



OPEN ACCESS

EDITED BY

Lucia Carboni,
University of Bologna, Italy

REVIEWED BY

Marisol Herrera Rivero,
University of Münster, Germany
Yunpeng Wang,
University of Oslo, Norway

*CORRESPONDENCE

Jingchun Chen
✉ jingchun.chen@unlv.edu

[†]These authors have contributed equally to this work and share first authorship

RECEIVED 20 March 2024

ACCEPTED 01 July 2024

PUBLISHED 23 July 2024

CITATION

Cammann DB, Lu Y, Rotter JI, Wood AC and Chen J (2024) Polygenic scores and Mendelian randomization identify plasma proteins causally implicated in Alzheimer's disease.

Front. Neurosci. 18:1404377.

doi: 10.3389/fnins.2024.1404377

COPYRIGHT

© 2024 Cammann, Lu, Rotter, Wood and Chen. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Polygenic scores and Mendelian randomization identify plasma proteins causally implicated in Alzheimer's disease

Davis B. Cammann^{1†}, Yimei Lu^{1†}, Jerome I. Rotter², Alexis C. Wood³ and Jingchun Chen^{1*}

¹Nevada Institute of Personalized Medicine, University of Nevada Las Vegas, Las Vegas, NV, United States, ²The Lundquist Institute for Biomedical Innovation, Harbor-UCLA Medical Center, Torrance, CA, United States, ³Baylor College of Medicine, Houston, TX, United States

Background: An increasing body of evidence suggests that neuroinflammation is one of the key drivers of late-onset Alzheimer's disease (LOAD) pathology. Due to the increased permeability of the blood–brain barrier (BBB) in older adults, peripheral plasma proteins can infiltrate the central nervous system (CNS) and drive neuroinflammation through interactions with neurons and glial cells. Because these inflammatory factors are heritable, a greater understanding of their genetic relationship with LOAD could identify new biomarkers that contribute to LOAD pathology or offer protection against it.

Methods: We used a genome-wide association study (GWAS) of 90 different plasma proteins ($n = 17,747$) to create polygenic scores (PGSs) in an independent discovery (cases = 1,852 and controls = 1,990) and replication (cases = 799 and controls = 778) cohort. Multivariate logistic regression was used to associate the plasma protein PGSs with LOAD diagnosis while controlling for age, sex, principal components 1–2, and the number of *APOE*-e4 alleles as covariates. After meta-analyzing the PGS-LOAD associations between the two cohorts, we then performed a two-sample Mendelian randomization (MR) analysis using the summary statistics of significant plasma protein level PGSs in the meta-analysis as an exposure, and a GWAS of clinically diagnosed LOAD (cases = 21,982, controls = 41,944) as an outcome to explore possible causal relationships between the two.

Results: We identified four plasma protein level PGSs that were significantly associated (FDR-adjusted $p < 0.05$) with LOAD in a meta-analysis of the discovery and replication cohorts: CX3CL1, hepatocyte growth factor (HGF), TIE2, and matrix metalloproteinase-3 (MMP-3). When these four plasma proteins were used as exposures in MR with LOAD liability as the outcome, plasma levels of HGF were inferred to have a negative causal relationship with the disease when single-nucleotide polymorphisms (SNPs) used as instrumental variables were not restricted to cis-variants (OR/95%CI = 0.945/0.906–0.984, $p = 0.005$).

Conclusion: Our results show that plasma HGF has a negative causal relationship with LOAD liability that is driven by pleiotropic SNPs possibly involved in other pathways. These findings suggest a low transferability between PGS and MR approaches, and future research should explore ways in which LOAD and the plasma proteome may interact.

KEYWORDS

polygenic score, Mendelian randomization, Alzheimer's disease, plasma proteins, pQTL

1 Introduction

Late-onset Alzheimer's disease (LOAD) is a progressive, neurodegenerative condition with no known cure and a diverse range of contributing pathologies that make it difficult to diagnose (Fang et al., 2020). Even after diagnosis, this treatment is hindered because LOAD patients often exhibit heterogeneity in their clinical symptoms at diagnosis (Devi and Scheltens, 2018), brain neuropathology (Murray et al., 2011), and comorbidities from other diseases such as type 2 diabetes (Santiago and Potashkin, 2021). Therefore, they likely require therapeutic options tailored to their individual needs. Due to the highly polygenic nature of LOAD, there are many potential contributing genetic factors to either disease risk or protection against it, none of which are individually necessary or sufficient for the development of LOAD (Baker and Escott-Price, 2020). Furthermore, many of the genes contributing to the risk of LOAD are considered to be pleiotropic, with upstream effects on multiple different traits that may give rise to some of the patterns of comorbidities seen in the LOAD patient population. Understanding the polygenic overlap of different traits with LOAD could help guide diagnosis and treatment by pinpointing which factors negatively or positively contribute to the overall health of patients, thus aiding in the management of risks.

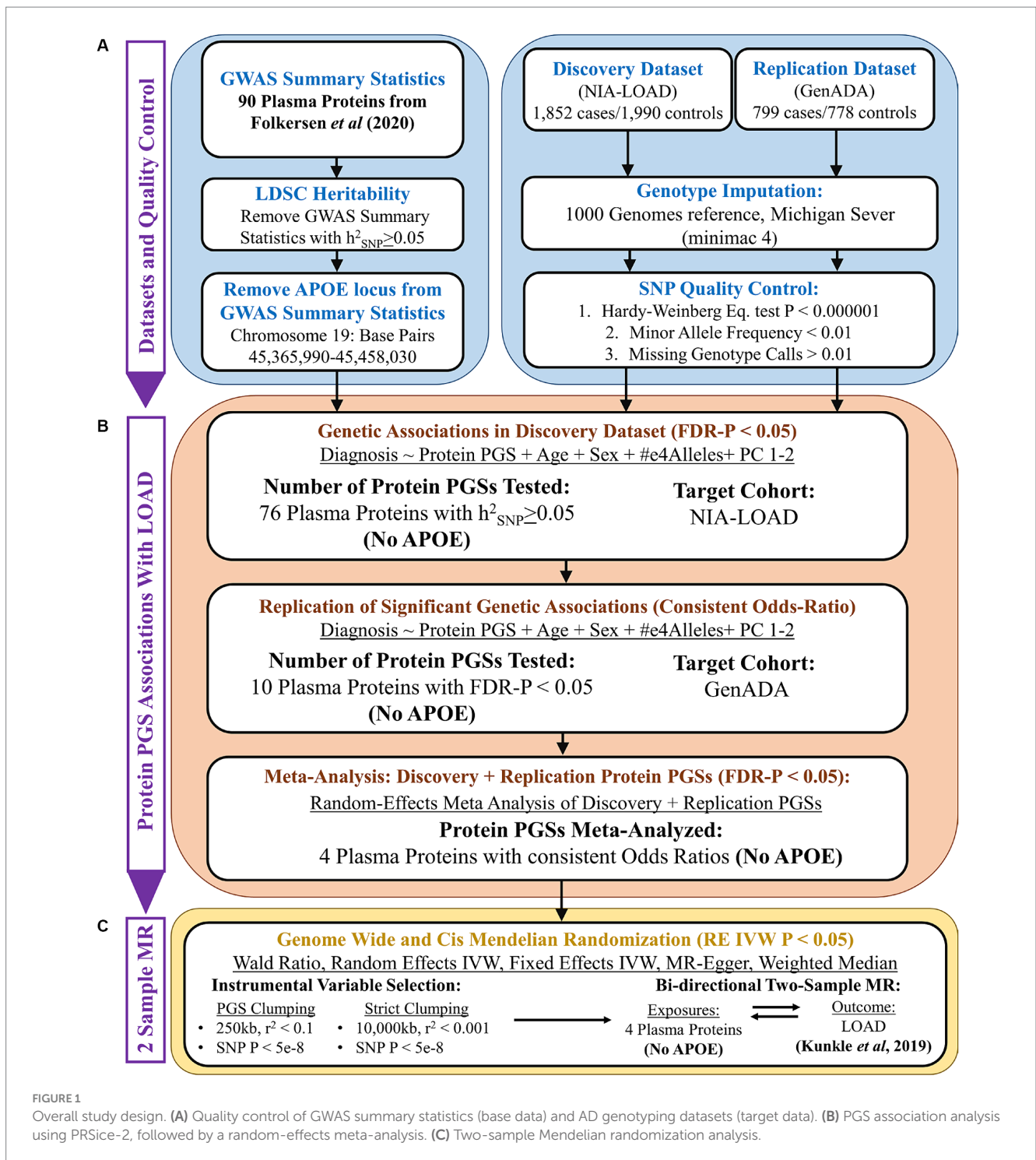
Like LOAD, circulating levels of cytokines and other plasma proteins are highly heritable and polygenic (de Craen et al., 2005; Li et al., 2016). Blood samples from children of parents with a history of LOAD show a higher production capacity for multiple pro-inflammatory cytokines than those from children without a familial history of LOAD, suggesting a shared genetic liability between inflammation and LOAD (van Exel et al., 2009). This notion is supported by recent genome-wide association studies (GWASs) for LOAD, in which several associated loci were implicated in inflammatory pathways (Harold et al., 2009) and candidate gene investigations (Desikan et al., 2015). Genetic risk factors for LOAD, including the *APOE*- ϵ 4 isoform and rare variants of *TREM2*, have further been shown to exacerbate neuroinflammation through their effect on the activation state of microglia (Colonna and Wang, 2016; Parhizkar and Holtzman, 2022). Recent studies have also suggested a genetic overlap between predictors of circulating proteins and LOAD, with some exploring inflammation-specific markers (van der Linden et al., 2021a) and others focusing on plasma proteins with evidence of a role in the disease (Handy et al., 2021).

The understanding of polygenic traits and their relationship with other phenotypes has been considerably improved by the implementation of polygenic scores (PGSs), which are single-unit estimates of an individual's genetic liability for a trait (Dudbridge, 2013). PGSs represent the sum of individual predisposing single-nucleotide polymorphisms (SNPs), which tend to be weighted by their effect size as drawn from a GWAS. A PGS is typically calculated using individual genotyping data in a target population

according to the effect allele of SNPs and their effect sizes provided by GWAS summary statistics. PGSs are most commonly developed for disease prediction, with genotyping information and GWAS summary statistics for the same trait. For example, LOAD-specific PGSs have achieved an area under the curve of up to 84% in distinguishing LOAD cases from controls (Leonenko et al., 2021). Using summary statistics of a trait different from the target trait provides a measure of shared genetic etiology between the two (Choi et al., 2020). This approach has seen recent success in identifying genetic associations between the blood levels of 31 different lipids consistent across two independent LOAD target cohorts (van der Linden et al., 2021b) and in showing associations between a LOAD PGS and markers of inflammation (Morgan et al., 2017). However, as shown by Handy et al. (2021), using a GWAS of plasma proteins to create PGSs of LOAD for association testing requires further causal validation, as even literature-selected proteins may show a weak association with the disease after accounting for factors such as population stratification.

A genetic association does not necessarily indicate that a trait causally contributes to disease pathology, as even at the genetic level, associations are subject to confounding. Mendelian randomization (MR) provides a means by which a causal relationship can be inferred between heritable traits by using SNPs as genetic instruments to test for a causal effect (Davey Smith and Ebrahim, 2003; Davey Smith and Hemani, 2014). In MR, a causal relationship between an exposure and an outcome is inferred where SNPs associated with an exposure trait, e.g., plasma protein levels, show a proportional association with an outcome trait, e.g., LOAD, under the assumption that they do not have an independent impact on the outcome (i.e., no pleiotropic effects across exposure and outcome), and are not associated with confounding variables (Davey Smith and Hemani, 2014). Two-sample MR leverages the MR approach using the summary statistics of two independent GWASs to serve as the exposure and outcome traits. For example, using well-powered GWASs across several measures of body mass, one study found a protective relationship between genetic predictors of lean body mass and LOAD (Daghlal et al., 2023), implicating a causal effect.

Using a recent GWAS of 90 different plasma proteins published by the SCALLOP consortium (Folkersen et al., 2020), we sought to develop plasma protein PGSs and test their association with LOAD diagnosis as a means of identifying novel genetic etiologies and potential pathways that may be associated with the disease. We then selected plasma protein PGSs that showed significant associations with LOAD as exposures in a two-sample MR analysis where LOAD liability served as the outcome to test whether they play a causal role in the disease. We expect that our unbiased selection of proteins and the larger participant sample size ($n = 17,747$) of the plasma protein GWAS used compared to prior studies will aid our goal of discovering novel relationships between plasma proteins and LOAD, which may assist future efforts in identifying diagnostic factors for the disease.



2 Materials and methods

2.1 Study design

The overall design of this study is given in Figure 1. The goal of this study was to discover plasma proteins with a genetic association with LOAD diagnosis using PGSs and then determine if any of them have a causal relationship with LOAD using two-sample MR. To do this, we first performed quality control on the GWAS summary statistics of plasma protein levels, using LDSC to calculate the

heritability of each plasma protein and retain summary statistics with an $h^2_{SNP} \geq 0.05$. To control the known strong effects of the APOE locus, we removed it from the GWAS summary statistics prior to their use in creating plasma proteins PGSs and as exposures in MR (Figure 1A).

Next, PGSs were calculated in our discovery cohort using the summary statistics of plasma proteins (Figure 1B). Plasma protein PGSs that had a significant association (FDR-adjusted $p < 0.05$) with LOAD diagnosis while accounting for the age, sex, APOE-e4 genotype, and first two genetic principal components (PCs) of discovery cohort participants were tested in our replication cohort. PGS-LOAD

TABLE 1 Demographic information of genotyping data.

	NIA-LOAD			GenADA		
	Cases	Controls	Total	Cases	Controls	Total
Sample Size	1852	1990	3,842	799	778	1,577
Age (Mean ± SD)	76.7 ± 6.99	70.8 ± 10.8	73.6 ± 9.61	72.2 ± 8.41	73.4 ± 7.92	72.8 ± 8.19
Sex (Male/Female)	645/1207	789/1201	1434/2408	339/460	276/502	615/962
APOE-ε4 Alleles, n (%)	0	575 (31.0)	1,303 (65.5)	1878 (48.9)	296 (37.0)	589 (56.1)
	1	1,003 (54.2)	634 (31.9)	1,637 (42.6)	397 (49.7)	574 (36.4)
	2	274 (14.8)	53 (2.66)	327 (8.51)	106 (13.3)	118 (7.48)

Age, age at onset (cases)/age at examination (controls); APOE-ε4 alleles, number of ε4 alleles in the genome of each individual.

associations with a consistent direction of effect by their odds ratio (OR) in the replication cohort were then meta-analyzed with their discovery cohort PGS-LOAD association in a random-effects model.

The GWAS summary statistics of significant plasma protein PGS-LOAD associations in the meta-analysis (FDR-adjusted $p < 0.05$) were used as exposures in a bidirectional two-sample MR analysis with LOAD as the outcome (Figure 1C). Plasma proteins with a p -value less than 0.05 by the random-effects inverse-variance weighted (RE IVW) or Wald ratio methods, and a consistent direction of effect in sensitivity analyses (fixed effects IVW, MR-Egger, weighted median), were inferred to have a causal effect on LOAD liability. As an additional sensitivity analysis, we used a LOAD GWAS as the exposure and the prior plasma protein exposures as outcomes to test possible bidirectional relationships between the plasma proteins and LOAD.

2.2 Genotyping (target) data

In this study, two LOAD genotyping datasets were requested from the database of Genotypes and Phenotypes (dbGaPs) for PGS association analyses. These include the National Institute of Aging Late-Onset Alzheimer's Disease (NIA-LOAD) study (phs000168.v2.p2) (Lee et al., 2008) and the Multi-Site Collaborative Study for Genotype-Phenotype Associations in Alzheimer's Disease (GenADA) (phs000219.v1.p1) (Li et al., 2008; Filippini et al., 2009). Both studies were conducted on European American (EA) individuals, except the NIA-LOAD study, which included a small cohort of African Americans (AA). Because the plasma protein GWASs were conducted in European-origin populations, we restricted our analyses of the NIA-LOAD study to EA individuals identified using principal component analysis (PCA). The NIA-LOAD study was used as our discovery cohort, while the GenADA study was used as our replication cohort. Information on the diagnosis, age, sex, and APOE-ε4 allele frequencies of individuals in the NIA-LOAD and GenADA datasets can be found in Table 1.

LOAD cases in both studies were defined as any individual with probable LOAD dementia by the National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) criteria. Controls were neurologically evaluated to be cognitively normal and matched to the age and sex of cases. NIA-LOAD participants (cases = 1,852 and controls = 1,990) had 601,273 SNPs genotyped using the Illumina Human610 QuadV1-B platform. In the GenADA study, participants (cases = 799 and controls = 778) were

genotyped using the Affymetrix 500k Set, which includes the Mapping 250k STY and Mapping 250k_NSP arrays. To fill in missing genetic information, both genotyping datasets were imputed to genome build 37 (hg19) with the 1,000 Genomes Phase 3v5 reference panel (Auton et al., 2015) on the Michigan Imputation Server¹ (Das et al., 2016). After imputation, we used Plink (v.1.9) to quality control SNPs with imputation quality (INFO) greater than 0.3, a minor allele frequency less than 0.01, Hardy-Weinberg equilibrium test p -value less than 0.000001, missing genotype rate, and missing rate per person less than 0.01 (Chang et al., 2015), resulting in a final total of 8,530,670 SNPs in both datasets.

2.3 GWAS summary statistics

In this study, we used summary statistics from GWASs of plasma levels of proteins and of LOAD. Prior to generating PGSs, we removed all SNPs within the APOE locus (genome build hg19-chromosome 19; base pairs 45,365,990–45,458,030) of the plasma protein summary statistics. This is because the APOE locus has a large effect on LOAD risk, and including many SNPs in linkage disequilibrium (LD) with APOE would falsely tag its effects and potentially bias any PGS results (Farrell and Brookes, 2022). Information on the two studies is summarized in Table 2.

2.3.1 LOAD summary statistics

We downloaded the Stage 12,019 Kunkle et al. (2019) GWAS summary statistics from the IEU GWAS catalog.² The International Genomics of Alzheimer's Project (IGAP) is a large three-stage study based on GWAS on individuals of European ancestry. In stage 1, IGAP used genotyped and imputed data on 11,480,632 SNPs to meta-analyze GWAS datasets consisting of 21,982 LOAD cases and 41,944 cognitively normal controls from four consortia: The Alzheimer Disease Genetics Consortium (ADGC); The European Alzheimer's disease Initiative (EADI); The Cohorts for Heart and Aging Research in Genomic Epidemiology Consortium (CHARGE); and The Genetic and Environmental Risk in AD Consortium Genetic and Environmental Risk in AD/Defining Genetic, Polygenic and

¹ <https://imputationserver.sph.umich.edu>

² <https://www.ebi.ac.uk/gwas/studies/GCST007511>

TABLE 2 GWAS summary statistics information.

Author (Year)	Consortia	Trait(s)	Ancestry	Sample Size	PMID
Kunkle et al. (2019)	IGAP	Clinically diagnosed LOAD	European	21,982 Cases, 41,944 Controls	30,820,047
Folkersen et al. (2020)	SCALLOP	90 Plasma Proteins	European	17,747	33,067,605

Author (year) = Primary author of GWAS study and year of study publication, consortia = consortia involved in the conduct of the GWAS study, trait(s) = traits studied by GWAS represented in summary statistics, ancestry = study population ethnicity, sample size = number of study participants, PMID = PubMed ID.

Environmental Risk for Alzheimer's Disease Consortium (GERAD/PERADES).

2.3.2 Plasma protein summary statistics

The summary statistics for each of the 90 blood plasma proteins by Folkersen et al. (2020) were downloaded from Zenodo.³ To discover genome-wide significant loci for each of the 90 proteins, a meta-analysis was performed on 21.4 million SNPs derived from 13 studies totaling 21,758 European individuals. Due to inter-cohort differences in genotype imputation, each protein had an average sample size of 17,747 individuals with 20.3 million SNPs. In each cohort, blood plasma levels of proteins were measured using the Olink proximity extension assay cardiovascular 1 panel (Assarsson et al., 2014). The log₂ normalized protein expression (NPX) values from the Olink assay for each protein had been ranked and either inverse normal transformed or standardized to unit variance to control for batch effects. In total, 467 genome-wide significant loci were reported in the original study to be associated with 85 of the 90 blood plasma proteins.

2.4 Heritability estimated from GWAS summary statistics

The heritability (h^2) of a trait is defined as the proportion of its phenotypic variance attributed to genetic variance (Evans et al., 2018). Estimates of heritability ascribed to SNPs (h^2_{SNP}) are important in ensuring the reliability of analyses using GWAS data. Following the recommendation by Choi et al. (2020), we required an $h^2_{\text{SNP}} \geq 0.05$ for each plasma protein GWAS before performing PGS analyses. We calculated the h^2_{SNP} from each GWAS using LD score regression (LDSC) (Bulik-Sullivan et al., 2015). LDSC calculates an "LD Score" for each SNP in a GWAS, which measures the amount of genetic variance tagged by the SNP. The χ^2 association test statistic for each SNP is then regressed against their LD Score, and the slope of this regression serves as an estimate of the GWAS's h^2_{SNP} . Before a GWAS's h^2_{SNP} calculation, its SNPs were limited to those on an ancestry-matched reference panel of ~1.2 million SNPs from the HapMap 3 project to avoid estimating the h^2_{SNP} with genetic variants of low imputation quality that were not reported by the original GWAS.

2.5 Plasma protein PGS modeling

In this study, we generated PGSs with the PRSice-2 software (Choi and O'Reilly, 2019). PRSice-2 takes the "Clumping and Thresholding"

(C + T) approach to create a PGS. First, SNPs are grouped across user-defined kilobase (kb) sized regions of the genome, and SNPs in LD above an r^2 threshold are pruned to remove those that are highly correlated (Wray et al., 2014). p -values from the GWAS summary statistics are then used to select a set of clumped SNPs under different p -value thresholds (P_T), which are then used to generate the PGSs. A PGS is generated as the sum of effect alleles in a target individual's genome that are weighted by the effect size of those alleles drawn from GWAS summary statistics. We used standardized PGSs of plasma protein levels in our association analysis:

$$\text{PGS}_j = \frac{\sum_i (S_i \times G_{ij}) - \text{Mean}(\text{PGS})}{\text{SD}(\text{PGS})}$$

Where S_i is the effect size of the effect allele for SNP i , G_{ij} is the genotype of SNP i (0, 1, 2) for individual j .

For the C + T approach, we clumped SNPs in 250 kb regions of the genome that had an r^2 greater than 0.1. PGS models for each protein were calculated using the "best-fit" approach implemented in the PRSice-2 program, where a range of P_T s was applied to the base data (the plasma protein GWASs), attempting to find a set of SNPs under a certain P_T that can explain the most of the target cohort's phenotype (LOAD diagnosis). In this study, a range of P_T s was assessed from 5×10^{-8} to 1 with an incremental interval of 5×10^{-5} to find the best PGS model for each protein.

2.6 Multivariate logistic regression and meta-analysis

Using PRSice-2, we evaluated the association of our plasma protein PGSs with LOAD diagnosis in a multivariate logistic regression model that included age, sex, *APOE-ε4* allele genotype, and the first two genetic PCs of the target data individuals as covariates. We used these same covariates when performing the PGS-LOAD association in both the discovery and replication cohort. PGS-LOAD associations were considered significant if their FDR-adjusted p -value was less than 0.05. The random-effects meta-analysis between the PGS-LOAD associations of the discovery and replication cohorts was performed using the metafor package in R (Viechtbauer, 2010). PGS-LOAD associations were considered significant in the meta-analysis if their FDR-adjusted summary estimate p -value was less than 0.05.

2.7 Mendelian randomization

For our MR analysis, we used the TwoSampleMR (v.0.6.2) package in R (Hemani et al., 2018). MR is used to infer a causal relationship between

³ <https://zenodo.org/records/2615265>

an exposure trait and outcome trait when a set of SNPs associated with the exposure [referred to as instrumental variables (IVs)] are also associated with the outcome through their effects on the exposure, assuming three key assumptions are met. SNPs used as IVs must be highly associated with the exposure, not associated with traits that confound the exposure or outcome, and not independently associated with the outcome except through the exposure (Davey Smith and Hemani, 2014).

2.7.1 IV selection

For our primary MR analysis, we used the GWAS summary statistics of significant plasma proteins from the PGS meta-analysis as exposures, and a GWAS of clinically diagnosed LOAD as our outcome (Kunkle et al., 2019). In each plasma protein exposure, SNPs used as IVs were genome-wide significant (GWAS $p < 5 \times 10^{-8}$) and had a first-stage F-statistic greater than 10 to ensure that they were highly associated with the exposure and considered strong IVs (Pierce et al., 2011). To ensure that SNPs were not correlated via LD, we opted to clump them under “strict” parameters ($r^2 < 0.001$, kb=10,000) and “PGS” parameters ($r^2 < 0.1$, kb=250) that match the default variant clumping strategies of the TwoSampleMR and PRSice-2 software, respectively. This was performed to allow comparison between the methods. For plasma protein exposures, we also tested SNPs under these two clumping strategies in a “Cis” analysis, where only SNPs on the same chromosome as the protein’s original gene were used, and a “Genome-Wide” analysis, where SNPs could come from any chromosome.

2.7.2 Causal effect estimation and sensitivity analyses

To calculate the causal effect of our exposure traits on our outcome, we used the RE IVW or Wald ratio methods as our primary analysis. We used the RE IVW method to match the random-effects meta-analysis done in our PGS association analysis and used the Wald ratio to account for exposures with only one SNP as a valid IV (Bowden et al., 2017; Burgess et al., 2017). We used the fixed effects IVW (FE IVW), MR Egger, and weighted median methods as sensitivity analyses to assess the effects of horizontal pleiotropy and invalid SNPs (Burgess et al., 2013; Bowden et al., 2015, 2016). As an additional analysis to assess potential bidirectional relationships between plasma proteins and LOAD liability, we repeated

our two-sample MR analysis using the LOAD GWAS as the exposure and each previously used plasma protein GWAS as the outcome. We considered an exposure to have a significant causal effect on an outcome when its RE IVW or Wald ratio p -value was less than 0.05 and its sensitivity analyses had a concordant direction of effect with the primary method.

2.8 Ethics approval statement

This study was approved by the University of Nevada Las Vegas (UNLV) Office of Research Integrity (IRB). Informed consent was obtained from all subjects and/or their legal guardian(s) in the contributing studies. Contributing studies received ethical approval from their respective institutional review boards (IRBs).

3 Results

3.1 Plasma protein PGS associations with LOAD diagnosis

Of the 90 plasma proteins in the original Folkersen et al. GWAS, 76 were identified as sufficiently heritable ($h^2_{SNP} \geq 0.05$) for use in the PGS association analysis (Supplementary Table S1). Notably, the plasma protein with the highest h^2_{SNP} in Folkersen et al. was galectin-3 (Gal-3), with its SNPs accounting for an estimated 43.6% of the trait’s heritability by LDSC. After calculating PGSs for each of the 76 viable plasma proteins in our discovery cohort, we found that 10 were significantly associated with LOAD diagnosis (FDR-adjusted $p < 0.05$) (Table 3). Seven of these plasma protein PGSs had a positive association with LOAD diagnosis, including prolactin (PRL), hepatocyte growth factor (HGF), interleukin-1 receptor agonist (IL-1ra), vascular endothelial growth factor D (VEGF-D), interleukin-18 (IL-18), angiopoietin-1 receptor (TIE2), and kidney injury molecule 1 (KIM-1). Three plasma protein PGSs had a negative association with LOAD diagnosis: platelet-derived growth factor subunit B (PDGF-B), matrix metalloproteinase-3 (MMP-3), and fractalkine (CX3CL1). The most significant plasma protein PGS-LOAD association

TABLE 3 Ten plasma protein PGSs from Folkersen et al. were significantly associated with LOAD diagnosis (FDR $p < 0.05$) in the discovery dataset.

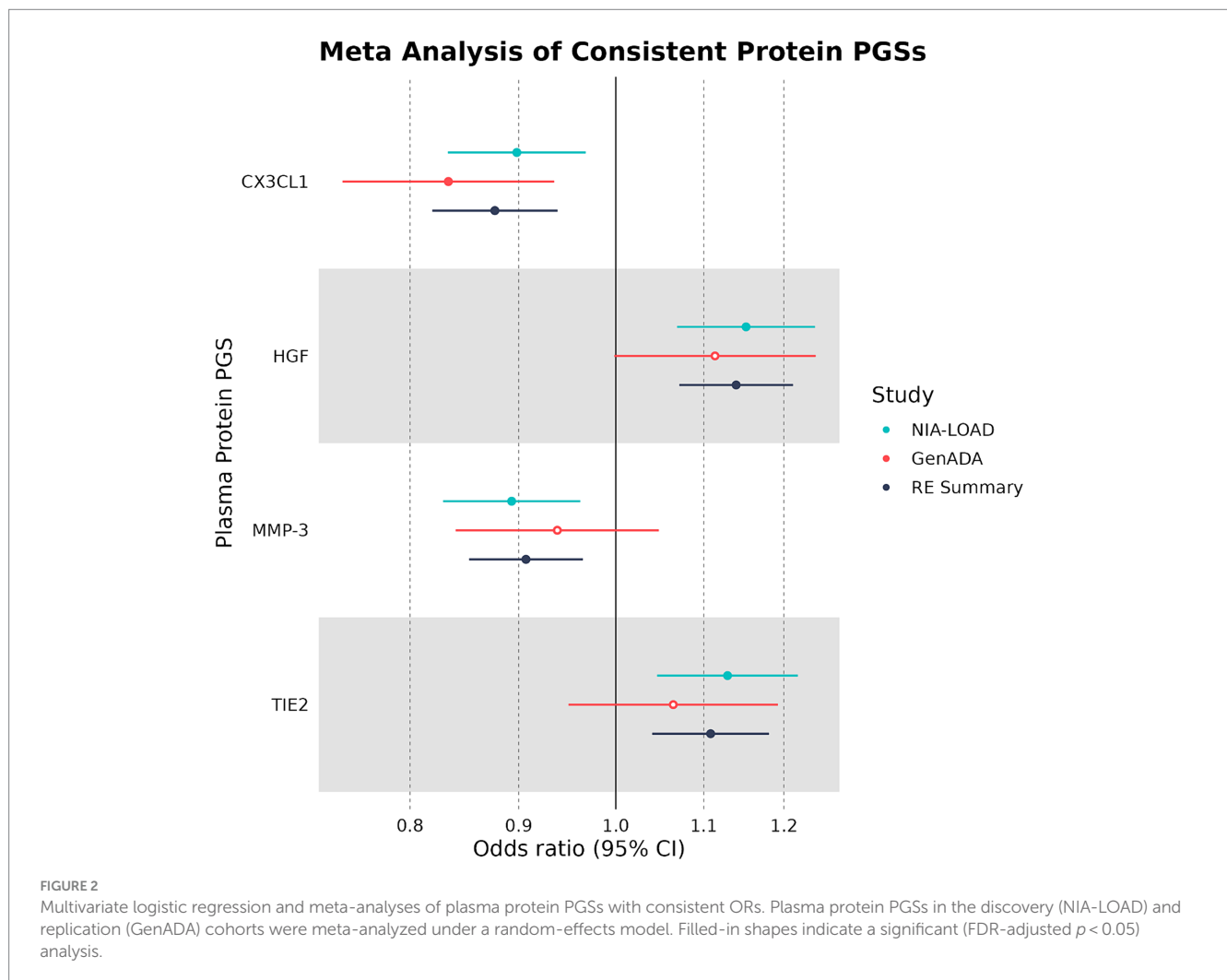
Protein	PVT	R ²	#SNPs	BETA	SE	OR (95%CI)	p	FDR
PRL	0.0023	0.0066	2,934	0.155	0.038	1.17 (1.09–1.24)	0.00004	0.003
HGF	0.01395	0.0053	17,367	0.141	0.038	1.15 (1.08–1.23)	0.00021	0.006
IL-1ra	0.00035	0.0052	701	0.138	0.038	1.15 (1.07–1.22)	0.00024	0.006
VEGF-D	0.01	0.0048	13,241	0.133	0.038	1.14 (1.07–1.22)	0.00045	0.009
PDGF-B	0.0036	0.0042	5,655	−0.126	0.039	0.882 (0.806–0.957)	0.00109	0.017
IL-18	0.06145	0.0038	56,531	0.121	0.039	1.13 (1.05–1.2)	0.00174	0.020
TIE2	0.091	0.0038	76,021	0.121	0.039	1.13 (1.05–1.21)	0.00188	0.020
MMP-3	0.0002	0.0034	567	−0.113	0.038	0.893 (0.819–0.968)	0.00293	0.028
CX3CL1	0.0014	0.0031	2,557	−0.107	0.038	0.898 (0.824–0.973)	0.00486	0.041
KIM-1	0.2858	0.0030	164,807	0.109	0.040	1.12 (1.04–1.19)	0.00576	0.044

Plasma protein PGS models generated by Folkersen et al. using multivariate logistic regression in the discovery (NIA-LOAD) cohort (covariates: age, sex, #APOE-ε4 alleles, principal components 1 and 2). PVT, “Best” p -value threshold from PRSice-2 used to select SNPs from GWAS summary statistics; R², Nagelkerke’s pseudo-R²; #SNPs, number of SNPs included at the PVT; BETA, PGS model coefficient; SE, PGS model standard error; OR (95%CI), odds ratio with 95% confidence interval; p, association p -value; FDR, false-discovery rate adjusted association p -value.

TABLE 4 Four plasma protein PGSs with consistent ORs in discovery and replication cohorts.

Protein	Study	PVT	R ²	BETA	SE	p	OR (95% CI)
CX3CL1	NIA-LOAD	0.0014	0.0031	-0.107	0.038	0.0049	0.898 (0.824–0.973)
	GenADA	0.0025	0.0088	-0.182	0.059	0.0019	0.834 (0.719–0.949)
HGF	NIA-LOAD	0.0140	0.0053	0.141	0.038	0.0002	1.15 (1.08–1.23)
	GenADA	0.0001	0.0034	0.108	0.056	0.0536	1.11 (1–1.22)
MMP-3	NIA-LOAD	0.0002	0.0034	-0.113	0.038	0.0029	0.893 (0.819–0.968)
	GenADA	0.0063	0.0012	-0.063	0.056	0.2593	0.939 (0.828–1.05)
TIE2	NIA-LOAD	0.091	0.0038	0.121	0.039	0.0019	1.13 (1.05–1.21)
	GenADA	0.1225	0.0010	0.062	0.058	0.2821	1.06 (0.951–1.18)

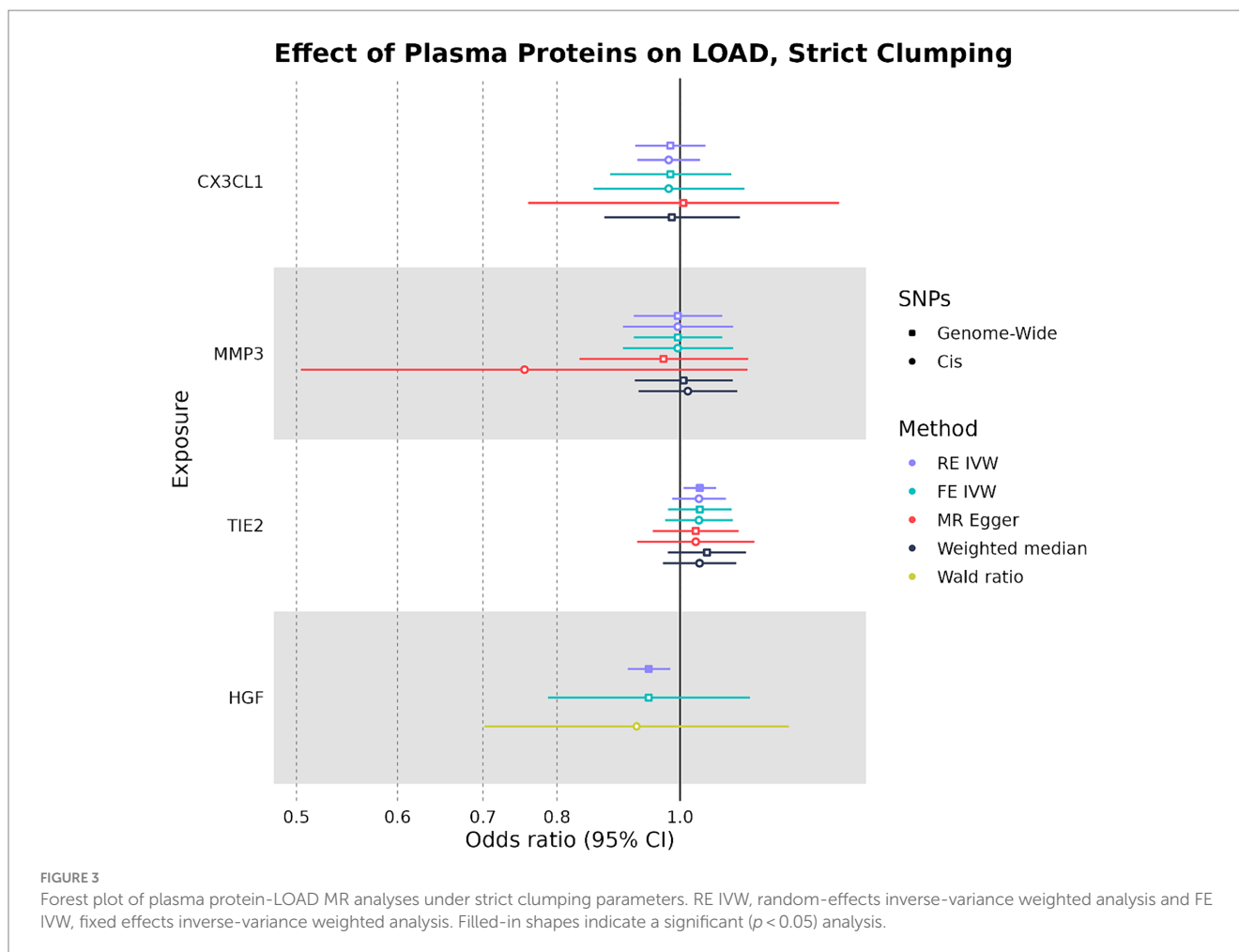
Plasma protein multivariate logistic regression models from Folkersen et al. in discovery (NIA-LOAD) and replication (GenADA) cohorts with consistent odds ratio. PGS models with $p < 0.05$ are bolded. Coefficient, logistic regression model coefficient; SE, standard error; p, association p-value; OR (95% CI), odds ratio with 95% confidence interval.



in the discovery cohort was PRL (OR: 1.17, 95%CI: 1.09–1.24, p -value: 0.00004).

Out of 10 plasma protein PGSs associated with LOAD diagnosis in the discovery cohort, 4 plasma protein PGSs had a consistent OR in the replication cohort (Table 4). Two of these plasma protein PGSs, CX3CL1 and MMP-3, had a consistent negative association with LOAD diagnosis, while HGF and TIE2 had a consistent positive association. Out of all plasma protein PGSs in the replication cohort, only CX3CL1

had a nominally significant association with LOAD diagnosis (OR: 0.842, 95%CI: 0.719–0.949, p -value: 0.0019). In a random-effects meta-analysis between the discovery and replication cohort plasma protein PGS-LOAD associations with consistent ORs, all associations remained significant (FDR-adjusted $p < 0.05$) (Figure 2). Overall, this analysis suggested that plasma protein PGSs of CX3CL1 and MMP-3 tended to be higher in control individuals, while PGSs of HGF and TIE2 were higher in individuals diagnosed with LOAD.



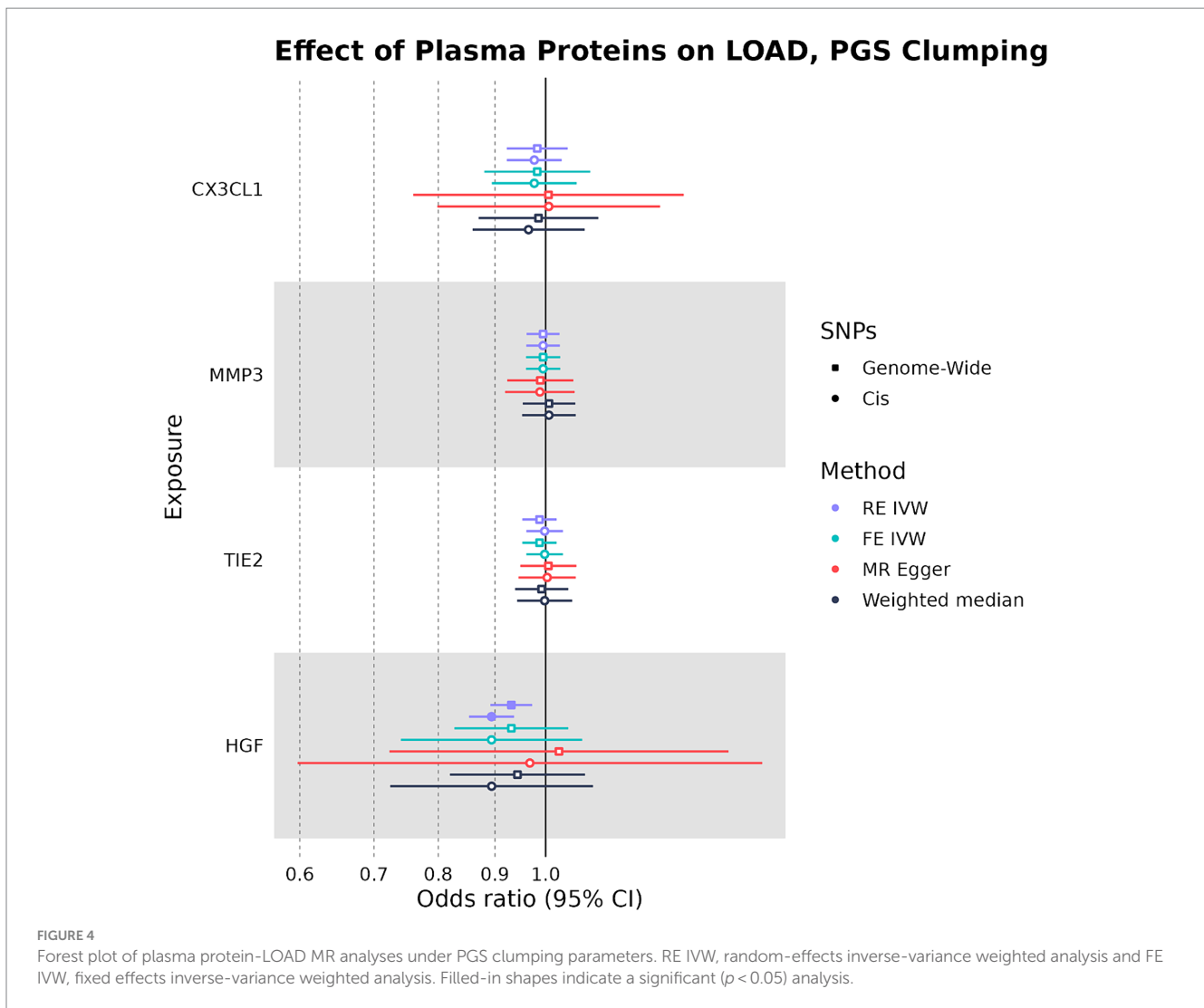
3.2 Two-sample MR analysis

We used the four plasma proteins from the random-effects meta-analysis as exposures in two-sample MR to see if any plasma protein had a causal effect on LOAD liability as an outcome. When we selected IVs under strict SNP clumping parameters, two plasma proteins were inferred to have a significant causal effect on LOAD liability, with HGF having a negative causal effect (OR: 0.945, 95%CI: 0.906–0.984, p -value: 0.004) and TIE2 having a positive causal effect (OR: 1.04, 95%CI: 1.01–1.07, p -value: 0.017) (Figure 3). As plasma HGF had only two valid SNPs for use as IVs in the strict clumping analysis, we were only able to test it under the RE IVW and FE IVW methods, which had a consistent direction of effect (Supplementary Table S3). Notably, plasma TIE2 and HGF were only significant under strict clumping parameters when SNPs used as IVs were sourced genome-wide, as restricting the SNPs to cis-pQTLs on the same chromosome as the plasma protein's original gene failed to replicate these observed causal effects (Figure 3).

When SNPs were clumped using PGS parameters, plasma HGF had a negative causal effect on LOAD liability that was significant by the RE IVW method in the genome-wide (OR: 0.931, 95%CI: 0.888–0.975, p -value: 0.0013) and cis (OR: 0.894, 95%CI: 0.847–0.940, p -value: 2.4×10^{-6}) analyses (Figure 4).

While plasma HGF sensitivity analyses had a consistent direction of effect in the cis analysis, the MR-Egger effect estimate of plasma HGF in the genome-wide analysis trended in the opposite direction (OR: 1.03) (Supplementary Table S4). In contrast to the strict clumping parameters, plasma TIE2 was not observed to have a causal effect on LOAD liability when using PGS clumping. Out of the original four plasma proteins used as exposures in both the PGS and strict clumping analyses, plasma HGF was the only one to have a consistent negative causal effect on LOAD liability.

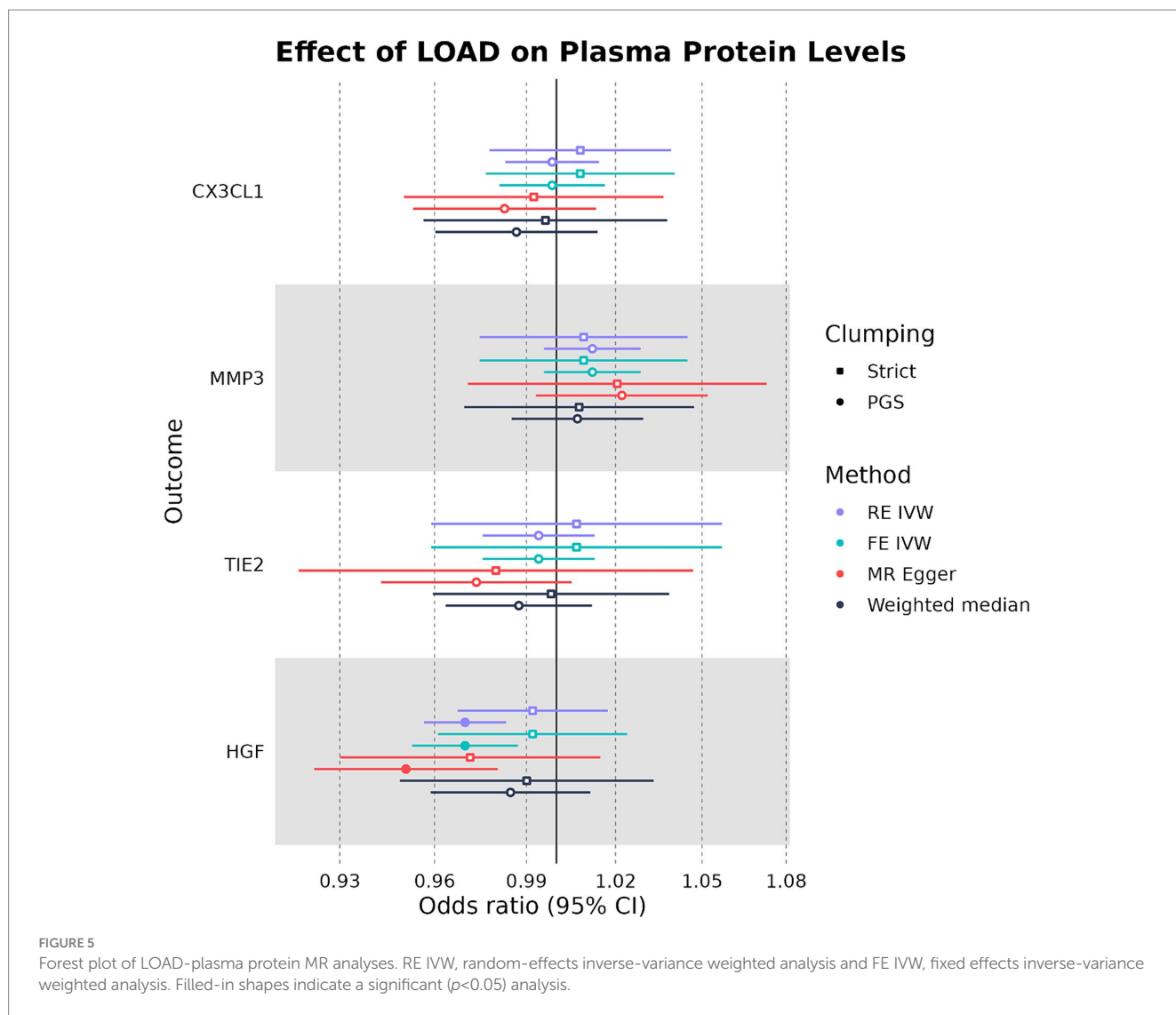
In our bidirectional analysis, where LOAD liability was used as an exposure and plasma CX3CL1, HGF, TIE2, and MMP-3 were used as outcomes, we found that LOAD liability had a significant negative causal effect on plasma HGF levels (OR: 0.956–0.984, p -value: 1.33×10^{-5}) (Figure 5). In addition to having a consistent direction of effect, the FE IVW (OR: 0.952–0.988, p -value: 0.00072) and MR-Egger (OR: 0.956–0.984, p -value: 1.33×10^{-5}) sensitivity analyses were also significant in the relationship between LOAD liability and plasma HGF (Supplementary Table S5). However, this effect on plasma HGF was only observed when SNPs were clumped with the PGS parameters. Overall, our MR analysis suggests that genetically proxied plasma HGF may exert a slight protective effect against LOAD liability, and in concordance, LOAD liability may have an effect on lowering levels of plasma HGF.



4 Discussion

Using GWAS summary statistics of the plasma levels of 76 different proteins ($n = 17,747$), we identified four plasma protein PGSs that had a significant association with LOAD diagnosis across both an independent discovery (cases = 1,852 and controls = 1,990) and replication (cases = 799 and controls = 778) cohort. Of these four plasma protein PGSs, CX3CL1 and MMP-3 had a negative association with the LOAD diagnosis, implicating a protective relationship. Plasma HGF and TIE2 PGSs had a positive association with LOAD diagnosis, suggesting a risk-factor relationship. Using two-sample MR, we found no bidirectional causal relationship between CX3CL1, MMP-3, or TIE2 and LOAD liability. We inferred a protective causal relationship between plasma HGF and LOAD liability when SNPs used as IVs were not restricted to the same chromosome as the HGF gene. Cis-IV selection for plasma HGF exposure only became significant when SNP clumping parameters were relaxed to those used during the PGS association analysis. Using this same clumping strategy, we also identified a negative causal effect of LOAD liability on plasma HGF levels, but not when the strict parameters were used to clump SNPs.

Combining plasma protein GWAS summary statistics from Folkersen et al. with LOAD genotyping from two independent cohorts, we identified four plasma proteins that were consistently associated with LOAD diagnosis through PGSs but did not translate into a consistent causal relationship between the levels of the plasma proteins and one's genetic liability for LOAD. Notably, plasma HGF PGSs had a positive association with LOAD diagnosis, but plasma HGF levels as an exposure were found to have a negative causal relationship with LOAD liability as an outcome in two-sample MR. HGF itself is a highly pleiotropic cytokine with functions across the body and CNS and is believed to play a role in the regulation of adult brain plasticity and learning (Shimamura et al., 2006; Kato et al., 2012; Desole et al., 2021). However, a prior MR investigation using plasma HGF as an exposure against LOAD and hippocampal volume outcomes found no evidence of a causal effect on either outcome using the IVW method, which is consistent with our findings using this same approach (Fani et al., 2021). HGF levels in the cerebrospinal fluid (CSF) have also shown positive correlations with mild cognitive impairment and other LOAD biomarkers in an observational study, further suggesting that any observed protective function of plasma HGF in LOAD may be pleiotropic rather than a direct effect on the brain (Zhao



et al., 2021). Given that plasma HGF as exposure was significant in our study when selecting genome-wide rather than cis SNPs under the RE IVW method, which is adjusted for possible pleiotropy, this could explain the observed causal effect (Bowden et al., 2017).

Our study has several strengths and limitations. A strength of our approach is the use of an independent discovery and replication cohort and meta-analysis to avoid false positives in the PGS association analysis, as well as filtering the GWAS summary statistics by their h^2_{SNP} beforehand to avoid the inclusion of plasma proteins with low SNP-based heritability. Another strength of our study is the use of an orthogonal statistical method, two-sample MR, to confirm results from our PGS analysis. This is because MR employs stricter assumptions for the SNPs proposed to drive an exposure–outcome relationship, most notably the use of SNPs that are significantly associated with the exposure, and the requirement that they do not have pleiotropic effects on the outcome or confounding factors. A potential limitation of developing PGSs for a trait separate from the phenotype of a target cohort is that the PGSs tend to explain a low amount of variance in the target phenotype. Although we were able to identify associations between the PGSs of plasma proteins and

LOAD diagnosis that were robust to confounding factors (age, sex, # *APOE*-e4 alleles, and genetic PCs), the PGSs themselves rarely explained more than 1% of the variance in the case/control phenotype. Additionally, we only considered the *APOE*-e4 genotype in our logistic regression models rather than the full *APOE* genotype, which leaves out the known protective effects of the *APOE*-e2 genotype against LOAD liability (Reiman et al., 2020). Other studies using PGSs to associate plasma proteins with LOAD have shown a similar trend in the variance explained by their PGSs (Handy et al., 2021; van der Linden et al., 2021a). This is likely because plasma proteins are involved in a myriad of other pathways unrelated to LOAD, leading to a large possibility of confounding and pleiotropy with a neurodegenerative disease of the CNS (Handy et al., 2021).

The potential for confounding in the plasma protein-LOAD relationship highlights the importance of two-sample MR to control potential false positives that may be influenced by these factors. Due to the aforementioned considerations about using a different “base” and “target” trait when generating PGSs, additional forms of verification are needed to ensure the validity of the association. To improve our approach for future studies, the inclusion of LOAD target cohorts in the PGS analysis with information available on the plasma levels of

proteins and metabolites in the participants could help to ensure the validity of the relevant protein PGSs. In addition, the inclusion of more diverse cohorts would improve the generalizability of the results. Future studies that seek to address the genetic relationship between plasma proteins and LOAD should focus on the role of confounders that may affect the interaction of these plasma proteins with the disease due to their role in different pathways. Additionally, future research should seek to understand the underlying mechanisms and pathways by which LOAD may induce changes in the plasma proteome.

Data availability statement

The summary statistics of the 90 plasma proteins by Folkersen et al are deposited in the Zenodo repository (<https://zenodo.org/records/2615265>). The summary statistics of the LOAD GWAS by Kunkle et al are deposited in the GWAS catalog, accession number GCST007511. The NIA-LOAD and GenADA genotyping datasets are deposited in the dbGaP database, with accession numbers phs000168.v2.p2 and phs000219.v1.p1 respectively.

Ethics statement

The studies involving humans were approved by University of Nevada Las Vegas (UNLV) Office of Research Integrity. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and institutional requirements.

Author contributions

DC: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. YL: Data curation, Methodology, Resources, Software, Visualization, Writing – review & editing. JR: Supervision, Writing – review & editing. AW: Supervision, Writing – review & editing. JC: Conceptualization, Data curation, Funding acquisition, Methodology, Project administration, Resources, Supervision, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This research

References

Assarsson, E., Lundberg, M., Holmquist, G., Björkstén, J., Thorsen, S. B., Ekman, D., et al. (2014). Homogenous 96-Plex PEA immunoassay exhibiting high sensitivity, specificity, and excellent scalability. *PLoS One* 9:e95192. doi: 10.1371/journal.pone.0095192

Auton, A., Abecasis, G. R., Altshuler, D. M., Durbin, R. M., Abecasis, G. R., Bentley, D. R., et al. (2015). A global reference for human genetic variation. *Nature* 526:7571. doi: 10.1038/nature15393

was funded to JC in part by NIH grants P20GM121325 and U54GM104944. This publication was also made possible by support from the Infectious Diseases Society of America (IDSA) Foundation (Reference #: 70823173). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the IDSA Foundation. This research was also supported in part by J.I.R. by the National Center for Advancing Translational Sciences, CTSI grant UL1TR001881, the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern California Diabetes Endocrinology Research Center, and the CHARGE Consortium by the National Heart, Lung, and Blood Institute (NHLBI) grant R01HL105756. ACW is supported, in part, by funding from NIH (R01AG058969), and from the USDA/ARS (Cooperative Agreement 58-3092-5-001). The contents of this publication do not necessarily reflect the views or policies of the USDA, nor does mention of trade names, commercial products, or organizations imply endorsement from the US government.

Acknowledgments

The authors thank the patients for their participation in the NIA/LOAD cohort and the GenADA cohort, as well as the original investigators who conducted these studies and made the data available.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnins.2024.1404377/full#supplementary-material>

Baker, E., and Escott-Price, V. (2020). Polygenic risk scores in Alzheimer's disease: current applications and future directions. *Front. Digital Health* 2:e14. doi: 10.3389/fdgh.2020.00014

Bowden, J., Davey Smith, G., and Burgess, S. (2015). Mendelian randomization with invalid instruments: effect estimation and bias detection through egger regression. *Int. J. Epidemiol.* 44, 512–525. doi: 10.1093/ije/dyv080

- Bowden, J., Davey Smith, G., Haycock, P. C., and Burgess, S. (2016). Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genet. Epidemiol.* 40, 304–314. doi: 10.1002/gepi.21965
- Bowden, J., Del Greco, M. F., Minelli, C., Davey Smith, G., Sheehan, N., and Thompson, J. (2017). A framework for the investigation of pleiotropy in two-sample summary data Mendelian randomization. *Stat. Med.* 36, 1783–1802. doi: 10.1002/sim.7221
- Bulik-Sullivan, B. K., Loh, P.-R., Finucane, H. K., Ripke, S., Yang, J., Patterson, N., et al. (2015). LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* 47, 291–295. doi: 10.1038/ng.3211
- Burgess, S., Butterworth, A., and Thompson, S. G. (2013). Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet. Epidemiol.* 37, 658–665. doi: 10.1002/gepi.21758
- Burgess, S., Small, D. S., and Thompson, S. G. (2017). A review of instrumental variable estimators for Mendelian randomization. *Stat. Methods Med. Res.* 26, 2333–2355. doi: 10.1177/0962280215597579
- Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., and Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* 4:e47. doi: 10.1186/s13742-015-0047-8
- Choi, S. W., Mak, T. S.-H., and O'Reilly, P. F. (2020). Tutorial: a guide to performing polygenic risk score analyses. *Nat. Protoc.* 15, 2759–2772. doi: 10.1038/s41596-020-0353-1
- Choi, S. W., and O'Reilly, P. F. (2019). PRSice-2: polygenic risk score software for biobank-scale data. *GigaScience* 8:giz082. doi: 10.1093/gigascience/giz082
- Colonna, M., and Wang, Y. (2016). TREM2 variants: new keys to decipher Alzheimer disease pathogenesis. *Nat. Rev. Neurosci.* 17, 201–207. doi: 10.1038/nrn.2016.7
- Daghlas, I., Nassan, M., and Gill, D. (2023). Genetically proxied lean mass and risk of Alzheimer's disease: Mendelian randomisation study. *BMJ Medicine* 2:e000354. doi: 10.1136/bmjmed-2022-000354
- Das, S., Forer, L., Schönherr, S., Sidore, C., Locke, A. E., Kwong, A., et al. (2016). Next-generation genotype imputation service and methods. *Nat. Genet.* 48, 1284–1287. doi: 10.1038/ng.3656
- Davey Smith, G., and Ebrahim, S. (2003). 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int. J. Epidemiol.* 32, 1–22. doi: 10.1093/ije/dyg070
- Davey Smith, G., and Hemani, G. (2014). Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum. Mol. Genet.* 23, R89–R98. doi: 10.1093/hmg/ddu328
- de Craen, A. J. M., Posthuma, D., Remarque, E. J., van den Biggelaar, A. H. J., Westendorp, R. G. J., and Boomsma, D. I. (2005). Heritability estimates of innate immunity: an extended twin study. *Genes Immunology* 6:167. doi: 10.1038/sj.gene.6364162
- Desikan, R. S., Schork, A. J., Wang, Y., Thompson, W. K., Dehghan, A., Ridker, P. M., et al. (2015). Polygenic overlap between C-reactive protein, plasma lipids, and Alzheimer disease. *Circulation* 131, 2061–2069. doi: 10.1161/CIRCULATIONAHA.115.015489
- Desole, C., Gallo, S., Vitacolonna, A., Montarolo, F., Bertolotto, A., Vivien, D., et al. (2021). HGF and MET: from brain development to neurological disorders. *Front Cell Dev Biol* 9:683609. doi: 10.3389/fcell.2021.683609
- Devi, G., and Scheltens, P. (2018). Heterogeneity of Alzheimer's disease: consequence for drug trials? *Alzheimers Res. Ther.* 10:122. doi: 10.1186/s13195-018-0455-y
- Dudbridge, F. (2013). Power and predictive accuracy of polygenic risk scores. *PLoS Genet.* 9:e1003348. doi: 10.1371/journal.pgen.1003348
- Evans, L. M., Tahmasbi, R., Vrieze, S. I., Abecasis, G. R., Das, S., Gazal, S., et al. (2018). Comparison of methods that use whole genome data to estimate the heritability and genetic architecture of complex traits. *Nat. Genet.* 50:5, 737–745. doi: 10.1038/s41588-018-0108-x
- Fang, J., Pieper, A. A., Nussinov, R., Lee, G., Bekris, L., Leverenz, J. B., et al. (2020). Harnessing endophenotypes and network medicine for Alzheimer's drug repurposing. *Med. Res. Rev.* 40, 2386–2426. doi: 10.1002/med.21709
- Fani, L., Georgakis, M. K., Ikram, M. A., Ikram, M. K., Malik, R., and Dichgans, M. (2021). Circulating biomarkers of immunity and inflammation, risk of Alzheimer's disease, and hippocampal volume: a Mendelian randomization study. *Transl. Psychiatry* 11:e1400. doi: 10.1038/s41398-021-01400-z
- Farrell, C., and Brookes, K. (2022). Utilising polygenic risk score analysis for AD to determine the "sphere of influence" of the APOE isoform SNPs. *J. Neurol. Neuromed* 6, 1–7. doi: 10.29245/2572.942X/2022/2.1284
- Filippini, N., Rao, A., Wetten, S., Gibson, R. A., Borrie, M., Guzman, D., et al. (2009). Anatomically-distinct genetic associations of APOE ε4 allele load with regional cortical atrophy in Alzheimer's disease. *NeuroImage* 44, 724–728. doi: 10.1016/j.neuroimage.2008.10.003
- Folkersen, L., Gustafsson, S., Wang, Q., Hansen, D. H., Hedman, Å. K., Schork, A., et al. (2020). Genomic and drug target evaluation of 90 cardiovascular proteins in 30,931 individuals. *Nat. Metab.* 2, 1135–1148. doi: 10.1038/s42255-020-00287-2
- Handy, A., Lord, J., Green, R., Xu, J., Aarsland, D., Velayudhan, L., et al. (2021). Assessing genetic overlap and causality between blood plasma proteins and Alzheimer's disease. *J. Alzheimers Dis.* 83, 1825–1839. doi: 10.3233/JAD-210462
- Harold, D., Abraham, R., Hollingworth, P., Sims, R., Gerrish, A., Hamshere, M. L., et al. (2009). Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. *Nat. Genet.* 41, 1088–1093. doi: 10.1038/ng.440
- Hemani, G., Zheng, J., Elsworth, B., Wade, K. H., Haberland, V., Baird, D., et al. (2018). The MR-base platform supports systematic causal inference across the human phenome. *eLife* 7:e34408. doi: 10.7554/eLife.34408
- Kato, T., Funakoshi, H., Kadoyama, K., Noma, S., Kanai, M., Ohya-Shimada, W., et al. (2012). Hepatocyte growth factor overexpression in the nervous system enhances learning and memory performance in mice. *J. Neurosci. Res.* 90, 1743–1755. doi: 10.1002/jnr.23065
- Kunkle, B. W., Grenier-Boley, B., Sims, R., Bis, J. C., Damotte, V., Naj, A. C., 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. doi: 10.1038/s41588-019-0358-2
- Lee, J. H., Cheng, R., Graff-Radford, N., Foroud, T., and Mayeux, R. National Institute on Aging Late-Onset Alzheimer's Disease Family Study Group (2008). Analyses of the National Institute on Aging late-onset Alzheimer's disease family study: implication of additional loci. *Arch. Neurol.* 65, 1518–1526. doi: 10.1001/archneur.65.11.1518
- Leonko, G., Baker, E., Stevenson-Hoare, J., Sierksma, A., Fiers, M., Williams, J., et al. (2021). Identifying individuals with high risk of Alzheimer's disease using polygenic risk scores. *Nat. Commun.* 12:e1. doi: 10.1038/s41467-021-24082-z
- Li, Y., Oosting, M., Smeekens, S. P., Jaeger, M., Aguirre-Gamboa, R., Le, K. T. T., et al. (2016). A functional genomics approach to understand variation in cytokine production in humans. *Cell* 167, 1099–1110.e14. doi: 10.1016/j.cell.2016.10.017
- Li, H., Wetten, S., Li, L., Jean, P. L., Upmanyu, R., Surh, L., et al. (2008). Candidate single-nucleotide polymorphisms from a Genomewide association study of Alzheimer disease. *Arch. Neurol.* 65, 45–53. doi: 10.1001/archneur.2007.3
- Morgan, A. R., Touchard, S., O'Hagan, C., Sims, R., Majounie, E., Escott-Price, V., et al. (2017). The correlation between inflammatory biomarkers and polygenic risk score in Alzheimer's disease. *J. Alzheimer's Disease* 56, 25–36. doi: 10.3233/JAD-160889
- Murray, M. E., Graff-Radford, N. R., Ross, O. A., Petersen, R. C., Duara, R., and Dickson, D. W. (2011). Neuropathologically defined subtypes of Alzheimer's disease with distinct clinical characteristics: a retrospective study. *Lancet Neurol* 10, 785–796. doi: 10.1016/S1474-4422(11)70156-9
- Parhizkar, S., and Holtzman, D. M. (2022). APOE mediated neuroinflammation and neurodegeneration in Alzheimer's disease. *Semin. Immunol.* 59:101594. doi: 10.1016/j.smim.2022.101594
- Pierce, B. L., Ahsan, H., and VanderWeele, T. J. (2011). Power and instrument strength requirements for Mendelian randomization studies using multiple genetic variants. *Int. J. Epidemiol.* 40, 740–752. doi: 10.1093/ije/dyq151
- Reiman, E. M., Arboleda-Velasquez, J. F., Quiroz, Y. T., Huentelman, M. J., Beach, T. G., Caselli, R. J., et al. (2020). Exceptionally low likelihood of Alzheimer's dementia in APOE2 homozygotes from a 5,000-person neuropathological study. *Nat. Commun.* 11:667. doi: 10.1038/s41467-019-14279-8
- Santiago, J. A., and Potashkin, J. A. (2021). The impact of disease comorbidities in Alzheimer's disease. *Front. Aging Neurosci.* 13:631770. doi: 10.3389/fnagi.2021.631770
- Shimamura, M., Sato, N., Waguri, S., Uchiyama, Y., Hayashi, T., Iida, H., et al. (2006). Gene transfer of hepatocyte growth factor gene improves learning and memory in the chronic stage of cerebral infarction. *Hypertension* 47, 742–751. doi: 10.1161/01.HYP.0000208598.57687.3e
- van der Linden, R. J., De Witte, W., and Poelmans, G. (2021a). Shared genetic etiology between Alzheimer's disease and blood levels of specific cytokines and growth factors. *Genes* 12:865. doi: 10.3390/genes12060865
- van der Linden, R. J., Reus, L. M., De Witte, W., Tijms, B. M., Olde Rikkert, M., Visser, P. J., et al. (2021b). Genetic overlap between Alzheimer's disease and blood lipid levels. *Neurobiol. Aging* 108, 189–195. doi: 10.1016/j.neurobiolaging.2021.06.019
- van Exel, E., Eikelenboom, P., Comijs, H., Frölich, M., Smit, J. H., Stek, M. L., et al. (2009). Vascular factors and markers of inflammation in offspring with a parental history of late-onset Alzheimer disease. *Arch. Gen. Psychiatry* 66, 1263–1270. doi: 10.1001/archgenpsychiatry.2009.146
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *J. Stat. Softw.* 36, 1–48. doi: 10.18637/jss.v036.i03
- Wray, N. R., Lee, S. H., Mehta, D., Vinkhuyzen, A. A. E., Dudbridge, F., and Middeldorp, C. M. (2014). Research review: polygenic methods and their application to psychiatric traits. *J. Child Psychol. Psychiatry* 55, 1068–1087. doi: 10.1111/jcpp.12295
- Zhao, L.-J., Wang, Z.-T., Ma, Y.-H., Zhang, W., Dong, Q., Yu, J.-T., et al. (2021). Associations of the cerebrospinal fluid hepatocyte growth factor with Alzheimer's disease pathology and cognitive function. *BMC Neurol.* 21:387. doi: 10.1186/s12883-021-02356-9