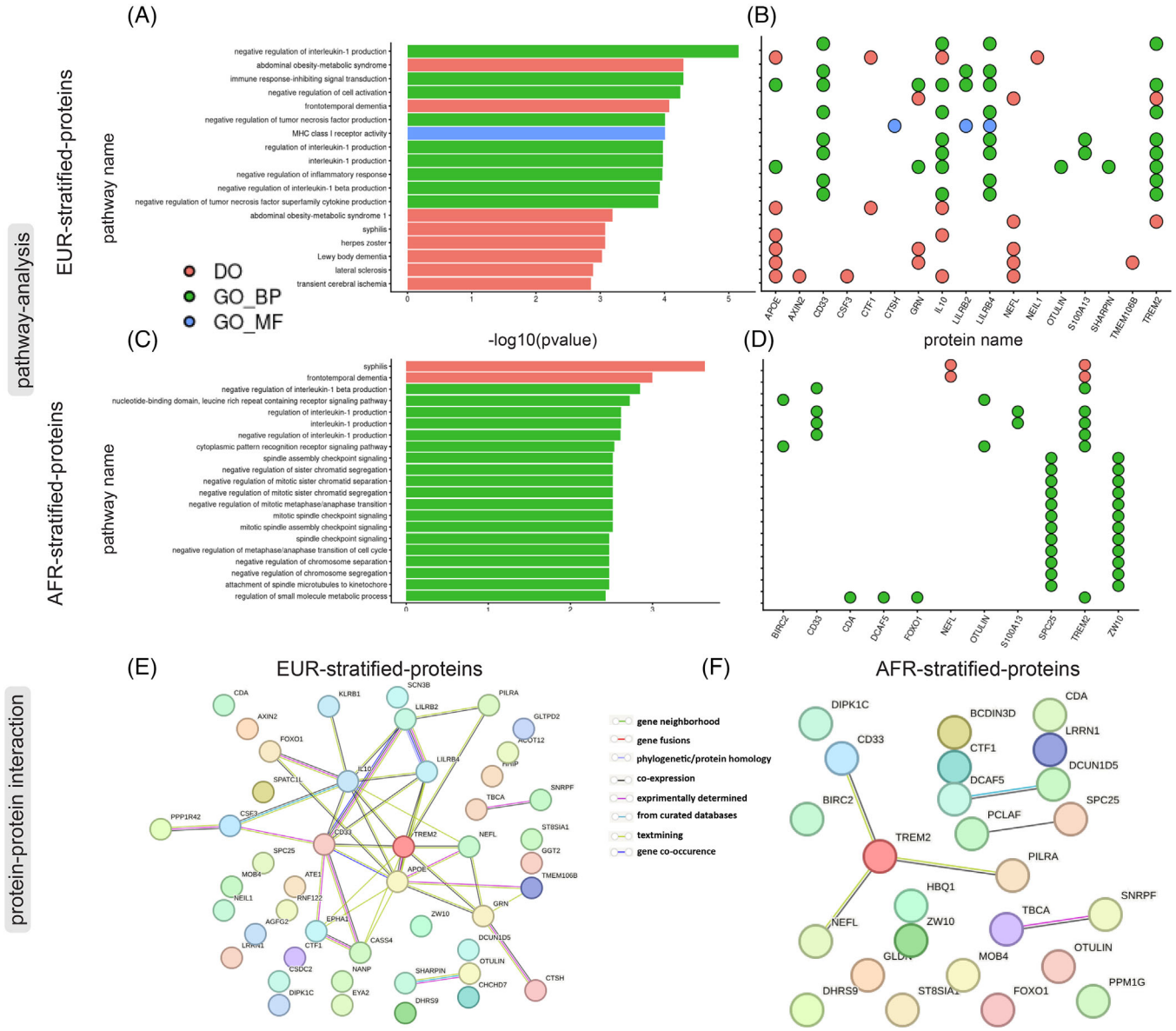


FIGURE 4 Biological insights of proteins implicated in AD. A, Bar plots of the significant GO and DO terms (q value < 0.1) enriched by proteins highly colocalized with AD risk from the EUR set (ranked by p value). The color code represents the ontology categories as gene ontology-BP as green, gene ontology-MF as blue, or disease ontology as red. B, Heatmap of the specific proteins as the X axis and the significant GO and DO terms as in (A). C, Same as (A), but proteins from the AFR set (q value < 0.2). D, Same as (C), but proteins from the AFR set (q value < 0.2). E, Network plot of all proteins highly colocalized with AD risk from the EUR set based on protein–protein interaction. F, Same as (E), but proteins are from the AFR set. The edges represent protein–protein associations, and the legend of the color is listed with different information from string-db. AD, Alzheimer's disease; AFR, African; DO, Disease Ontology; EUR, European; GO, Gene Ontology; GWAS, genome-wide association study.

choline, autophagy, and mitochondrial biogenesis. The PE metabolism has been reported in association with multiple diseases,[43] such as AD, Parkinson's disease, and non-alcoholic liver disease.

For the AFR-metabolite findings (Figure 5C; Table S15), 6-oxopiperidine-2-carboxylate was enriched in the lysine metabolism super-pathway, whereas 1-palmitoyl-2-arachidonoyl-GPE (16:0/20:4) was enriched in the lipid super-pathway. This indicates that the two ancestral groups share the PE sub-pathway under the lipid super pathway as a common enriched metabolite pathway.

3.7 | Comparison of plasma findings with the CSF findings on AD

To investigate whether there is a tissue-specific effect of the same molecular features on AD, we turned to cross-reference the EUR findings[3,4] from CSF. In brief, the CSF proteomic finding was based on 3506 samples and 7008 protein aptamers, while the CSF metabolomic finding was based on 2602 samples and 440 metabolites. Especially for the AD-related findings, 37 proteins and 5 metabolites were pinpointed with the same colocalization evidence (PP.H4 > 0.8).

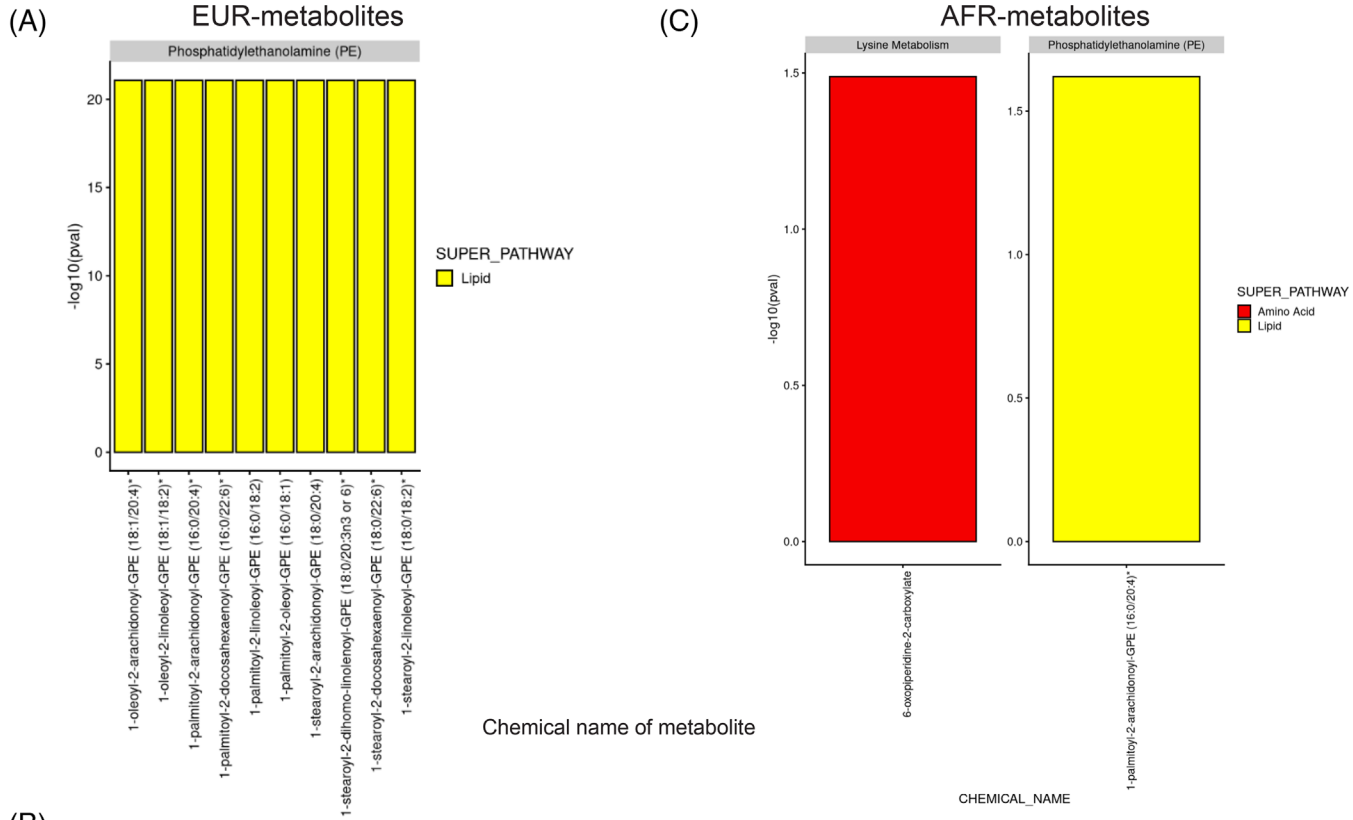


FIGURE 5 Biological insights of metabolites implicated in AD. A, Bar plot of the significant sub-pathway enriched by metabolites highly colocalized with AD risk from the EUR set (*Y* axis as −log10 [*p* value], *X* axis as the metabolite name). The color code represents the super-pathway annotated by Metabolon. B, PathBank view of the PE pathway enriched by the EUR metabolite findings. C, Same as (A), but metabolites from the AFR set. AD, Alzheimer's disease; AFR, African; EUR, European; PE, phosphatidyl–ethanolamine

In the CSF–plasma proteomic comparison, we found nine proteins in the EUR-stratified analyses (Table S16) and three in the AFR analyses (Table S17) that overlapped with the EUR CSF proteins nominated based on the AD risk GWAS by Bellenguez *et al*., alone. In the cross-tissue metabolomic comparison, we found no overlap between EUR or AFR plasma metabolites and EUR CSF metabolites nominated for the same AD risk GWAS.

4 | DISCUSSION

In this study, we implemented a systematic biology approach to identify proteins and metabolites implicated in AD by integrating four xQTL datasets with AD GWASs via colocalization. We identified 28 proteins in the AFR-stratified analyses and 52 proteins in EUR that colocalized with the AD risk, with 21 proteins common to both groups. Additionally, one metabolite in the AFR group and ten metabolites in the EUR group were uniquely colocalized with AD risk, with one metabolite being shared across ancestries. These identified proteins and metabolites were found to be enriched in convergent pathways across the AFR and EUR groups. More than half of our protein (Table S6, AFR: 72%; Table S7, EUR: 61%) and metabolite (Table S8, AFR: 50%; Table S9, EUR: 83%) findings, irrespective of ancestry stratification, have not been previously reported, demonstrating the strength of our study. It is worth noting that these findings observed in participants of AFR ancestry require validation with AFR-matched GWAS.

Plasma is the most widely used tissue type, given its greater availability compared to CSF or other body fluids. Consequently, current AD biomarker studies[44] are increasingly based on plasma rather than CSF. In our study, we provided evidence that plasma proteins and metabolites can be used to identify AD effectors underlying disease risk loci. The EUR plasma proteomic analysis revealed 46 proteins (7 in *cis* and 39 in *trans*, Table S18) that were not detected in the CSF, whereas the AFR plasma proteomic analysis identified 25 proteins (2 in *cis* and 23 in *trans*, Table S19) that were also not found in the CSF. For the metabolites, neither EUR nor AFR plasma findings were observed in the CSF. These additional proteins and metabolites helped bridge the gap from the genetic variants to AD. These molecular phenotypes serve as the product of the upstream functional genes, as well as the effectors of the downstream disease endpoint. As only limited samples were from AFR ancestry, no AFR-stratified analyses in both proteomics and metabolomics were performed in CSF to date. We thus cannot perform ancestry-matched comparisons between the plasma and CSF findings from the AFR ancestry.

We identified 21 ancestry-shared proteins, including *cis*-associated proteins CD33, TREM2, and PILRA, as well as *trans*-associated proteins, encompassing NEFL, ZW10, OTULIN, and TREM2. The shared proteomic pathways involved in the regulation of interleukin-1 production and disease ontology terms related to infectious diseases and frontotemporal dementia. One metabolite was shared between the two ancestral groups, and the lipid pathway was shared.

The extent of ancestry-specific molecular phenotypes may be attributed to the sample size (N_EUR ≈ 2300; N_AFR ≈ 400), as the EUR ancestry findings were four times higher than the AFR ancestry proteomic findings and seven times higher in metabolites. In total, we identified 2848 EUR and 954 AFR pQTLs. Additionally, we found 954 EUR and 65 AFR mQTLs. Using these xQTLs as anchors, we performed colocalization with all AD GWAS loci. Overall, we identified 60% (31/52) EUR-specific and 25% (7/28) AFR-specific proteomic findings and 91% (10/11) EUR- and 50% (1/2) AFR-specific metabolomic findings. The ancestry-specific findings could also be attributed to variant frequency, as similar observations have been reported by other groups in different molecular phenotypes, such as gene expression.[45,46]

One important caveat of this study is that even though we used ancestry-stratified xQTL sets, we linked those with AD risk via GWAS that included only EUR-ancestry individuals. A more proper analysis would involve conducting colocalization studies of AFR-specific xQTLs with AFR-specific AD GWAS loci, and similarly for EUR. We performed the AFR-stratified findings using the AFR AD risk GWAS by Kunkle *et al*.[47] This study included only 2784 AD cases and 5222 healthy controls (total $N = 8006$), which is ≈ 1% of Bellenguez *et al*. (encompassing 111,326 clinically diagnosed or proxy AD cases and 677,663 controls, total $N = 788{,}989$), and thus identified one genome-wide locus near *APOE* (p value $= 1.8\times10^{-25}$). When performing these analyses, we were not able to identify any proteins (Figure S7) or metabolites (Figure S8). The absence of findings is probably due to the significantly reduced power resulting from the 100-fold smaller sample size. We also want to note that a more recent and larger scale study by Sherva *et al*.,[48] cannot be used here, as the full summary statistics on all variants with both effect size and p value have not been released due to restrictions from the Million Veteran Program, as of November 18, 2025 (only p value for variants with $p < 1\times10^{-4}$ were released). However, we expect that AFR-stratified analyses will lead to the identification of additional proteins and metabolites implicated in AD that otherwise would not be identified using only EUR populations.

In the protein enrichment pathway analyses, we found both convergence and divergence between the EUR and AFR sets. For the convergent pathways: both EUR and AFR protein findings on AD were enriched in negative regulation of interleukin-1 production through proteins TREM2 and CD33; both ancestral groups showed disease enrichment in frontotemporal dementia and infectious diseases, given the proteins of NEFL and TREM2. For the divergent pathways, EUR proteins were enriched in neuroinflammatory response (per six proteins: OTULIN, TREM2, IL10, SHARPIN, GRN, apoE) and abdominal obesity-metabolic syndrome (per four proteins: IL10, CTF1, NEIL1, apoE), whereas AFR proteins were enriched in: attachment of spindle microtubules to kinetochore (by two proteins: SPC25 and ZW10) and regulation of small molecule metabolic process (by four proteins: FOXO1, DCAF5, TREM2, CDA). This indicates that both shared and unique pathways may contribute to AD pathogenesis in different genetic ancestral groups. Similar observations were found in metabolite pathway enrichment analyses. Both the EUR and AFR metabolomic findings on AD shared the PE sub-pathway, though for the EUR set, more metabolites were identified. For the AFR set, another significant pathway was lysine metabolism. Furthermore, both the plasma metabolite and CSF metabolite sets highlighted the lipid super path-

way, with the plasma focusing on PE and the CSF on phosphatidylcholine (PC) or sterol. This may indicate tissue-dependent effects on AD.

For the PPI result, we found an interaction between LILRB2 and TREM2. A recent study[49] dissected the molecular mechanisms of how LILRB2 mediates the TREM2 inhibition in human microglia. Moreover, the authors blocked *LILRB2* with an antibody, Ab29, and rescued the TREM2 signaling inhibition. Similarly, we found another interaction between LILRB4 and apoE, and another more recent study[50] revealed that an antibody targeting the human LILRB4 blocked its interaction with mouse apoE in a mouse model. Moreover, we identified a GRN–TMEM106B interaction. Notably, three studies[51–53] reported that knocking out both GRN and TMEM106B in a mouse model exacerbated multiple phenotypes of neurodegeneration.[54]

To provide clinical relevance to our findings, we discussed the implications for drug development (Tables S20 and S21) and biomarkers (Tables S22–S24). See Supplementary Notes in supporting information for details.

Our study has several limitations: First, our AFR set did not fully match the LD structure of the AD GWAS. We noted the rationale for not using the AFR-based AD risk studies[47,48] above. Second, our results do not encompass all genomic features, as the proteomics technology used in our study captured 7000 proteins with 2000 genes for the EUR set and 1000 genes for the AFR set used for disease integration. Metabolomics quantification further reduced the features to 1500, linking only hundreds or even tens of genes through metabolite–gene associations. Third, our study was underpowered in the AFR-stratified analyses due to a 6-fold difference in sample sizes between the EUR and AFR sets. Fourth, our study did not examine AD risk loci nor pass genome-wide significance. This may indeed miss some variants with strong xQTL evidence but with weaker AD risk association. We chose the strategy due to a much heavier computational burden than the *cis*-only eQTL studies. Fifth, we did not merge proteomics and metabolomics data before colocalizing with AD risk. Instead, we could still integrate two omic layers by comparing our pairwise trait analyses (namely, protein–AD and metabolite–AD) using locus ID as an anchor. Sixth, we did not conduct colocalization analyses of CSF versus plasma xQTLs to obtain tissue-shared/specific xQTL findings before integrating with the disease. Instead, our current strategy focuses on post hoc comparisons of proteins and metabolites prioritized for AD with the consistent threshold (PP.H4 > 0.8 per colocalization) rather than genetic variants. While this strategy might overlook tissue-specific QTLs, it avoids biases related to sample size between tissues. Seventh, our pQTL identified did not adjust for AD case–control status. This may bias *trans* findings driven by *APOE*; however, we found the three largest plasma pQTL studies[39,55,56] from other non-AD cohorts also identify the *APOE* pleiotropy region associated with 90 to 200 proteins. We argue that disease status did not play a major role in identifying such a pleiotropic region of *APOE*.

Taken together, our multi-ancestry, multi-omic findings on AD reveal both convergent and divergent molecular pathways between two ancestral populations at protein and metabolite levels. Our study underscores the importance of including participants of non-European ancestry, given the distinct proteomic and metabolomic findings observed in the African ancestral group.

ACKNOWLEDGMENTS

We thank Ciyang Wang for the helpful discussion on metabolomics-relevant findings. We thank all the participants and their families for giving consent for the research purpose. This work was supported by access to equipment made possible by the Hope Center for Neurological Disorders, the Neurogenomics and Informatics Center (NGI: https://neurogenomics.wustl.edu/), and the Departments of Neurology and Psychiatry at Washington University School of Medicine. This work was supported by grants from the National Institutes of Health (NIH; R01AG044546 to C.C., RF1AG053303 to C.C., RF1AG058501 to C.C., U01AG058922 to C.C., and RF1AG074007 to Y.J.S.), the Chan Zuckerberg Initiative, and the Alzheimer's Association Zenith Fellows Award (ZEN-22-848604, C.C.). The recruitment and clinical characterization of research participants at Washington University were supported by P30AG066444 (J.C.M., D.M.H.), P01AG03991 (J.C.M.), and P01AG026276 (J.C.M.).

CONFLICT OF INTEREST STATEMENT

C.C. has received research support from GSK and EISAI and is a member of the advisory board of Circular Genomics and owns stocks. The funders of the study had no role in the collection, analysis or interpretation of data; in the writing of the report; or in the decision to submit the paper for publication. D.M.H. co-founded, has equity, and is on the scientific advisory board of C2N Diagnostics. He is on the scientific advisory boards of Denali, Genentech, Cajal Neuroscience, and Switch Therapeutics and consults for Asteroid and Pfizer. All other authors have no conflicts of interest. Author disclosures are available in the supporting information.

CONSENT STATEMENT

The institutional review board of Washington University School of Medicine in St. Louis approved this study. All participants provided informed consent for the datasets being used in the present study.

ORCID

Carlos Cruchaga https://orcid.org/0000-0002-0276-2899

REFERENCES

1. Kunkle BW, Grenier-Boley B, Sims R, et al. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Abeta, tau, immunity and lipid processing. *Nat Genet*. 2019;51(3):414-430. doi:10.1038/s41588-019-0358-2
2. Bellenguez C, Kucukali F, Jansen IE, et al. New insights into the genetic etiology of Alzheimer's disease and related dementias. *Nat Genet*. 2022;54(4):412-436. doi:10.1038/s41588-022-01024-z
3. Western D, Timsina J, Wang L, et al. Proteogenomic analysis of human cerebrospinal fluid identifies neurologically relevant regulation and implicates causal proteins for Alzheimer's disease. *Nat Genet*. 2024;56(12):2672-84. doi:10.1038/s41588-024-01972-8
4. Wang C, Yang C, Western D, et al. Genetic architecture of cerebrospinal fluid and brain metabolite levels and the genetic colocaliza-

tion of metabolites with human traits. *Nat Genet*. 2024;56(12):2685-95. doi:10.1038/s41588-024-01973-7
5. Yang C, Farias FHG, Ibanez L, et al. Genomic atlas of the proteome from brain, CSF and plasma prioritizes proteins implicated in neurological disorders. *Nat Neurosci*. 2021;24(9):1302-1312. doi:10.1038/s41593-021-00886-6
6. Andrews SJ, Renton AE, Fulton-Howard B, Podlesny-Drabiniok A, Marcora E, Goate AM. The complex genetic architecture of Alzheimer's disease: novel insights and future directions. *EBioMedicine*. 2023;90:104511. doi:10.1016/j.ebiom.2023.104511
7. Giambartolomei C, Vukcevic D, Schadt EE, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet*. 2014;10(5):e1004383. doi:10.1371/journal.pgen.1004383
8. Hormozdiari F, van de Bunt M, Segre AV, et al. Colocalization of GWAS and eQTL signals detects target genes. *Am J Hum Genet*. 2016;99(6):1245-1260. doi:10.1016/j.ajhg.2016.10.003
9. Hukku A, Pividori M, Luca F, Pique-Regi R, Im HK, Wen X. Probabilistic colocalization of genetic variants from complex and molecular traits: promise and limitations. *Am J Hum Genet*. 2021;108(1):25-35. doi:10.1016/j.ajhg.2020.11.012
10. Wallace C. A more accurate method for colocalisation analysis allowing for multiple causal variants. *PLoS Genet*. 2021;17(9):e1009440. doi:10.1371/journal.pgen.1009440
11. Wen X, Pique-Regi R, Luca F. Integrating molecular QTL data into genome-wide genetic association analysis: probabilistic assessment of enrichment and colocalization. *PLoS Genet*. 2017;13(3):e1006646. doi:10.1371/journal.pgen.1006646
12. Zheng J, Haberland V, Baird D, et al. Phenome-wide Mendelian randomization mapping the influence of the plasma proteome on complex diseases. *Nat Genet*. 2020;52(10):1122-1131. doi:10.1038/s41588-020-0682-6
13. Zhang W, Lu T, Sladek R, Li Y, Najafabadi H, Dupuis J. SharePro: an accurate and efficient genetic colocalization method accounting for multiple causal signals. *Bioinformatics*. 2024;40(5):btae295. doi:10.1093/bioinformatics/btae295
14. Gusev A, Ko A, Shi H, et al. Integrative approaches for large-scale transcriptome-wide association studies. *Nat Genet*. Mar 2016;48(3):245-252. doi:10.1038/ng.3506
15. Gamazon ER, Wheeler HE, Shah KP, et al. A gene-based association method for mapping traits using reference transcriptome data. *Nat Genet*. 2015;47(9):1091-1098. doi:10.1038/ng.3367
16. Hemani G, Zheng J, Elsworth B, et al. The MR-Base platform supports systematic causal inference across the human phenome. *Elife*. 2018;7:e34408. doi:10.7554/eLife.34408
17. Yang C, Gorijala P, Timsina J, et al. European and African ancestry-specific plasma protein-QTL and metabolite-QTL analyses identify ancestry-specific T2D effector proteins and metabolites. *Nat Commun*. 2025;16(1):7412. doi:10.1038/s41467-025-62463-w
18. Khan AT, Gogarten SM, McHugh CP, et al. Recommendations on the use and reporting of race, ethnicity, and ancestry in genetic research: experiences from the NHLBI TOPMed program. *Cell Genom*. 2022;2(8):100155. doi:10.1016/j.xgen.2022.100155
19. Fernandez MV, Liu M, Beric A, et al. Genetic and multi-omic resources for Alzheimer disease and related dementia from the Knight Alzheimer Disease Research Center. *Sci Data*. 2024;11(1):768. doi:10.1038/s41597-024-03485-9
20. Gold L, Ayers D, Bertino J, et al. Aptamer-based multiplexed proteomic technology for biomarker discovery. *PLoS One*. 2010;5(12):e15004. doi:10.1371/journal.pone.0015004
21. Evans AM, DeHaven CD, Barrett T, Mitchell M, Milgram E. Integrated, nontargeted ultrahigh performance liquid chromatography/electrospray ionization tandem mass spectrometry platform for the identification and relative quantification of the small-molecule complement of biological systems. *Anal Chem*. 2009;81(16):6656-6667. doi:10.1021/ac901536h
22. Taliun D, Harris DN, Kessler MD, et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature*. 2021;590(7845):290-299. doi:10.1038/s41586-021-03205-y
23. 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. doi:10.1186/s13742-015-0047-8
24. Genomes Project C, Auton A, Brooks LD, et al. A global reference for human genetic variation. *Nature*. 2015;526(7571):68-74. doi:10.1038/nature15393
25. Wang G, Sarkar A, Carbonetto P, Stephens M. A simple new approach to variable selection in regression, with application to genetic fine mapping. *J R Stat Soc Series B Stat Methodol*. 2020;82(5):1273-1300. doi:10.1111/rssb.12388
26. Zou Y, Carbonetto P, Wang G, Stephens M. Fine-mapping from summary data with the "sum of single effects" model. *PLoS Genet*. 2022;18(7):e1010299. doi:10.1371/journal.pgen.1010299
27. Staley J, Palmer T, GENI plots.
28. Wu T, Hu E, Xu S, et al. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innovation*. 2021;2(3):100141. doi:10.1016/j.xinn.2021.100141
29. Fujita M, Gao Z, Zeng L, et al. Cell subtype-specific effects of genetic variation in the Alzheimer's disease brain. *Nat Genet*. 2024;56(4):605-614. doi:10.1038/s41588-024-01685-y
30. Hoffman GE, Lee D, Bendl J, et al. Efficient differential expression analysis of large-scale single cell transcriptomics data using dreamlet. *bioRxiv*. 2024;20:2023-03. doi:10.1101/2023.03.17.533005
31. Sayols S. rrvgo: a Bioconductor package for interpreting lists of gene ontology terms. *MicroPubl Biol*. 2023;2023. doi:10.17912/micropub.biology.000811
32. Pietzner M, Stewart ID, Raffler J, et al. Plasma metabolites to profile pathways in noncommunicable disease multimorbidity. *Nat Med*. 2021;27(3):471-479. doi:10.1038/s41591-021-01266-0
33. Wishart DS, Kruger R, Sivakumaran A, et al. PathBank 2.0-the pathway database for model organism metabolomics. *Nucleic Acids Res*. 2024;52(D1):D654-D662. doi:10.1093/nar/gkad1041
34. Szklarczyk D, Kirsch R, Koutrouli M, et al. The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res*. 2023;51(D1):D638-D646. doi:10.1093/nar/gkac1000
35. Rezwani M, Pourfathollah AA, Noorbakhsh F. rbioapi: user-friendly R interface to biologic web services' API. *Bioinformatics*. 2022;38(10):2952-2953. doi:10.1093/bioinformatics/btac172
36. Deming Y, Filipello F, Cignarella F, et al. The MS4A gene cluster is a key modulator of soluble TREM2 and Alzheimer's disease risk. *Sci Transl Med*. 2019;11(505):eaau2291. doi:10.1126/scitranslmed.aau2291
37. Wang L, Nykanen NP, Western D, et al. Proteo-genomics of soluble TREM2 in cerebrospinal fluid provides novel insights and identifies novel modulators for Alzheimer's disease. *Mol Neurodegener*. 2024;19(1):1. doi:10.1186/s13024-023-00687-4
38. Yao M, Miller GW, Vardarajan BN, Baccarelli AA, Guo Z, Liu Z. Deciphering proteins in Alzheimer's disease: a new Mendelian randomization method integrated with AlphaFold3 for 3D structure prediction. *Cell Genom*. 2024;4(12):100700. doi:10.1016/j.xgen.2024.100700
39. Sun BB, Chiou J, Traylor M, et al. Plasma proteomic associations with genetics and health in the UK Biobank. *Nature*. 2023;622(7982):329-338. doi:10.1038/s41586-023-06592-6
40. Jansen IE, Savage JE, Watanabe K, et al. Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nat Genet*. 2019;51(3):404-413. doi:10.1038/s41588-018-0311-9
41. Chen T, Pan F, Huang Q, et al. Metabolic phenotyping reveals an emerging role of ammonia abnormality in Alzheimer's disease. *Nat Commun*. 2024;15(1):3796. doi:10.1038/s41467-024-47897-y

42. Liu Y, Thalamuthu A, Mather KA, et al. Plasma lipidome is dysregulated in Alzheimer's disease and is associated with disease risk genes. *Transl Psychiatry*. 2021;11(1):344. doi:10.1038/s41398-021-01362-2
43. Calzada E, Onguka O, Claypool SM. Phosphatidylethanolamine metabolism in health and disease. *Int Rev Cell Mol Biol*. 2016;321:29-88. doi:10.1016/bs.ircmb.2015.10.001
44. Blennow K, Galasko D, Perneczky R, et al. The potential clinical value of plasma biomarkers in Alzheimer's disease. *Alzheimers Dement*. 2023;19(12):5805-5816. doi:10.1002/alz.13455
45. Chen Y, Liu S, Ren Z, et al. Cross-ancestry analysis of brain QTLs enhances interpretation of schizophrenia genome-wide association studies. *Am J Hum Genet*. 2024;111(11):2444-2457. doi:10.1016/j.ajhg.2024.09.001
46. Saitou M, Dahl A, Wang Q, Liu X. Allele frequency impacts the cross-ancestry portability of gene expression prediction in lymphoblastoid cell lines. *Am J Hum Genet*. 2024;111(12):2814-2825. doi:10.1016/j.ajhg.2024.10.009
47. Kunkle BW, Schmidt M, Klein HU, et al. Novel Alzheimer disease risk loci and pathways in African American individuals using the African genome resources panel: a meta-analysis. *JAMA Neurol*. 2021;78(1):102-113. doi:10.1001/jamaneurol.2020.3536
48. Sherva R, Zhang R, Sahelijo N, et al. African ancestry GWAS of dementia in a large military cohort identifies significant risk loci. *Mol Psychiatry*. Mar 2023;28(3):1293-1302. doi:10.1038/s41380-022-01890-3
49. Zhao P, Xu Y, Jiang LL, et al. LILRB2-mediated TREM2 signaling inhibition suppresses microglia functions. *Mol Neurodegener*. 2022;17(1):44. doi:10.1186/s13024-022-00550-y
50. Hou J, Chen Y, Cai Z, et al. Antibody-mediated targeting of human microglial leukocyte Ig-like receptor B4 attenuates amyloid pathology in a mouse model. *Sci Transl Med*. 2024;16(741):eadj9052. doi:10.1126/scitranslmed.adj9052
51. Feng T, Mai S, Roscoe JM, et al. Loss of TMEM106B and PGRN leads to severe lysosomal abnormalities and neurodegeneration in mice. *EMBO Rep*. 2020;21(10):e50219. doi:10.15252/embr.202050219
52. Werner G, Damme M, Schludi M, et al. Loss of TMEM106B potentiates lysosomal and FTLD-like pathology in progranulin-deficient mice. *EMBO Rep*. 2020;21(10):e50241. doi:10.15252/embr.202050241
53. Zhou X, Brooks M, Jiang P, et al. Loss of Tmem106b exacerbates FTLD pathologies and causes motor deficits in progranulin-deficient mice. *EMBO Rep*. 2020;21(10):e50197. doi:10.15252/embr.202050197
54. Clayton EL, Isaacs AM. Progranulin and TMEM106B: when two become wan. *EMBO Rep*. 2020;21(10):e51668. doi:10.15252/embr.202051668
55. Ferkingstad E, Sulem P, Atlason BA, et al. Large-scale integration of the plasma proteome with genetics and disease. *Nat Genet*. 2021;53(12):1712-1721. doi:10.1038/s41588-021-00978-w
56. Pietzner M, Wheeler E, Carrasco-Zanini J, et al. Mapping the proteo-genomic convergence of human diseases. *Science*. 2021;374(6569):eabj1541. doi:10.1126/science.abj1541

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Yang C, Timsina J, Liu M, et al. Analyses of plasma multi-omic data across ancestries identify novel pathways implicated in Alzheimer's disease. *Alzheimer's Dement*. 2026;22:00-00e71164. https://doi.org/10.1002/alz.71164