OXFORD

Human Molecular Genetics, 2023, Vol. 32, 15, 2532–2543
https://doi.org/10.1093/hmg/ddad087
Advance access publication date 19 May 2023
Association Studies Article

Genetic architecture of plasma Alzheimer disease biomarkers

Joseph Bradley[1,2,3], Priyanka Gorijala[1,2], Suzanne E. Schindler[3,4], Yun J. Sung[1,2,3], Beau Ances[3,4] the Alzheimer's Disease Neuroimaging Initiative, the Human Connectome Project, Maria V. Fernandez[1,2] and Carlos Cruchaga[1,2,3,*]

[1]Department of Psychiatry, Washington University School of Medicine, St. Louis, MO 63110, USA
[2]NeuroGenomics and Informatics Center, Washington University School of Medicine, St. Louis, MO 63110, USA
[3]Hope Center for Neurologic Diseases, Washington University in St. Louis, St. Louis, MO 63110, USA
[4]Department of Neurology, Washington University School of Medicine, St. Louis, MO 63110, USA

*To whom correspondence should be addressed at: 4444 Forest Park Ave, Washington University School of Medicine, St. Louis, MO 63110, USA. Tel: +1 3142860546; Fax: +1 3143622244; Email: cruchagac@wustl.edu

Abstract

Genome-wide association studies (GWAS) of cerebrospinal fluid (CSF) Alzheimer's Disease (AD) biomarker levels have identified novel genes implicated in disease risk, onset and progression. However, lumbar punctures have limited availability and may be perceived as invasive. Blood collection is readily available and well accepted, but it is not clear whether plasma biomarkers will be informative for genetic studies. Here we perform genetic analyses on concentrations of plasma amyloid-β peptides Aβ40 ($n = 1{,}467$) and Aβ42 ($n = 1{,}484$), Aβ42/40 ($n = 1467$) total tau ($n = 504$), tau phosphorylated (p-tau181; $n = 1079$) and neurofilament light (NfL; $n = 2{,}058$). GWAS and gene-based analysis was used to identify single variant and genes associated with plasma levels. Finally, polygenic risk score and summary statistics were used to investigate overlapping genetic architecture between plasma biomarkers, CSF biomarkers and AD risk. We found a total of six genome-wide significant signals. *APOE* was associated with plasma Aβ42, Aβ42/40, tau, p-tau181 and NfL. We proposed 10 candidate functional genes on the basis of 12 single nucleotide polymorphism-biomarker pairs and brain differential gene expression analysis. We found a significant genetic overlap between CSF and plasma biomarkers. We also demonstrate that it is possible to improve the specificity and sensitivity of these biomarkers, when genetic variants regulating protein levels are included in the model. This current study using plasma biomarker levels as quantitative traits can be critical to identification of novel genes that impact AD and more accurate interpretation of plasma biomarker levels.

Introduction

As the field of Alzheimer's Disease (AD) genome-wide association studies (GWAS) advances, a seemingly ever-increasing number of loci have been identified. Most recently 75 loci were reported by Bellenguez *et al.* (1) in a large GWAS analyses ($n = 788{,}989$). Although increasing sample size and newer tools for analysis have greatly improved our ability to identify common and rare variants associated with AD, there is still much work to be done in understanding the biological consequences of these genetic underpinnings. One way of addressing this knowledge gap in GWAS analyses is to use AD-associated fluid biomarkers as a proxy for case-control status (2,3), which can offer increased statistical power (2–4) as well as insights into how disease affects the various tissues for measuring biomarker levels.

Previous analyses have identified multiple loci associated with risk, onset or progression by using with concentrations of Aβ42, tau, p-tau181, clusterin (*CLU*), apolipoprotein E (*APOE*) and soluble *TREM2* (*sTREM2*) (4–12). Our group also leveraged these endophenotypes, disease associated biomarkers with a demonstrated connection to genetics, to understand pathologic events. We used cerebrospinal fluid (CSF) p-tau181 and Aβ concentrations to demonstrate that *APOE* also affects tau pathology independently of AD (2) that was further validated using animal models (13). Some notable findings were that CSF *sTREM2* levels are associated with AD and that CSF *sTREM2* levels correlate with CSF tau and p-tau181, but not Aβ42, suggesting a post amyloid accumulation relationship between CSF *sTREM2* and AD (14). In addition, Mendelian Randomization analyses demonstrated that *TREM2* affects AD risk in most individuals, not just in carriers of *TREM2* variants, and that higher *TREM2* levels are protective. Additional genetic analysis of CSF *TREM2* levels also found that the *MS4A* gene cluster modulates CSF *sTREM2* and implicates *TREM2* as having a key biological role in sporadic late-onset AD.

Recently, additional studies have investigated if plasma biomarkers are also informative endophenotypes for genetic studies as in the case of the CSF biomarkers. Damotte *et al.* (15), performed meta-analyses on over 12,000 non-demented individuals to find variants associated with Aβ40, Aβ42 and Aβ42/40 ratio. Their analysis found variants in the *BACE1*, *APP*, *PSEN2* and *APOE* loci to be associated with plasma Aβ levels, suggesting that plasma Aβ levels are a potential endophenotype for AD. Additionally, a recent study by Sarnowski *et al.* (16) ($n = 14{,}721$) identified 14 novel loci and 10 genes, which had been previously associated with neurological diseases, associated with circulating total tau. Specifically, they find strong associations with the *MAPT* locus for Europeans, suggesting that variants in the *MAPT* locus are the primary risk factor in tau-mediated

Received: December 23, 2022. **Revised:** April 29, 2023. **Accepted:** May 16, 2023

Table 1. Summary of demographics for each analyte per cohort

	N	ADNI	Knight-ADRC	HCP	Age	Range	% female	% *APOE4* +	% controls
Aβ40	1964	683	1121	157	71.55 ± 9.46	37–91	50.00%	42.92%	57.35%
Aβ42	1985	704	1121	157	71.56 ± 9.43	37–91	50.05%	43.04%	57.05%
Aβ42/40	1991	704	1124	160	71.59 ± 9.45	37–91	50.13%	43.04%	57.05%
NfL	2380	1577	803	0	72.28 ± 8.08	37–91	45.75%	44.08%	45.21%
tau	563	563	0	0	75.41 ± 6.79	55–91	41.92%	48.49%	32.68%
p-tau181	1186	1186	0	0	73.25 ± 7.07	55–91	45.70%	43.68%	36.26%
Cases	1457	1166	291	NA	73.78 ± 7.58	50–91	42.76%	56.49%	NA
Controls	1355	525	830	NA	69.94 ± 9.11	37–91	53.73%	32.77%	NA
Total	3010	1695	1162	157	72.03 ± 8.79	37–91	47.77%	43.32%	44.95%

This project uses plasma Aβ40, Aβ42, NfL, tau and p-tau181 from 1162 unique Knight-ADRC subjects, 1695 unique ADNI subjects and 157 unique HCP subjects. Demographics, which include number of subjects who has data for each analyte (N), Age, Age range, % female subjects, % of subjects with *APOE* ε4 and % of controls in data set are shown here.

neurodegeneration in that ethnic group. Finally, A separate study of p-tau181 levels (11) ($n = 714$) identified novel relationships between p-tau181 and *APOE* risk allele rs769449. These findings demonstrate the power of using AD endophenotypes to address unanswered questions about the underlying biology of AD. These studies were performed in a large dataset but with previous generations of Aβ and tau assays and not the newer assays generation that provide higher sensitivity and perform better to predict AD risk. Recent studies have found that recent liquid chromatography/mass-spectrometry-based (LC–MS) methods of measuring Aβ, such as the methods used by Kirmess *et al*. have shown an ability to associate status with a decrease in Aβ42 and Aβ42/40 ratio with a similar accuracy to CSF biomarkers, as well as they have better concordance with amyloid positron emission tomography (PET) compared with immunoassays (17). In this study, we will focus on plasma AD biomarkers generated using the new generation of assays.

Although CSF endophenotypes have been demonstrated to be a powerful analysis tool, they are more difficult to obtain because of invasiveness and costs of the procedure; thus, opening a need for accessible, less invasive assays like plasma biomarkers. Recently, sensitive and precise LC–MS methods for measuring plasma Aβ42/Aβ40 (Aβ_{1-40}/Aβ_{1-42}) have opened the door for novel studies and clinical application of AD plasma biomarkers (18–20). Additional immunoassays such as SiMoA have been adapted for increased detection of other AD plasma biomarkers (19): including tau, p-tau181 and neurofilament light (NfL). These biomarkers' epidemiological relevance to AD makes analyses of them especially useful to understanding how genetic markers are contributing to AD outcomes. Biomarkers such as tau and NfL could be especially useful in this case, as they are relevant to the etiology of several neurological diseases such as frontal-temporal dementia (18) and multiple sclerosis (20), respectively. There are many overlapping features of neurodegenerative diseases. Identifying overlapping genetic markers can improve our ability to understand the causal genes and functional pathways implicated disease, or, importantly, allow for the identification of new drug targets. These six biomarkers are additionally important as they have shown power to differentiate AD cases and controls with more consistency than some other AD biomarkers (21). This study aims to identify genetic variants associated with AD plasma concentrations of Aβ40, Aβ42, Aβ42/40, NfL, tau, and p-tau181; as well as use this relationship to gain a better understanding of how genetics influence the relationship between AD and these plasma biomarkers. Additionally, we hypothesize that it may be possible to improve the accuracy of these plasma biomarkers by incorporating genetic information on the association model, as in our previous study for *YKL40* levels in CSF (7).

Results

Study design

We performed GWAS of six plasma biomarkers plasma (Aβ40, Aβ42, Aβ42/40, NfL, tau and p-tau181) measured in the Alzheimer's Disease Neuroimaging initiative (ADNI), the Knight-ADRC and the Human Connectome Project (HCP) cohorts. Analyses were performed using plasma concentrations of Aβ40 ($n = 1{,}964$), Aβ42 ($n = 1{,}985$), total tau ($n = 563$), p-tau181 ($n = 1{,}186$) and NfL ($n = 2{,}380$) in unrelated individuals of European ancestry. Within ADNI, individuals had an average age of 74, with the total population ranging from 55 to 91. Knight-ADRC participants were 69 years old on average, ranging between 37 and 91. Finally, HCP participants were 74 years old on average ranging between 51 and 90 years old, for a total population average of 72 (37–91). Across the three datasets, subjects were, on average, 47.77% female, 43.32% APOE ε4 + and 43.32% controls. The full population demographics are summarized in Table 1 and Supplementary Material, Table S1. Plasma biomarker levels were measured with either C2N LC–MS-based methods (Aβ40, Aβ42 and NfL), Luminex immunoassay-based methods, (Aβ40, and Aβ42) or SiMoA immunoassay-based methods (tau, p-tau181, NfL). More information about each cohort and quality control (QC) steps applied are available in the methods section and in Figure 1, Table 1, and Supplementary Material, Table S1.

We performed the following steps for this study (Fig. 1): (i) Baseline plasma biomarker levels were normalized in each cohort separately by log10 transformation and calculating a z-score. Outliers, defined as z-score values greater than three standard deviations above or z-score values less than three standard deviations below the mean, were removed (Supplementary Material, Figs S1–S3). (ii) The data were cleaned by excluding individuals without genotype data, excluding related and duplicated individuals, and excluding individuals not of European ancestry (Supplementary Material, Figs S4 and S5). (iii) Linear regression analyses were run using Plink v2.0 using an additive model of single nucleotide polymorphism (SNP) dosages with covariates age, sex, CohortArray and the first 10 principal components. (iv) Post-GWAS analyses included comparisons across CSF biomarker and AD risk phenotypes, annotation, gene based and fine mapping and using plasma biomarker levels corrected by its top hit to determine that variants genetic contribution to AD.

Single variant analysis

Single variant analysis (SVA) of our five biomarkers identified six loci passing the genome-wide significance threshold (5×10^{-08}) and eight SNP-biomarker pairs (Table 2). Forest plots (Supplementary Material, Figs S7–S19) showed consistent direction

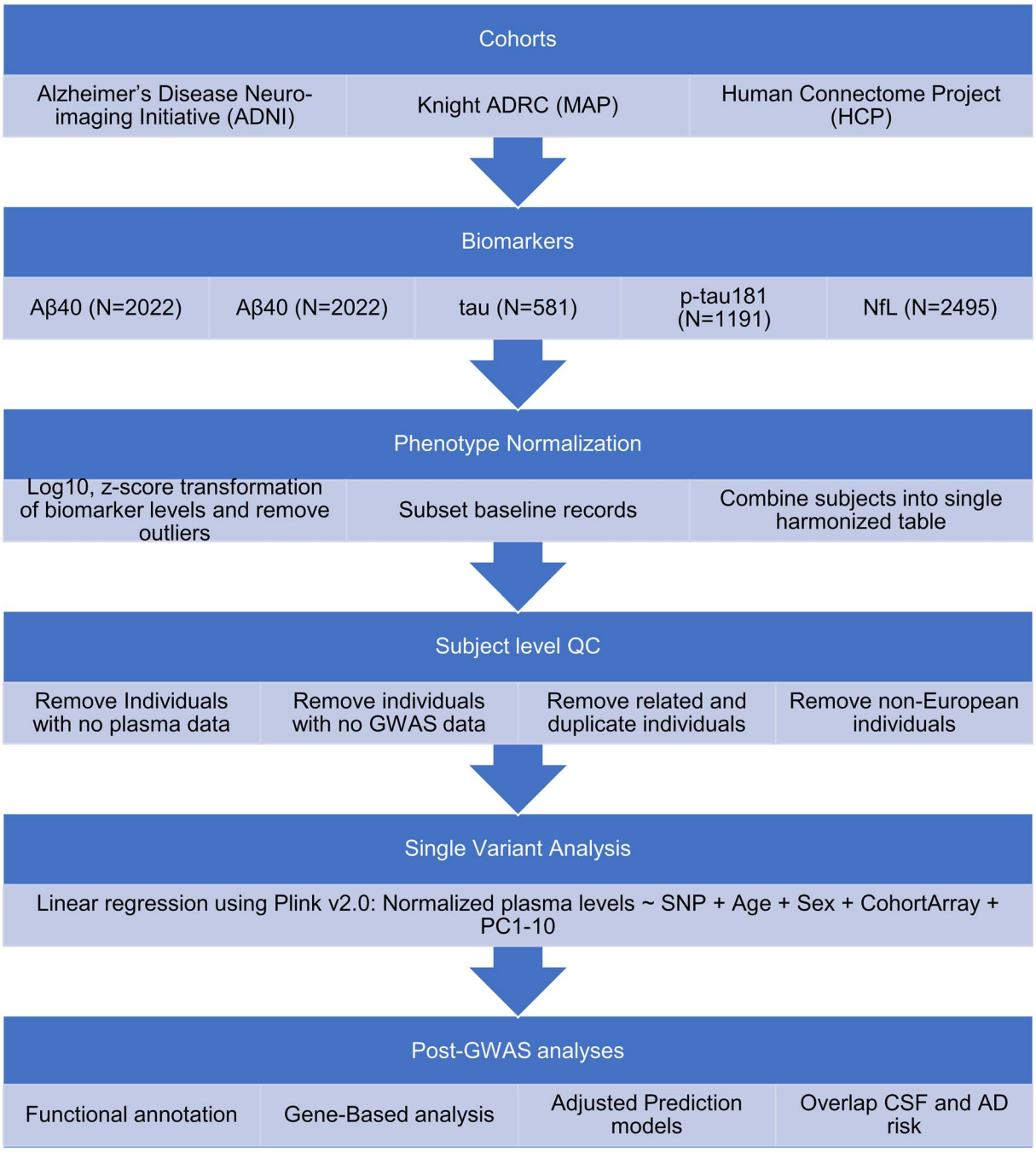

Figure 1. Study Design. The project follows the figure above top-to-bottom, left-to-right order. We normalize phenotype by log10, z-score transformation, subset baseline records and harmonize the three cohort tables. Next, we perform subject-level QC by removing individuals with no plasma or GWAS data, are related/duplicates, and are not of European ancestry. We perform single variant analysis using Plink v2.0. Finally, we perform post-GWAS analyses including functional annotation, gene-set analysis, modified prediction models and overlap with other phenotypes.

of effect for the sentinel SNPs across datasets. The *APOE* $\varepsilon 4$ risk variant rs429358 minor allele frequency (MAF = 22.4%) was the most strongly associated signal for Aβ42 ($P = 2.84 \times 10^{-08}$), Aβ42/40 ($P = 6.26 \times 10^{-18}$), p-tau181 ($P = 3.91 \times 10^{-22}$) and NfL ($P = 7.63 \times 10^{-10}$; Fig. 2 and supplementary figures).

There were three additional loci associated with plasma NFL and all of them were rare variants (MAF < 0.01; rs76576243, rs62476358 and rs149377334, Table 2 and Fig. 2E, Supplementary Material, Figs S15–17). rs76576243 ($P = 2.09 \times 10^{-08}$; MAF = 0.7%) on chromosome 20 is an intergenic variant ~290 kb away from *SULF2* and 420 kb away from *NCOA3*. We performed functional annotation, and we did not find any coding variant, eQTL or pQTL in this locus. The next most significant hit for NfL, rs62476358 ($P = 2.91 \times 10^{-08}$, MAF = 0.3%) on chromosome 7, is an intronic variant of the gene *PTPRN2* which is the only gene in the region. Lastly, rs149377334 ($P = 3.24 \times 10^{-08}$, MAF = 0.5%), another intergenic variant on chromosome 8, is 79 kb away from its nearest gene *UNC5D*. No other genes are within the 500 kb flanks of the variant.

Two additional significant loci were associated with tau levels (Table 2 and Fig. 2C, Supplementary Material, Figs S9 and S10): rs1795325153 and rs7790638 both of which are also rare variants. Rs1795325153 ($P = 1.05 \times 10^{-08}$; MAF = 0.4%), is an intergenic variant near the *CHRM2* gene, whereas rs7790638 ($P = 4.41 \times 10^{-08}$; MAF = 0.6%) is an intronic variant of *UMAD1*. Functional annotation (coding variants, and eQTL and pQTL) did not find

Table 2. Summary of top hits and overlap with CSF p-tau181, tau and AD risk

Biomarker	rsID	Nearest Gene	MAF	BETA	Plasma_P	CSF Aβ42_P	CSF p-tau181_P	AD_Risk_P
Aβ40	rs113183385	*RP11-292B1.2*	0.025	−0.599	5.65×10^{-08}	0.788	**0.033**	6.88×10^{-01}
Tau	rs7790638	*UMAD1*	0.006	−2.776	4.41×10^{-08}	NA	NA	6.35×10^{-01}
p-tau181	rs554323670	*ETV1*	0.009	−1.159	8.26×10^{-08}	0.806	0.534	1.63×10^{-01}
Tau	rs1795325153	*CHRM2*	0.004	−2.823	1.05×10^{-08}	NA	NA	NA
NfL	rs62476358	*PTPRN2*	0.003	1.122	2.91×10^{-08}	0.284	0.457	8.86×10^{-01}
NfL	rs149377334	*UNC5D*	0.005	0.93	3.24×10^{-08}	0.249	0.605	4.47×10^{-01}
p-tau181	rs189447682	*KCNB2*	0.004	−1.328	8.02×10^{-08}	0.947	0.376	2.95×10^{-01}
NfL	rs10809277	PTPRD	0.46	0.132	8.37×10^{-08}	0.124	0.943	4.53×10^{-01}
Aβ42	rs6857	*APOE*	0.251	−0.211	2.84×10^{-08}	$\mathbf{6.29 \times 10^{-143}}$	**0.028**	$\mathbf{4.00 \times 10^{-123}}$
NfL	rs429358	*APOE*	0.224	0.184	7.63×10^{-10}	$\mathbf{4.09 \times 10^{-299}}$	$\mathbf{2.93 \times 10^{-77}}$	$<\mathbf{10^{-300}}$
Aβ42/40 ratio	rs429358	*APOE*	0.213	−0.363	6.26×10^{-18}	$\mathbf{4.09 \times 10^{-299}}$	$\mathbf{2.93 \times 10^{-77}}$	$<\mathbf{10^{-300}}$
p-tau181	rs56131196	*APOE*	0.277	0.39	3.97×10^{-22}	$\mathbf{1.69 \times 10^{-321}}$	$\mathbf{9.94 \times 10^{-94}}$	$\mathbf{4.00 \times 10^{-123}}$
NfL	rs76576243	*SULF2*	0.007	1.07	2.09×10^{-08}	0.551	NA	$\mathbf{3.79 \times 10^{-02}}$

Top hits from each plasma analyte were cross referenced with CSF Aβ42, p-tau181 and tau results from previous studies and AD risk results from Bellenguez *et al.* to look for associations in common. Only results in *APOE* region replicate suggesting our dataset may be underpowered. Bolded values represent a nominally significant *P*-value ($P \leq 0.05$).

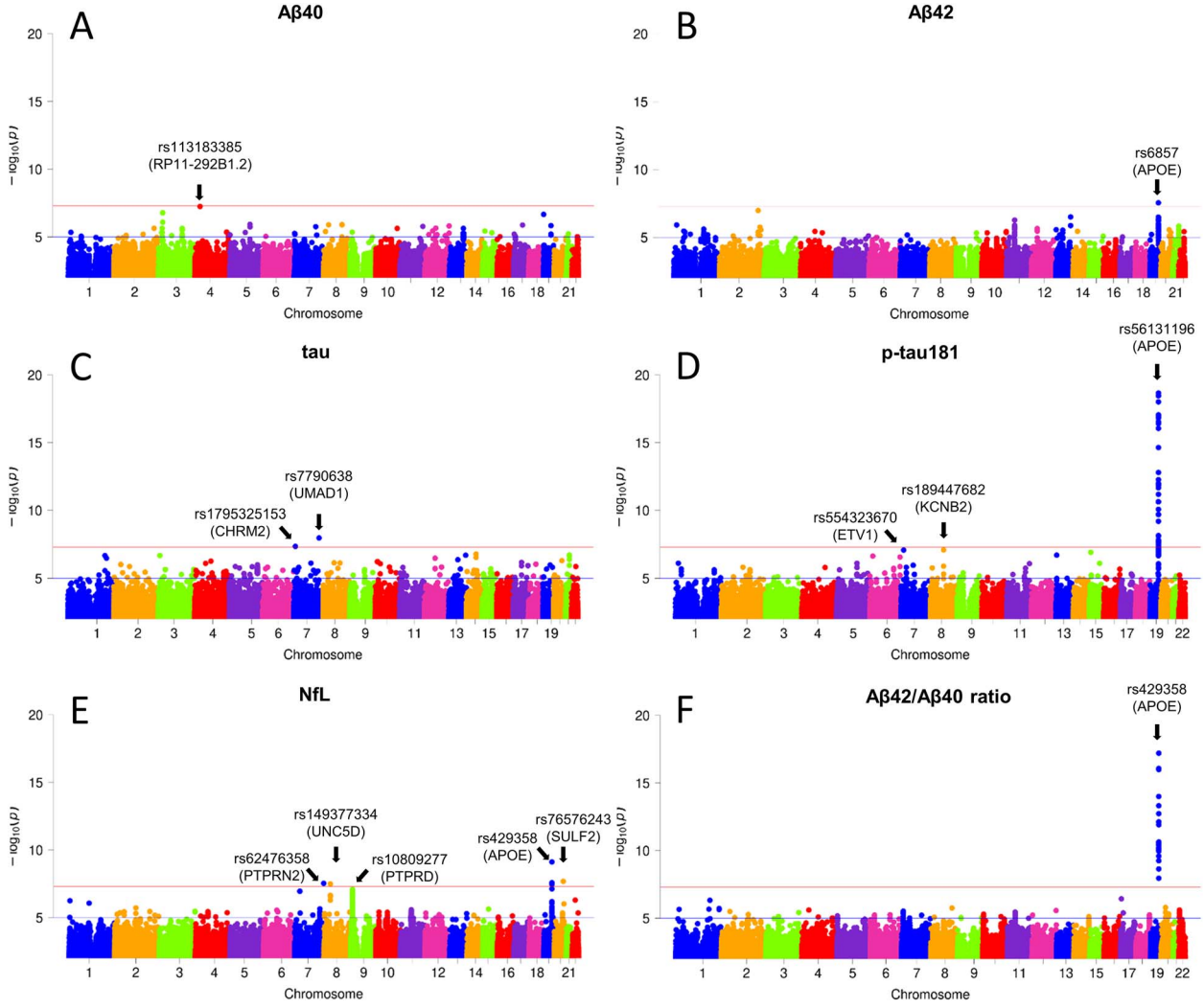

Figure 2. Manhattan plots from SVA. Manhattan plots of SVA results for Aβ40 (**A**), Aβ42 (**B**), tau (**C**), p-tau181 (**D**) and NfL (**E**). The red line represents the genome-wide significance threshold of 5×10^{-08} and the blue line represents the suggestive threshold of 1×10^{-05}.

any association in these loci. A search of candidate genes within the AMP-AD database, Agora, revealed that candidate genes *APOE* (chr19 locus for Aβ42, p-tau181, NfL and Aβ42/40), *CHRM2* (chr7 locus for tau), *ETV1* (chr7 locus for p-tau181), *KCNB2* (chr8 locus for p-tau181), *PTPRD* (chr9 locus for NfL), *PTPRN2* (chr7 locus for NfL), *SULF2* (chr29 locus for NfL), *UMAD1* (chr7 locus for tau), and *UNC5D* (chr8 locus for NfL) were all differentially expressed in at least one brain region in AD cases compared with cognitively normal controls. A summary of these results can be found in the Supplementary Material, Table S2.

We also identified four loci of interest (LOI) at near genome-wide significance ($P < 1 \times 10^{-07}$), which requires larger studies for validation. The rs113183385 SNP ($P = 5.65 \times 10^{-08}$; MAF = 2.5%), an intergenic variant on chromosome 4 not near any major genes, was associated with plasma Aβ40 levels (Table 2 and Fig. 2A and Supplementary Material, Fig. S6). For plasma p-tau181 levels, there were two LOI which are both intergenic variants: rs189447682 ($P = 8.02 \times 10^{-08}$; MAF = 0.4%) near *KCNB2* on chromosome 8 and rs55432670 ($P = 8.26 \times 10^{-08}$; MAF = 0.9%) near *ETV1* on chromosome 7 (Table 2 and Fig. 2D and Supplementary Material, Fig. S12–S13). For plasma NfL levels, a LOI was observed at a common intergenic variant, rs10809277 ($P = 8.37 \times 10^{-08}$; MAF = 46%), ~400 kb away from *PTPRD* on chromosome 9 (Table 2 and Fig. 2E and Supplementary Material, S18). All LOI (except rs10809277) were rare variants (or in gene poor regions in the case of rs113183385) and no eQTLs were identified in connection with these variants, or variants in LD ($D' = 1$, $P \leq 0.01$). None of these were in any previously reported AD risk loci; however, the nearest gene for the two p-tau181 loci, *KCNB2 and ETV1*, and the NfL locus, *PTPRD*, have been investigated with respect to AD.

Overlap genetic architecture of plasma biomarkers with other AD phenotypes

Next, we sought to determine if the variants associated with plasma AD biomarkers were associated with CSF biomarkers or AD risk. In comparing plasma and CSF biomarker analyses (Table 2) we identified two loci in common between the two tissues. The strongest association for both markers was found at the *APOE* locus.

We first compared the statistical power of CSF versus plasma by looking at the strength of the association of a known region; APOE, with Aβ and p-tau181. This study of plasma biomarkers has a lower sample size ($N_{plasma\_A\beta42} = 1{,}485$, $N_{plasma\_ptau181} = 1{,}078$) than CSF genetics studies ($N_{CSF\_A\beta42} = 13{,}116$, $N_{CSF\_ptau181} = 12{,}553$), making it difficult to compare the power of the two tissues. To compare the power of plasma GWAS more accurately to CSF, we can also look at the earlier analysis of CSF Aβ42, tau and p-tau181 by Cruchaga *et al.* (2), which included 1,269 individuals. We compared the *P*-values of the top hit in the *APOE* locus (rs769449) reported in the main text of Cruchaga *et al.* (2) with *P*-values from our analysis of Aβ42 and p-tau181. We used the proxy variant rs41377151 in place of the *APOE2* variant rs769449 as it was not present in the analysis of Aβ42. The signal in plasma Aβ42 was less significant than that for CSF (2) ($P_{plasma} = 4.85 \times 10^{-07}$, $P_{CSF} = 9.02 \times 10^{-47}$). On the other hand, the strength of the signal in the *APOE* locus for plasma p-tau181 had a bit more significant association compared with CSF ($P_{plasma} = 2.15 \times 10^{-19}$, $P_{CSF} = 2.56 \times 10^{-18}$). This juxtaposition between Aβ42 and p-tau181 may suggest that plasma p-tau181 may be a better endophenotype for genetic studies than Aβ42. Aβ40 and Aβ42 on their own are known to be not as good biomarkers as their ratio (Aβ42/Aβ40). We found that the association of *APOE* with plasma Aβ42/Aβ40 had a more significant association than Aβ42 alone and similar to that of p-tau181 ($P_{A\beta42/A\beta40} = 6.23 \times 10^{-18}$).

Then, we compared the genome-wide signals between CSF and plasma versus AD risk. We find a similar overlap between plasma versus risk and CSF versus risk when comparing our results with the latest AD risk paper by Bellenguez *et al.* (1), Variants in the *APOE* locus have a common association with plasma phenotypes and AD risk (Table 2). To analyze the overlap between AD risk with plasma biomarkers, we first combined the summary statistics from Bellenguez *et al.* 2022 and Kunkle *et al.* 2019 (21) and performed LD clumping within a 2 MB window. In Kunkle *et al.* there were 57 signals, and 83 in the Bellenguez *et al.* paper. A total of 41 of 83 signals from Bellenguez overlap with Kunkle *et al.* and 42 were novel leading to a total of 99 loci. Out of these 99 SNPs, we found 23 SNPs across 10 loci that have at least a nominally significant association with at least one of our plasma phenotypes (Supplementary Material, Table S3). When comparing plasma GWAS results to results from Jansen *et al.* (22) we find that the SNPs in the chr1 (*CR1*) locus for CSF Aβ42 also shown a nominal significance in plasma Aβ40 and Tau. The SNPs in the *GMNC* and *C16orf195* loci associated with CSF levels also have nominal significance in plasma Aβ42 and NfL and Tau, respectively. The *APOE* locus from CSF Aβ42 and p-tau181 is also associated with plasma Aβ40 and Tau and shows a genome-wide significance in plasma Aβ42, Aβ42/40 ratio, NfL and p-tau181 (Supplementary Material, Table S4, Table S5). Looking in the other direction, only the *APOE* locus from plasma analyses shows a genome-wide significance in Jansen *et al.* whereas rs113183385 ($P = 0.033$) and rs6857 ($P = 0.028$) replicate at nominal significance in CSF p-tau181 (Table 2). These results suggest that plasma biomarkers, in a larger dataset could identify previously confirmed AD risk loci.

To expand on the genetic overlap between AD risk, CSF biomarkers and this study, we performed linear regression of plasma phenotypes as a response variable to AD risk polygenic risk score (PRS) and CSF PRS with and without *APOE*. We found that the AD PRS with *APOE* to be significantly associated with Aβ42 ($P_{\text{with } APOE} = 6.08 \times 10^{-05}$), tau ($P_{\text{with } APOE} = 1.85 \times 10^{-02}$), p-tau181 ($P_{\text{with } APOE} = 7.21 \times 10^{-14}$, NFL ($P_{\text{with } APOE} = 9.79 \times 10^{-04}$, and, A$\beta$42/40 ($P_{\text{with } APOE} = 2.16 \times 10^{-07}$; Supplementary Material, Table S6) but not Aβ40 (Supplementary Material, Table S6). On the other hand, PRS CSF Aβ42 ($P = 9.46 \times 10^{-09}$) and p-tau181 ($P = 2.17 \times 10^{-06}$) were significantly associated with their plasma counterparts, but not tau ($P = 1.52 \times 10^{-01}$). However, CSF PRS without *APOE* was not significantly associated with any plasma biomarkers, suggesting that there is little genetic overlap between CSF and plasma beyond *APOE*. However, analyses by Deming *et al.* 2017 (23) only found a handful of loci outside of the *APOE* locus, so a small number of SNPs are being used to generate risk scores when *APOE* is not included; therefore, the risk scores may not be reliable. The results using AD risk PRS demonstrate that AD risk variants are associated with plasma biomarker levels—specifically levels of plasma Aβ42, Aβ42/40, tau, p-tau181 and NfL. The weak association ultimately echoes the previous point, that more statistical power is needed in plasma biomarker datasets to identify AD risk loci.

Improving biomarker accuracy by including genetics variant

As several loci associated with plasma biomarkers were not associated with AD risk, besides *APOE*, we hypothesize that by correcting the plasma levels with those loci, the sensitivity and specificity of these biomarkers could increase. We also hypothesize that if we

Table 3. Summary of SNP-adjusted ROC

Analyte	SNP	rsID	AUC	SNP adjusted AUC	Delta AUC	P
p-tau181	chr19:44919589	rs56131196	0.636	0.610	−2.60%	7.18×10^{-04}
Aβ42	chr19:44888997	rs6857	0.536	0.511	−2.50%	1.11×10^{-08}
Tau	chr7:7824214	rs7790638	0.570	0.545	−2.50%	4.64×10^{-04}
Tau	chr7:136835925	rs1795325153	0.570	0.546	−2.40%	3.22×10^{-04}
Aβ40	chr4:28949717	rs113183385	0.546	0.529	−1.70%	4.84×10^{-03}
Aβ42/40 ratio	Chr19:44908684	rs429358	0.600	0.591	−0.90%	3.60×10^{-02}
p-tau181	chr8:72434155	rs189447682	0.636	0.641	0.50%	3.05×10^{-01}
p-tau181	chr7:14036424	rs554323670	0.636	0.643	0.70%	1.95×10^{-01}
NFL	chr19:44908684	rs429358	0.618	0.625	0.70%	3.74×10^{-01}
NFL	chr7:158022052	rs62476358	0.618	0.640	2.20%	2.73×10^{-03}
NFL	chr8:35156150	rs149377334	0.618	0.640	2.20%	2.83×10^{-03}
NFL	chr20:48076998	rs76576243	0.618	0.640	2.20%	2.42×10^{-03}
NFL	chr9:11008173	rs10809277	0.618	0.641	2.30%	2.08×10^{-03}

To help determine if significant variants are truly associated with AD, we compare ROC curves of a base model—Status ~ normalized protein level—versus an alternative model which uses the studentized residuals of normalized protein levels ~ SNP in place of the normalized protein level (Status ~ studentized residuals). An increased AUC corresponds to genetic noise; whereas, a decreased AUC corresponds to the variant having a true association with AD. P-value determined by DeLong test for comparing AUC (H_0 = difference between AUC is 0).

correct by a known AD-risk variant, like *APOE*, the sensitivity and specificity of these biomarkers would decrease.

We were able to validate this hypothesis and establish a baseline by using *APOE* locus, which is a known AD locus as a reference. We corrected the protein levels by including *APOE* in the model. When doing this, we found that the difference in the area under the receiver operating characteristics curve (ΔAUC), between the model including only the protein levels and that correcting by *APOE* to be −2.6%, −2.5%, −0.90% and 0.7% for p-tau181, Aβ42, Aβ42/40 and NfL, respectively. This indicates that the sensitivity and specificity of these proteins, when correcting for *APOE*, decreases, as we hypothesized. However, we also hypothesize that if we corrected by variants not associated with AD risk, the sensitivity and specificity would improve. At the same time, these analyses suggest that we should only expect to see very modest changes, lower than 3% in the AUC from testing any of the variants associated with plasma levels.

When correcting with the variants found in this study, i.e. rs7790638 (tau), rs1795325153 (tau), rs7790638 (tau) and rs113183385 (Aβ40), we found that the AUC were significantly different from the unadjusted model including protein levels (Fig. 3 and Table 3). The ΔAUC for both tau-associated variants were 2.4% and 2.5% (rs7790638 and rs1795325153, respectively) indicating nearly as strong of an effect on tau as *APOE ε4* has on Aβ40 and p-tau181. Similarly, when the variants associated with NFL outside of the *APOE* region (rs76576243, rs62476358, rs149377334 and rs10809277) were used to correct NFL levels, the AUC showed a significant increase compared with the default model (ΔAUC of 2.2–2.3% and $P < 2.83 \times 10^{-03}$, respectively).

Discussion

We aimed to identify novel loci associated with AD by using plasma biomarkers as endophenotypes and to aid a better interpretation of plasma biomarker levels. To that end, it is important to perform these analyses and compare with CSF biomarkers and AD risk to determine how informative plasma biomarkers are for genetic analysis. To the former, we detected associations with the *APOE* locus across Aβ42, Aβ42/40, tau, p-tau181 and NFL. We were additionally able to identify several rare variants associated with those four biomarkers. Fine mapping of these variants identified one potentially novel gene, *KCNB2*, although it is not clear if this variant is truly associated with AD. That said, we believe that our ability to detect at least *APOE* suggests that these plasma biomarkers are informative and that they can identify variants associated with AD risk. It is important to note; however, that larger studies and/or are needed to identify additional signals and validate the rare variants findings from this study.

With respect to the interpretation of plasma biomarkers GWAS, our comparison to CSF shows that, currently, CSF still provides clearer and stronger results for some biomarkers like Aβ42. However, for a similar sample size, we found that plasma p-tau181 had comparable power as CSF. In combination with the relative ease of attaining plasma from more individuals, these findings would suggest plasma to be the better for analysis regarding p-tau181 levels. Additional analyses need to be performed to determine if this may also be true for other p-tau species.

Additionally, our analysis of overlap with AD risk using the latest GWAS shows that the genetic architecture of plasma biomarkers overlaps with that of AD risk. A total of 23 out of 99 sentinel SNPs identified by Bellenguez *et al.* and Kunkle *et al.* were found to have a nominally significant *P*-value in association with at least one plasma biomarker (Supplementary Material, Table S3). We also found AD risk PRS to be significantly associated with plasma biomarkers Aβ42, Aβ42/40, p-tau181 and NfL when the *APOE* region is included (Supplementary Material, Table S6). These findings reinforce the conclusion that plasma biomarkers can be used to detect AD risk associations, but also suggest a mechanism of AD biology driving the original association. For example, rs679515 in the *CR1* locus, identified by Bellenguez *et al.* with *P*-value 1.40×10^{-23} is associated with nominal significance to Aβ40 (4.09×10^{-02}) and tau181 (3.25×10^{-02}; Supplementary Material, Table S3). This suggests that the functional gene in that locus is mainly affecting tau and amyloid biology with respect to AD. The results from the PRS can be used to build upon this. On the basis of the correlations between plasma biomarker levels and AD risk PRS without *APOE*, it seems that AD risk loci outside of *APOE* are primarily associated with and affecting plasma p-tau181 and NfL. Because the correlation with plasma Aβ42, Aβ42/40 and tau levels is attenuated when removing the *APOE* region, the data suggests that Aβ42, Aβ42/40 and tau have an *APOE* driven relationship to AD. That is, AD-associated loci outside of *APOE* do not seem to be strongly contributing to plasma Aβ42, Aβ42/40 and tau levels. Alternatively, we must also consider that we simply do not have the resolution necessary to detect significant correlations in Aβ.

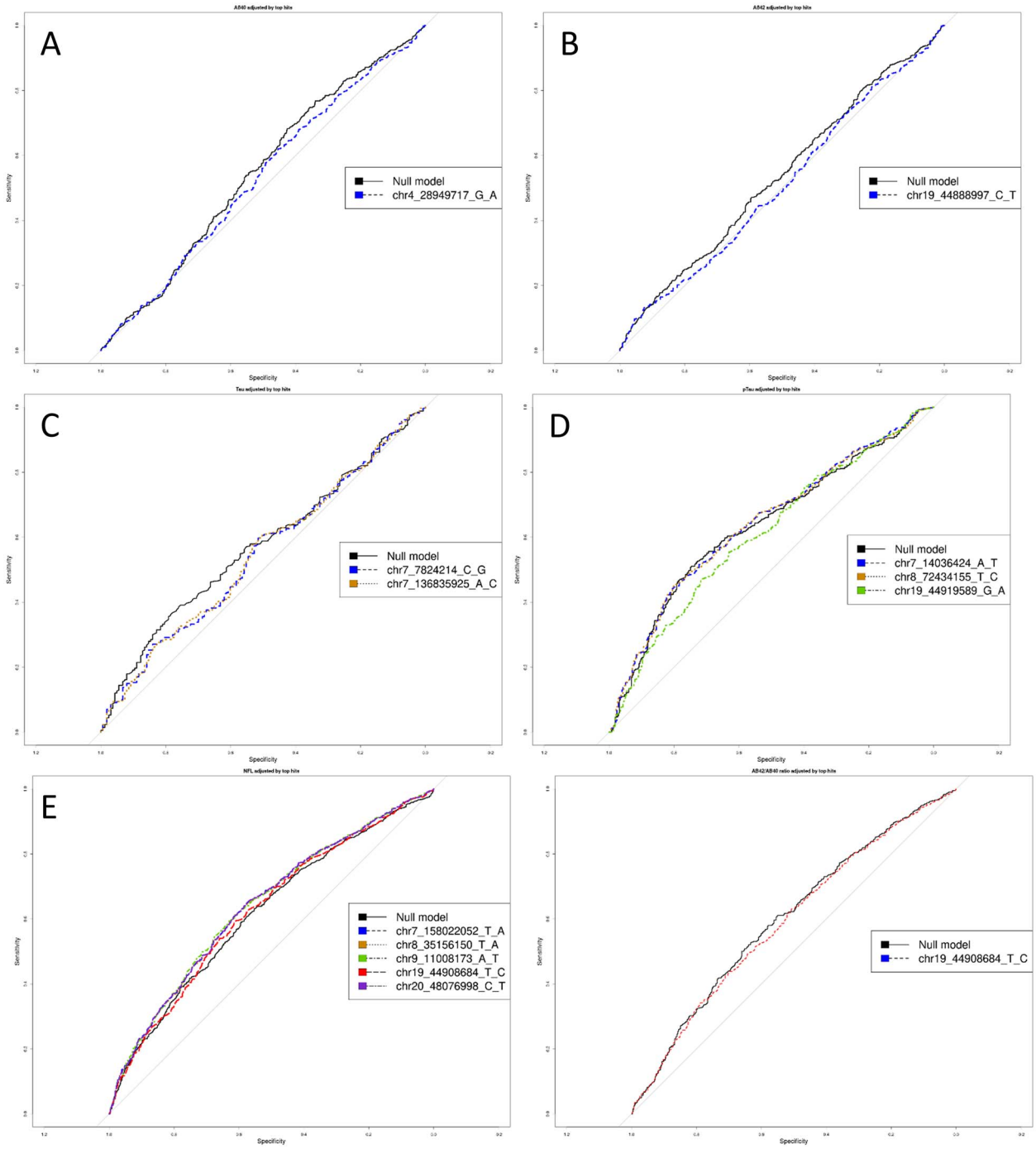

Figure 3. ROC plots from SNP-adjusted prediction models. ROC plots for Aβ40 (**A**), Aβ42 (**B**), tau (**C**), p-tau181 (**D**) and NfL (**E**). The default model in each plot is indicated by a solid black line. Each SNP adjusted model is represented by a dashed line and color, shown in the legend to the right of each plot. AUCs and *P*-values are summarized in Table 3.

To add to this, and incorporate the context of other AD literature, *APOE* (associated with plasma Aβ42, p-tau181 and NfL) and *UMAD1* (associated with plasma tau on chr7)—candidate genes identified through fine mapping—have been previously reported to be associated with AD risk (1). The variant driving the association in *APOE* locus is the known *APOE* ε4 allele rs429358 (24). Five of the other GWAS loci from this analysis, *CHRM2* (chr7, tau), *ETV1*/ER81 (chr7, p-tau181), *PTPRD* (chr9, NfL), *PTPRN2* (chr7, NfL) and *SULF2* (chr20, NfL), have also been associated with Alzheimer's disease in at least one study. One study has investigated links between *CHRM2* (chr7, tau) and AD and found that rs6962027 in *CHRM2* and rs7511970 in *CHRM3* may interact to affect AD risk (25). A study by Pastorcic and Das (26) that investigated *ETS* transcription factors ER81 (*ETV1* chr7, p-tau181) and Elk1 showed that both regulate transcription of *PSEN1* gene promoter. *PTPRN2* (chr7, NfL), which was identified on the basis of whole-genome sequencing of *APOE* ε4 carriers (27), has not yet been studied functionally in connection to Alzheimer's, but it may function similarly to *PTPRD* (chr9, NfL), which 'has likely roles as a neuronal cell adhesion molecule' and is associated with neurofibrillary tangle density (28). With regards to *SULF2* (chr20, NfL), Roberts *et al.* (29) found that it has decreased expression in

specific regions of the brain in AD cases compared with controls. As such, these results also emphasize the need for increased sample size in plasma biomarker studies with power to detect more risk loci.

Neither *UNC5D* (chr8, NfL) nor *KCNB2* (chr8, p-tau181), have been previously associated with AD by any studies. This is combination with our fine mapping suggests that these are novel associations to plasma NfL and p-tau181 levels, respectively, and potentially AD. In addition to this, AMP-AD results show that *UNC5D* is significantly differentially expressed in five out of nine brain tissues assayed, whereas *KNCB2* is significantly differentially expressed in seven out of nine brain tissues assayed. For four of the seven tissues (temporal cortex, superior temporal gyrus, para-hippocampal gyrus and the inferior frontal gyrus) wherein *KCNB2* is significantly differentially expressed, the log2 of the fold change is less than −0.263 or >0.263, which is meaningful change by AMP-AD definition (Supplementary Material, Table S2). This also suggests that p-tau181 levels in cases are affecting expression changes in *KCNB2* or vice versa.

When comparing plasma and CSF biomarkers, we found a very strong overlap in the *APOE* locus with our biomarker GWAS results. In this locus, the most significant SNP from Cruchaga *et al.* (2), rs34404554 associated with CSF p-tau181 had a *P*-value of 1.33×10^{-16} as compared with $P = 2.34 \times 10^{-11}$ in our analysis of plasma p-tau181. Outside of the *APOE* locus, we do not see an overlap of the CSF genome-wide hits, except for one variant near *GLIS3*, associated with CSF p-tau181, which has a nominally significant *P*-value ($P < 0.05$) in our plasma dataset (rs623295, $P = 0.034$). We, similarly, only found overlap in the *APOE* locus when comparing results of Aβ42 by Cruchaga *et al.* (6) with our current analysis, albeit with much lower *P*-values. We also came to a similar conclusion comparing our plasma biomarkers to CSF Aβ42, tau and p-tau181 by Deming *et al.* PRS (23) and CSF Aβ42 and p-tau181 summary statistics in the most recent large-scale CSF biomarker GWAS by Jansen *et al.* (22). Using CSF biomarker PRS we only found associations with plasma biomarkers when the *APOE* region was included. This indicates that there is relatively low genetic overlap outside of this locus. However, it should be taken into consideration that the low number of SNPs outside the *APOE* locus at genome-wide significance could mean the results of the PRS without *APOE* for CSF biomarkers are unreliable. Additionally, the lack of genome-wide significant replication outside of the *APOE* locus implies a need for larger sample sizes for strong associations; however, that we find previously identified AD risk and CSF biomarker loci that replicate at nominal significance encourages us that this is possible. Taken together, these findings suggest that plasma p-tau181 could be as informative as CSF p-tau181 for genetic analyses with a similar sample size. It should be noted; however, that the analysis by Cruchaga *et al.* (2) has partial sample overlap to this analysis so that could explain why there is similar power for p-tau181.

To compare our results with more recent studies of plasma p-tau181, we can look to GWAS of plasma p-tau181 levels in ADNI individuals by Huang *et al.* (11). Their analysis used a much smaller sample size ($n = 714$) and potentially up to 60% sample overlap with our analysis. Comparing the two, we again confirm the *APOE* locus as being associated with plasma p-tau181 levels, but our study provides larger statistical power: The top rs4420638 and had a *P*-value of 6.26×10^{-08}, in Huang *et al.* versus $P = 1.83 \times 10^{-21}$ in our analysis. This more directly illustrates the point that increasing sample size can improve our ability to detect signals in association analyses of plasma p-tau181 levels. It also highlights the ability of plasma p-tau181 assay to find consistent results across at least partially independent datasets. It is important to note; however, the limitations of plasma p-tau181 to identify loci outside of the *APOE* region thus far. Additionally, simply increasing sample size might not be a sufficient catchall solution to gaining utility out of plasma biomarkers. These limitations could be because of the quality and consistency of plasma p-tau immunoassays needing some improvement. A recent publication comparing 10 assays for measuring p-tau181 and p-tau217 (30) found that mass spectrometry-based measurements of p-tau217 consistently outperformed immunoassays of either p-tau217 or p-tau181. Therefore, it is important that consistent platforms are used across studies to be able to combine data from different studies, which will lead to larger statistical power.

Our analysis of plasma tau did not find any genome-wide significant results in common with Sarnowski *et al.* (16) analysis of circulating total Tau, which included 14,721 samples. However, there are still important similarities between the two worth considering. Firstly, even in our analyses it did not reach the genome-wide significant, we were able to confirm the association with the *MAPT* locus, which was the second genome-wide locus and in analyses showed a $P = 6.7 \times 10^{-07}$. We additionally found that rs4607127, one of the lead variants in their trans-ancestry GWAS, replicated at nominal *P*-value ($P = 0.012$) in our dataset. The latter finding is unexpected given that it was only significant in their analysis of African Americans, but it highlights the importance of not only conducting large GWASs, but also doing analyses across ancestries, as they highlighted in their discussion. We also see some slight overlap in our MAGMA gene-based results and their analysis of GWAS catalog traits for the main genes identified in their meta-analysis. Specifically, their GWAS meta-analysis implicates *IL15*, which is associated with neurofibrillary tangles in GWAS catalog results, as a candidate functional gene for circulating total tau levels. The same gene reached nominal significance ($P = 0.049$) in our MAGMA gene-based analysis, supporting the indication of *IL15* may have a tau mediated functional role in AD.

We do find genome-wide significance level replication with results from Damotte *et al.* (15), but only in the *APOE* region. None of our analyses of Aβ levels finds significant results for the rs650585 variant identified by Damotte *et al.* (15), nor for the *BACE1/RNF214* locus as a whole. Damotte's group also did gene-based analysis using their main GWAS results and found genome-wide significant results for *PSEN2*, *CCK*, *RNF214*, *BACE1* and *APP* in association with Aβ40; *APOE* and *APOC1* in association with Aβ42; and *ZNF397*, *APOE* and *APOC1* in association with Aβ42/40 ratio. We did not have power to generate significant gene-based analysis for Aβ40, but there was a strong signal in the *APP* locus in the main GWAS ($P = 5.88 \times 10^{-06}$). Gene-based analyses for Aβ42, and Aβ42/40 ratio successfully replicated signals for *APOE* and *APOC1*. It should be noted that although our analysis is nearly nine times smaller than the dataset used by Damotte *et al.* the analyses performed here used the new generation of plasma assays for these proteins. Although our dataset has both cases and controls, which offers us more statistical power compared with their model using only non-demented subjects, the sample size required to generate results in their analysis and weak results in the *APOE* locus suggests that much larger studies with these more sensitive analyses are needed to make any definite determination on the matter.

Finally, we attempted to improve the sensitivity and specificity of our plasma biomarkers by including our most significant variant in our model. Our group has used similar methods previously in an analysis of CSF *YKL-40* levels (7) where we demonstrate that when correcting *YKL40* levels with the variants associated

with *YKL40* levels but not AD risk, the sensitivity and specificity increases. In our current analysis, we did not see any improved sensitivity and specificity for plasma p-tau181, tau or Aβ42 levels when including out top hits in the model, likely because of the low frequency of the identified variants, or because those variants may have some residual effect on AD risk. On the other hand, we show a significant and consistent increase in the sensitivity and specificity power when NfL levels were corrected by the genome-wide signals. This is an important finding, as it demonstrates that we can improve the sensitivity and specificity of these biomarkers by including variants associated with the plasma levels but not the disease.

It is important to note that these results have some limitations. Firstly, as has been indicated in other sections of the discussion, our power to detect signals outside of the *APOE* region for both common and rare variants is lacking. Future plasma biomarker studies would require larger sample sizes or biomarker assays with greater sensitivity, such as the p-tau217 assay MS assay that is mentioned by Janelidze *et al.* (30). Following this point, some significant loci and LOI identified in our main GWAS are very close to, and in some cases below the target minor allele count of 10 for our given sample size. As such, some very signals identified here may be false positives. In addition to that, we do not have strong results for determining if these variants are truly associated with AD. More studies will need to be done to validate that these variants influence both AD and biomarker levels.

In conclusion, this analysis demonstrates potential power in using plasma biomarker levels as endophenotypes. Taking into consideration the genetic overlap of plasma and CSF biomarker, and the context of recent plasma biomarker studies, these findings suggest that plasma biomarkers can identify variants associated with AD risk and can do so with some level of consistency. Plasma is a much easier, cheaper and less invasive biospecimen to use for clinical assessment and genetic analyses and will be a useful tool for further expanding upon our understanding of AD biology and pathophysiology.

Materials and Methods

Data and demographics

Plasma Aβ40 ($n = 1964$), Aβ42 ($n = 1985$), Tau ($n = 563$), p-tau181 ($n = 1186$), and NFL ($n = 2380$) were measured for a total of 3010 individuals (in three different studies; ADNI ($n = 1695$), Knight-ADRC ($n = 1162$), and the HCP ($n = 157$). Within ADNI, individuals had an average age of 74, with the total population ranging from 55 to 91. Knight-ADRC participants were 69 years old on average, ranging between 37 and 91. Finally, HCP participants were 74 years old on average ranging between 51 and 90 years old, for a total population average of 72 (37–91). Across the three datasets, there subjects were, on average, 47.77% female, 43.32% *APOE* $\varepsilon 4+$ and 43.32% controls. The full population demographics are summarized in Table 1 and Supplementary Material, S1.

Knight-ADRC

Knight-ADRC samples were collected in the morning after an overnight fast, immediately centrifuged and stored at −80°C. Measures of plasma Aβ40, Aβ42 and NfL were performed by C2N diagnostics (17), which uses Mass spectrometry-based tests to measure concentrations of protein in plasma. Plasma Aβ40, Aβ42 and NFL levels were measured in 1121 (74.04% cases), 1121 (74.04% cases) and 803 (75.22% cases) Knight-ADRC individuals respectively ($n = 1162$; Supplementary Material, Table S1).

CSF and blood samples were collected following overnight fasting. Plasma Aβ42 and Aβ40 were measured in the C2N Diagnostics commercial laboratory with an immunoprecipitation-mass spectrometry assay (St. Louis, MO, USA). Plasma NfL was measured with Quanterix Nf-Light assay kits on a HD-X analyzer. Concentrations of CSF Aβ40, Aβ42, total tau (tau) and tau phosphorylated at 181 (p-tau181) were measured by chemiluminescent enzyme immunoassay using a fully automated platform (LUMIPULSE G1200, Fujirebio, Malvern, PA, USA). CSF NfL was measured via commercial ELISA kit (UMAN Diagnostics, Umeå, Sweden).

The human connectome project

Aβ40 and Aβ42 for human connectome subjects was measured by C2N diagnostics by the same methods as previously described. Plasma Aβ40 and Aβ42 were measured for 157 HCP individuals (Supplementary Material, Table S1). Case control status was unknown for these individuals.

ADNI

Plasma Aβ40 and Aβ42 for ADNI subjects was measured using Luminex immunoassay platform and was performed on 2454 plasma samples collected from 733 ADNI subjects. Alternatively, measurement of plasma p-tau181 tau, and NfL were done using the Single Molecule array (SiMoA) technique. After removing outlier and missing data from ADNI plasma Aβ40, Aβ42, Tau, p-tau181 and NFL was measures, there were 683 (29.58% Cases), 704 (29.11% cases), 1577 (32.72% cases), 563 (32.86% cases) and 1186 (36.42% cases) individuals for each analyte, respectively ($n = 1659$; Supplementary Material, Table S1).

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public–private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging, PET, other biological markers and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment and early Alzheimer's disease (AD). For up-to-date information, see www.adni-info.org.

Plasma data harmonization

Because of the variation in measurement devices and methods across datasets, measures of each analyte were normalized by a z-scored, log10 transformation, as previously reported (23,31). Outliers, defined as three standard deviations above or below the mean, were removed. Data normality before and after normalization is checked simply visually by histogram to see that the data follow a normal distribution centered at 0. No significant differences were found in biomarker levels across cohorts after normalization (Supplementary Material, Fig. S1).

Genotyping QC

Genetic data were available for a total of 2692 unique individuals. Genotyping data come from several different rounds of genotyping on Illumina platforms. Stringent quality thresholds were applied to the genotype data for each platform separately. SNPs were kept if they met the following criteria: (i) had a genotyping rate ≥98%; (ii) had a MAF ≥0.3% and (iii) were in Hardy–Weinberg equilibrium (HWE) ($P > =1 \times 10^{-6}$). After removing low quality SNPs and individuals, genotype imputation was performed using the Impute2 program with haplotypes derived from the 1000 Genomes Project (released June 2012). Genotype imputation was

performed separately on the basis of the genotype platform used. SNPs were removed if they failed any of the following criteria: (i) an impute2 info-score quality of <0.3; (ii) a MAF < 2% or (iii) out of HWE. After Imputation and QC, the different imputed plink files were merged. A total of nearly 14 million (13 791 029) imputed and directly genotyped SNPs and 2388 individuals were used for final analyses.

To determine relatedness, Z0 and Z1 from IBD analysis for all individuals are plotted (Supplementary Material, Fig. S2). Individuals which fall outside of the selected range ($Z0 \geq 0.65$, $Z1 \leq 0.4$) are considered relatives or duplicates. A single individual from each relative/duplicate pair with lowest call rate was removed. Finally, this analysis only uses data from subjects of a European genetic background. Genetic background for all individuals is determined by plotting the first two principal component analyses and identifying the European cluster. Individuals whose first two principal components fall within the selected range (Supplementary Material, Fig. S3) are carried forward in QC steps. A summary of subjects removed at each QC step can be found in Supplementary Material, Table S7. Only baseline measures of each dataset are used for SVA. The final analysis dataset, after QC steps consists of 2388 total individuals (54.20% cases; Supplementary Material, Table S1).

Statistical analyses

SVA of normalized plasma analyte levels were performed using PLINK v2.0. The following covariates were included in the analysis; age, sex, PC1-PC10 and CohortArray—A categorical dummy variable of the combined Cohort and Array for each subject—with the largest group removed. It is suggested within the Plink2.0 documentation to address collinearity by having multiple categorical covariates. Variants are considered statistically significant if they have an SVA unadjusted *P*-value $\leq 5 \times 10^{-08}$. Some signals which are nearly significant ($P < 1 \times 10^{-07}$) with strong LD trails are also considered for post-GWAS analyses. After SVA, forest plots were generated to look at the direction of effect of top SNPs across each phenotype. Following this, significant hits were further explored using regional association plots [generated by locuszoom (32)] to interrogate genes underneath each analytes top hit. Additional gene mapping and functional annotation was done using FUnctional Mapping and Annotation (FUMA) software (33). Candidate genes indicated from fine mapping were interrogated for differential expression in AD cases versus controls by looking up genes in Agora, a publicly available database of genes and whether they are associated with AD. Agora and the data within are made available by studies performed through the Accelerating Medicines Partnership Program for Alzheimer's Disease (AMP-AD).

Post-GWAS

Much of the post-GWAS analysis and processing was done in R with the following packages loaded; qqman (34), dplyr, pROC (35), MASS (36), snpStats, data.table, ggplot2. Additional information on how each was used will be described further in relevant sections.

Functional annotation and Gene-based analyses were done using FUMA. Because FUMA currently only works with hg19 based results, our summary statistics were lifted over to hg19 from hg38. The final tables were generated in R. Additional post-GWAS analyses were done to investigate the relationship between the results found in this study, the results of the latest AD risk GWAS (1), and the previous GWASs using CSF endophenotypes from our group.

Quantifying the overlap with AD risk and CSF analytes Aβ, tau and p-tau181

Top SNPs from this study were cross-checked in AD risk (1) and CSF GWAS (22) summary statistics to look for common associations. The reverse was also done. To evaluate the genetic overlap of AD risk variants and plasma biomarkers, and CSF biomarker risk variants and plasma biomarkers, PRS were calculated using AD risk GWAS (1) and CSF biomarker GWAS (23) summary statistics with and without variants in the *APOE* region as reported previously (37). Using the PRS as a response variable in linear regression against plasma analytes with and without *APOE* allows us to interpret how the combined effect of AD risk or CSF biomarker SNPs contribute to our plasma phenotypes both with and without *APOE* background.

Improving biomarkers performance by correcting by genetic associations

Finally, we wanted to try improving the sensitivity and specificity of our plasma AD biomarkers for case control status. To do this we compare two models using receiver operating characteristic (ROC) curve analysis; a model testing sensitivity and specificity of normalized plasma biomarker levels (status ~ normalized protein level), and a model testing sensitivity and specificity of normalized plasma biomarker levels after they are adjusted for genetic effects [status ~ residuals (normalized protein level ~ SNP dosage)]. SNP dosages were extracted from plink.bed files using the read.plink function from the snpStats package. To correct for genetic effect, we calculated Studentized residuals—using the 'studres' function from the MASS package—from a linear regression of plasma biomarker levels and each of the top hits for each phenotype. As was done in the plink analysis, age, gender, CohortArray and the first 10 PCs were used as covariates. We summarize the results by reporting the change in AUC (ΔAUC)—calculated by taking the difference in the uncorrected AUC (using z-score) and the corrected AUC (using studentized residuals)—and the *P*-value—calculated by the DeLong test to compare AUC.

Supplementary Material

Supplementary Material is available at HMG online.

Acknowledgements

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.

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) and DOD ADNI). 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.; Cogstate; 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 Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH and NINDS. The data used for the analyses described in this manuscript were obtained from: V8 the GTEx Portal as part of FUMA analyses on 05/13/22.

The results published here are in whole or in part on the basis of the data obtained from the AD Knowledge Portal (https://adknowledgeportal.org). The AD Knowledge Portal is a platform for accessing data, analyses and tools generated by the Accelerating Medicines Partnership (AMP-AD) Target Discovery Program and other National Institute on Aging (NIA)-supported programs to enable open-science practices and accelerate translational learning. We thank the participants of the Religious Order Study and Memory and Aging projects for the generous donation, the Sun Health Research Institute Brain and Body Donation Program, the Mayo Clinic Brain Bank, and the Mount Sinai/JJ Peters VA Medical Center NIH Brain and Tissue Repository. Data and analysis contributing investigators include Nilüfer Ertekin-Taner, Steven Younkin (Mayo Clinic, Jacksonville, FL), Todd Golde (University of Florida), Nathan Price (Institute for Systems Biology), David Bennett, Christopher Gaiteri (Rush University), Philip De Jager (Columbia University), Bin Zhang, Eric Schadt, Michelle Ehrlich, Vahram Haroutunian, Sam Gandy (Icahn School of Medicine at Mount Sinai), Koichi Iijima (National Center for Geriatrics and Gerontology, Japan), Scott Noggle (New York Stem Cell Foundation), Lara Mangravite (Sage Bionetworks).

Conflict of Interest statement. CC has received research support from: GSK and EISAI. 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. CC is a member of the advisory board of Vivid Genomics and Circular Genomics and owns stocks. C2N Diagnostics was co-founded by Drs. Randall Bateman and David Holtzman, who are faculty members at Washington University. The PrecivityAD test was developed in the laboratory of Dr. Randall Bateman at Washington University and licensed to C2N Diagnostics. Washington University has a financial interest in the PrecivityAD test.

Funding

This work was supported by grants from the National Institutes of Health, R01AG044546 (CC), RO1AG070941 (S.E. Schindler), P01AG003991 (CC, JCM), RF1AG053303 (CC), RF1AG058501 (CC), U01AG058922 (CC), NIH R44 AG059489 (C2N Diagnostics), Bright Focus (CA2016636), The Gerald and Henrietta Rauenhorst Foundation, the Alzheimer's Drug Discovery Foundation (GC-201711-2013978) and the Chan Zuckerberg Initiative (CZI), and the Alzheimer's Association Zenith Fellows Award (ZEN-22-848604, awarded to CC). The recruitment and clinical characterization of research participants at Washington University were supported by NIH P30AG066444 (JCM), P01AG03991 (JCM) and P01AG026276 (JCM).

Data availability

Proteomic and genetic data from the Knight ADRC participants are available at the NIAGADS and can be accessed at https://www.niagads.org/KnightADRC-collection.

References

1. Bellenguez, C., Küçükali, F., Jansen, I.E., Kleineidam, L., Moreno-Grau, S., Amin, N., Naj, A.C., Campos-Martin, R., Grenier-Boley, B., Andrade, V. *et al.* (2022) New insights into the genetic etiology of Alzheimer's disease and related dementias. *Nat. Genet.*, **54**, 412–436.
2. Cruchaga, C., Kauwe, J.S.K., Harari, O., Jin, S.C., Cai, Y., Karch, C.M., Benitez, B.A., Jeng, A.T., Skorupa, T., Carrell, D. *et al.* (2013) GWAS of cerebrospinal fluid tau levels identifies risk variants for alzheimer's disease. *Neur.*, **78**, 256–268.
3. Cruchaga, C., Kauwe, J.S.K., Nowotny, P., Bales, K., Pickering, E.H., Mayo, K., Bertelsen, S., Hinrichs, A., Fagan, A.M., Holtzman, D.M. *et al.* (2012) Cerebrospinal fluid *APOE* levels: an endophenotype for genetic studies for Alzheimer's disease. *Hum. Mol. Genet.*, **21**, 4558–4571.
4. Kauwe, J.S.K., Cruchaga, C., Karch, C.M., Sadler, B., Lee, M., Mayo, K., Latu, W., Su'a, M., Fagan, A.M., Holtzman, D.M. *et al.* (2011) Fine mapping of genetic variants in BIN1, CLU, CR1 and PICALM for association with cerebrospinal fluid biomarkers for Alzheimer's disease. *PLoS One*, **6**, e0015918.
5. Cruchaga, C., Ebbert, M.T.W. and Kauwe, J.S.K. (2014) Genetic discoveries in AD using CSF amyloid and tau. *Curr. Genet. Med. Rep.*, **2**, 23–29.
6. Cruchaga, C., Kauwe, J.S.K., Mayo, K., Spiegel, N., Bertelsen, S., Nowotny, P., Shah, A.R., Abraham, R., Hollingworth, P., Harold, D. *et al.* (2010) SNPs associated with cerebrospinal fluid Phospho-tau levels influence rate of decline in alzheimer's disease. *PLoS Genet.*, **6**, e31039.
7. Deming, Y., Black, K., Carrell, D., Cai, Y., Del-Aguila, J.L., Fernandez, M.V., Budde, J., Ma, S.M., Saef, B., Howells, B. *et al.* (2016) Chitinase-3-like 1 protein (CHI3L1) locus influences cerebrospinal fluid levels of YKL-40. *BMC Neurol.*, **16**, 217.
8. Cruchaga, C., Chakraverty, S., Mayo, K., Vallania, F.L.M., Mitra, R.D., Faber, K., Williamson, J., Bird, T., Diaz-Arrastia, R., Foroud, T.M. *et al.* (2012) Rare variants in APP, PSEN1 and PSEN2 increase risk for AD in late-onset Alzheimer's disease families. *PLoS One*, **7**, e31039.
9. Piccio, L., Deming, Y., Del-Águila, J.L., Ghezzi, L., Holtzman, D.M., Fagan, A.M., Fenoglio, C., Galimberti, D., Borroni, B. and Cruchaga, C. (2016) Cerebrospinal fluid soluble TREM2 is higher in Alzheimer disease and associated with mutation status. *Acta Neuropathol.*, **131**, 925–933.
10. Blennow, K., Hampel, H., Weiner, M. and Zetterberg, H. (2010) (2010) cerebrospinal fluid and plasma biomarkers in Alzheimer disease. Cerebrospinal fluid and plasma biomarkers in Alzheimer disease. *Nat. Rev. Neurol.*, **6**, 131–144.
11. Huang, Y.-Y., Yang, Y.-X., Wang, H.-F., Shen, X.-N., Tan, L. and Yu, J.-T. (2022) Genome-wide association study identifies *APOE* locus influencing plasma p-tau181 levels. *J. Hum. Genet.*, **67**, 459–463.
12. Deming, Y., Xia, J., Cai, Y., Lord, J., Holmans, P., Bertelsen, S., Holtzman, D., Morris, J.C., Bales, K., Pickering, E.H. *et al.* (2016) A potential endophenotype for Alzheimer's disease: cerebrospinal fluid clusterin. *Neurobiol. Aging*, **37**, 208.e1–208.e9.

13. Shi, Y., Yamada, K., Liddelow, S.A., Smith, S.T., Zhao, L., Luo, W., Tsai, R.M., Spina, S., Grinberg, L.T., Rojas, J.C. *et al.* (2017) *APOE4* markedly exacerbates tau-mediated neurodegeneration in a mouse model of tauopathy. *Nat.*, **549**, 523–527.
14. Deming, Y., Filipello, F., Cignarella, F., Cantoni, C., Hsu, S., Mikesell, R., Li, Z., Del-Aguila, J.L., Dube, U., Farias, F.G. *et al.* (2019) The MS4A gene cluster is a key modulator of soluble TREM2 and Alzheimer's disease risk. *Sci. Transl. Med.*, **11**, eaau2291.
15. Damotte, V., van der Lee, S.J., Chouraki, V., Grenier-Boley, B., Simino, J., Adams, H., Tosto, G., White, C., Terzikhan, N., Cruchaga, C. *et al.* (2021) Plasma amyloid β levels are driven by genetic variants near *APOE*, BACE1, APP, PSEN2: a genome-wide association study in over 12,000 non-demented participants. *Alz. and Dem.*, **17**, 1663–1674.
16. Sarnowski, C., Ghanbari, M., Bis, J.C., Logue, M., Fornage, M., Mishra, A., Ahmad, S., Beiser, A.S., Boerwinkle, E., Bouteloup, V. *et al.* (2022) Meta-analysis of genome-wide association studies identifies ancestry-specific associations underlying circulating total tau levels. *Commun. Biol.*, **5**, 336.
17. Kirmess, K.M., Meyer, M.R., Holubasch, M.S., Knapik, S.S., Hu, Y., Jackson, E.N., Harpstrite, S.E., Verghese, P.B., West, T., Fogelman, I. *et al.* (2021) The PrecivityAD™ test: accurate and reliable LC-MS/MS assays for quantifying plasma amyloid beta 40 and 42 and apolipoprotein E proteotype for the assessment of brain amyloidosis. *Clin. Chim. Acta*, **519**, 267–275.
18. Rissin, D.M., Kan, C.W., Campbell, T.G., Howes, S.C., Fournier, D.R., Song, L., Piech, T., Patel, P.P., Chang, L., Rivnak, A.J. *et al.* (2010) Single-molecule enzyme-linked immunosorbent assay detects serum proteins at subfemtomolar concentrations. *Nat. Biotechnol.*, **28**, 595–599.
19. Aschenbrenner, A.J., Li, Y., Henson, R.L., Volluz, K., Hassenstab, J., Verghese, P., West, T., Meyer, M.R., Kirmess, K.M., Fagan, A.M. *et al.* (2022) Comparison of plasma and CSF biomarkers in predicting cognitive decline. *Ann. Clin. Transl. Neurol.*, **9**, 1739–1751.
20. Olsson, B., Lautner, R., Andreasson, U., Öhrfelt, A., Portelius, E., Bjerke, M., Hölttä, M., Rosén, C., Olsson, C., Strobel, G. *et al.* (2016) CSF and blood biomarkers for the diagnosis of Alzheimer's disease: a systematic review and meta-analysis. *Lancet Neurol.*, **15**, 673–684.
21. Kunkle, B.W., Grenier-Boley, B., Sims, R., Bis, J.C., Damotte, V., Naj, A.C., Boland, A., Vronskaya, M., van der Lee, S.J., Amlie-Wolf, A. *et al.* (2019) 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.
22. Jansen, I.E., van der Lee, S.J., Gomez-Fonseca, D., de Rojas, I., Dalmasso, M.C., Grenier-Boley, B., Zettergren, A., Mishra, A., Ali, M., Andrade, V. *et al.* (2022) Genome-wide meta-analysis for Alzheimer's disease cerebrospinal fluid biomarkers. *Acta Neuropathol.*, **144**, 821–842.
23. Deming, Y., Li, Z., Kapoor, M., Harari, O., Del-Aguila, J.L., Black, K., Carrell, D., Cai, Y., Fernandez, M.V., Budde, J. *et al.* (2017) Genome-wide association study identifies four novel loci associated with Alzheimer's endophenotypes and disease modifiers. *Acta Neuropathol.*, **133**, 839–856.
24. Corder, E.H., Saunders, A.M., Strittmatter, W.J., Schmechel, D.E., Gaskell, P.C., Small, G.W., Roses, A.D., Haines, J.L. and Pericak-Vance, M.A. (1979) (1993) gene dose of Apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Sci.*, **261**, 921–923.
25. Ya Chee, L. and Cumming, A. (2017) Polymorphisms in the cholinergic receptors muscarinic (CHRM2 and CHRM3) genes and Alzheimer's disease. *Avicenna J. Med. Biotechnol.*, **10**, 196–199.
26. Pastorcic, M. and Das, H.K. (2003) Ets transcription factors ER81 and Elk1 regulate the transcription of the human presenilin 1 gene promoter. *Mol. Brain Res.*, **113**, 57–66.
27. Park, J.H., Park, I., Youm, E.M., Lee, S., Park, J.H., Lee, J., Lee, D.Y., Byun, M.S., Lee, J.H., Yi, D. *et al.* (2021) Novel Alzheimer's disease risk variants identified based on whole-genome sequencing of *APOE* ε4 carriers. *Transl. Psychiatry*, **11**, 296.
28. Uhl, G.R. and Martinez, M.J. (2019) PTPRD: neurobiology, genetics, and initial pharmacology of a pleiotropic contributor to brain phenotypes. *Ann. N. Y. Acad. Sci.*, **1451**, 112–129.
29. Roberts, R.O., Kang, Y.N., Hu, C., Moser, C.D., Wang, S., Moore, M.J., Graham, R.P., Lai, J.-P., Petersen, R.C. and Roberts, L.R. (2017) Decreased expression of sulfatase 2 in the brains of Alzheimer's disease patients: implications for regulation of neuronal cell Signaling. *J. Alz. Dis. Rep.*, **1**, 115–124.
30. Janelidze, S., Bali, D., Ashton, N.J., Barthélemy, N.R., Vanbrabant, J., Stoops, E., Vanmechelen, E., He, Y., Dolado, A.O., Triana-Baltzer, G. *et al.* (2022) Head-to-head comparison of 10 plasma phospho-tau assays in prodromal Alzheimer's disease. *Brain*, **146**, 1592–1601.
31. Ali, M., Sung, Y.J., Wang, F., Fernández, M.V., Morris, J.C., Fagan, A.M., Blennow, K., Zetterberg, H., Heslegrave, A., Johansson, P.M. *et al.* (2022) Leveraging large multi-center cohorts of Alzheimer disease endophenotypes to understand the role of klotho heterozygosity on disease risk. *PLoS One*, **17**, e0267298.
32. Boughton, A.P., Welch, R.P., Flickinger, M., VandeHaar, P., Taliun, D., Abecasis, G.R. and Boehnke, M. (2021) LocusZoom.Js: interactive and embeddable visualization of genetic association study results. *Bioinf.*, **37**, 3017–3018.
33. Watanabe, K., Taskesen, E., Van Bochoven, A. and Posthuma, D. (2017) Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.*, **8**, 1826.
34. D. Turner, S. (2018) Qqman: an R package for visualizing GWAS results using Q-Q and Manhattan plots. *J. Open Source Softw.*, **3**, 731.
35. Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.C. and Müller, M. (2011) pROC: an open-source package for R and S+ to analyze and compare ROC curves. *Bioinformatics*, **12**(1), 77.
36. Venables, W.N. and Ripley, B.D. (2002, 2002) In Chambers, J., Eddy, W., Härdle, W., Sheather, S. and Tierney, L. (eds), *Modern Applied Statistics with S. Modern Applied Statistics with S*. Springer New York, New York, NY.
37. Cruchaga, C., Del-Aguila, J.L., Saef, B., Black, K., Fernandez, M.V., Budde, J., Ibanez, L., Deming, Y., Kapoor, M., Tosto, G. *et al.* (2018) Polygenic risk score of sporadic late-onset Alzheimer's disease reveals a shared architecture with the familial and early-onset forms. *Alz. and Dem.*, **14**, 205–214.