Cerebrospinal fluid proteome profiling using machine learning shows a unique protein signature associated with APOE4 genotype

Artur Shvetcov, PhD, [1,2#], Shannon Thomson, PhD, [3,4], Ann-Na Cho, PhD, [5], Heather M. Wilkins, PhD, [6,7,8], Joanne H. Reed, PhD, [3,4], Russell H. Swerdlow, MD, [6,7,8,9], David A. Brown, PhD, [2,10,11], the Alzheimer's Disease Neuroimaging Initiative [12], & Caitlin A. Finney, PhD [3,4#*]

1. Department of Psychological Medicine, Sydney Children's Hospital Network, Sydney, NSW, Australia

2. Discipline of Psychiatry and Mental Health, School of Clinical Medicine, Faculty of Medicine and Health, University of New South Wales, Sydney, NSW, Australia

3. Centre for Immunology and Allergy Research, Westmead Institute for Medical Research, Sydney, NSW, Australia

4. School of Medical Sciences, Faculty of Medicine and Health, The University of Sydney, Sydney, NSW, Australia

5. School of Biomedical Engineering, Faculty of Engineering, The University of Sydney, Sydney, NSW, Australia

6. University of Kansas Alzheimer's Disease Centre, Kansas City, KS, USA

7. Department of Biochemistry and Molecular Biology, University of Kansas Medical Centre, Kansas City, KS, USA

8. Department of Neurology, University of Kansas Medical Centre, Kansas City, KS, USA

9. Department of Molecular and Integrative Physiology, University of Kansas Medical Centre, Kansas City, KS, USA

10. Department of Immunopathology, Institute for Clinical Pathology and Medical Research-New South Wales Health Pathology, Sydney, NSW, Australia

11. Westmead Clinical School, Faculty of Medicine and Health, The University of Sydney, Sydney, NSW, Australia

12. Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

# These authors contributed equally to this work

* Author to whom correspondence should be addressed. E: caitlin.finney@wimr.org.au; A: 176 Hawkesbury Road, Westmead, New South Wales, Australia 2145.

Abstract

APOE4 is the biggest genetic risk factor for Alzheimer's disease (AD). Proteome-wide changes independent of AD brain pathology and whether these extend to APOE4 carriers irrespective of cognitive status remain unknown. To investigate APOE4-associated proteome changes in people with and without AD. Patient clinical, APOE genotype, and cerebrospinal fluid (CSF) proteome data for 735 participants was sourced from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. Participants with no cognitive impairment, mild cognitive impairment (MCI), and AD were included. Diagnosis was defined using cognitive assessments. Supervised machine learning (classification and regression trees; CART) was used for proteome profiling to identify protein signatures. Enrichment analyses for brain regions and cells and peripheral immune cells were performed using NetworkAnalyst 3.0 and the Human Protein Atlas. CART revealed an APOE4 specific proteome signature consisting of 7 proteins that were independent predictors of APOE4 carriers (sensitivity, specificity, PPV, NPV, and AUC = 1.0) and 50 proteins that interacted together for prediction (sensitivity = 0.99, specificity = 0.74, PPV = 0.86, NPV = 0.98, AUC = 0.92). This signature was found across APOE4 carriers independently of diagnosis and was associated with an increased risk of progression to cognitive impairment over time. Proteins within the signature were enriched in brain regions including the caudate and cortex. Enriched brain cells included endothelial cells, oligodendrocytes, and astrocytes. Enriched peripheral immune cells included T cells, macrophages, and B cells. We identified an APOE4 specific CSF proteome signature of 57 proteins that was independent of cognitive status and was associated with an increased risk of progression to cognitive impairment. It was also associated with a strong immune and inflammatory phenotype across the brain and periphery. Our findings suggest that the APOE4 proteome signature is independent of AD-specific brain pathology and likely underlies APOE4 carriers' vulnerability to cognitive decline and AD.

Introduction

A mutation in the apolipoprotein E gene called ε4 (APOE4) remains the single biggest genetic risk factor for late onset Alzheimer's disease (AD) [1]. APOE4 carriers with AD show increased tau and amyloid-β accumulation, atrophy and worse cognitive phenotype [2,3]. Yet we still have a poor understanding of the mechanisms underlying the APOE4 risk allele.

Recent evidence suggests that APOE4 carriers with AD have differences in plasma and cerebrospinal fluid (CSF) proteins relative to non-carriers with AD [4-6]. These studies, however, have been limited by including only a small number of proteins (<300) and examining protein changes through the lens of brain-specific AD pathology including aggregated amyloid-β. Proteome-wide changes independent of AD brain pathology and whether these extend to APOE4 carriers irrespective of cognitive status remains unknown. Here, we use machine learning to profile the proteome of APOE4 carriers with and without cognitive impairment from the Alzheimer's Disease Neuroimaging Initiative (ADNI) cohort.

Methods

Data and Participants

We sourced patient clinical, APOE genotype, and CSF proteome data for 735 patients (Supplementary Table 1) through the ADNI database (https://ida.loni.usc.edu/). Diagnostic criteria included Mini-Mental State Examination (MMSE) scores of 24-30 for non-impaired (NI) and mild cognitive impairment (MCI) patients and 20-26 for AD and a Clinical Dementia Rating (CDR) score of 0 for NI, 0.5 for MCI, and 0.5-1 for AD [7]. DNA was extracted from blood samples for APOE genetic testing [7]. CSF proteomics was performed using the SomaScan 7k assay. Data was normalized and protein levels are reported as relative fluorescence unit (RFU) [8] (Supplementary Methods). The ADNI cohort study was approved by the institutional review boards of the participating ADNI centers, and all participants provided informed consent.

Statistical Analyses

Weight-of-evidence and information value were used for feature selection to identify a subset of proteins (1,534/6,082) that were the strongest predictors (>0.3) of group differentiation and classification and regression trees (CART) to identify the proteome signature with the highest predictive performance (see Supplementary Methods). Reported performance metrics included sensitivity, specificity, positive predictive value (PPV),

negative predictive value (NPV), and area under the curve (AUC). Data analysis and figures were done in R v 4.3.1 (packages: 'Information', 'caret', 'ggplot2') and GraphPad Prism v9.4.1. Protein-protein interaction (PPI) network and enrichment analyses were done using NetworkAnalyst 3.0 [9] and the Human Protein Atlas database [10,11].

Results

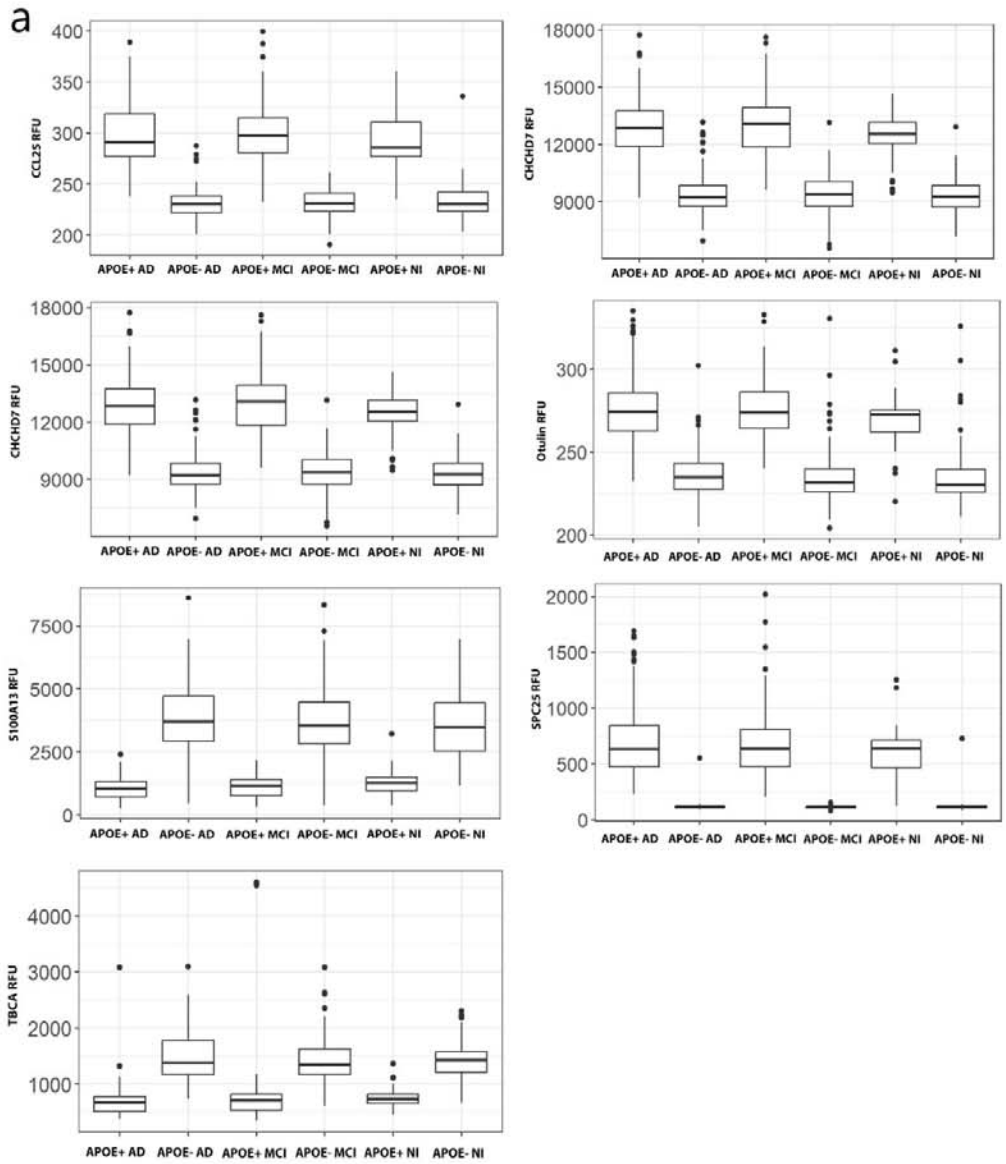
APOE4 carriers have a unique proteome signature irrespective of cognitive status

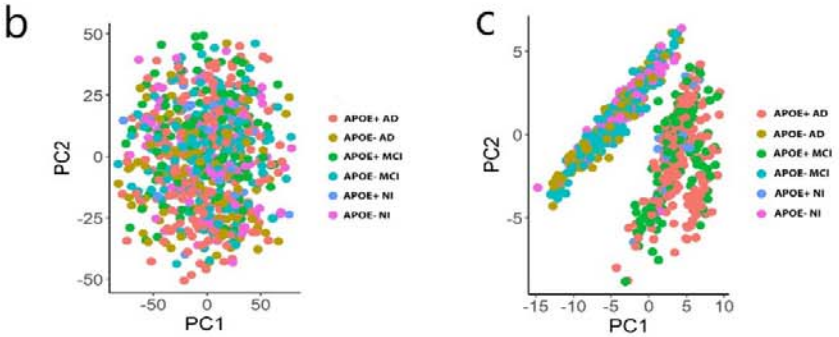
We first assessed if APOE4 carriers with AD ($APOE4^+$ AD) had a unique proteome signature relative to APOE4 non-carriers with AD ($APOE4^-$ AD) using CART. Two groups of proteins had a strong predictive performance (sensitivity >0.75). The first group was characterized by seven proteins that independently predicted $APOE4^+$ AD and $APOE4^-$ AD (sensitivity, specificity, PPV, NPV, and AUC = 1.0 for each of the seven) and included CCL25, CHCHD7, LRRN1, Otulin, S100A13, SPC25, and TBCA (independent proteins; Figure 1a; Supplementary Table 3). The second group was characterized by 50 proteins that interacted together to predict the groups (interactive proteins; Supplementary Table 4) and showed strong CART performance metrics (sensitivity = 0.99, specificity = 0.74, PPV = 0.86, NPV = 0.98, AUC = 0.92).

Next, we examined whether this proteome signature was unique to $APOE4^+$ AD. To do this, we first tested whether our CART models and protein signatures were specific to $APOE4^+$ AD or whether they were generalizable to APOE4 carriers with MCI or no impairment ($APOE4^+$ MCI or $APOE4^+$ NI). Models using either the independent or interactive proteins performed poorly and lost their predictive power (Table 1). CART models showed that APOE4+ MCI and APOE4+ NI were similarly indistinguishable (Table 1). We then tested whether proteome signatures were able to differentiate between APOE4 carriers and non-carriers independent of cognitive status. CART models comparing $APOE4^+$ MCI and $APOE4^-$ MCI as well as $APOE4^+$ NI to $APOE4^-$ NI had a high level of performance (Table 1). We performed a PCA to further validate this finding, showing no group separation when looking at all 6,082 identified CSF proteins (Figure 1b) but clear group separation delineated by APOE4 status based on our 57-protein proteome signature (Figure 1c). This was again confirmed by heat map based on measured protein levels (Figure 1d). These findings indicated that the CSF proteome signature was indeed unique to APOE4 carriers irrespective of cognitive status.

Table 1. Performance metrics of CART models testing the uniqueness of the identified proteome signature to APOE4 carriers.

	Sensitivity	**Specificity**	**PPV**	**NPV**	**AUC**
APOE4^{+} AD vs. APOE4^{+} MCI					
Independent Proteins	0.67	0.44	0.78	0.28	0.47
Interactive Proteins	0.67	0.38	0.69	0.37	0.53
APOE4^{+} AD vs. APOE4^{+} no impairment					
Independent Proteins	0.84	0.15	0.47	0.50	0.54
Interactive Proteins	0.88	0.28	0.65	0.50	0.49
APOE4^{+} MCI vs. APOE4^{+} no impairment					
Independent Proteins	0.83	0.13	0.62	0.34	0.53
Interactive Proteins	0.85	0.22	0.74	0.34	0.47
APOE4^{+} MCI vs. APOE4^{-} MCI					
Independent Proteins	1.00	1.00	1.00	1.00	1.00
Interactive Proteins	0.91	0.94	0.93	0.92	0.95
APOE4^{+} no impairment vs. APOE4^{-} no impairment					
Independent Proteins	0.78	0.97	0.90	0.93	0.91
Interactive Proteins	0.73	0.94	0.75	0.93	0.84


a
CCL25 RFU
CHCHD7 RFU
Otulin RFU
S100A13 RFU
SPC25 RFU
TBCA RFU
APOE+ AD
APOE- AD
APOE+ MCI
APOE- MCI
APOE+ NI
APOE- NI


b
c
PC1
PC2
APOE+ AD
APOE- AD
APOE+ MCI
APOE- MCI
APOE+ NI
APOE- NI

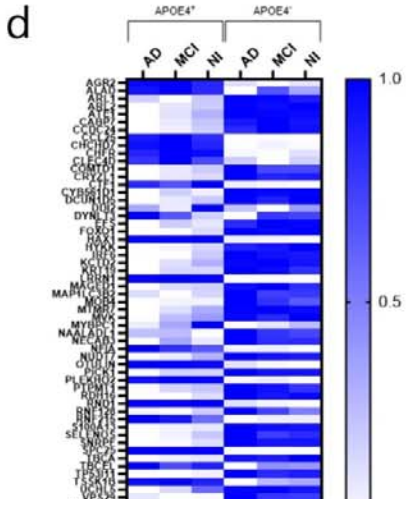

d
APOE4+
APOE4-
AD
MCI
NI
1.0
0.5

Figure 1. Identification and confirmation of a unique CSF proteome among APOE4 carriers irrespective of cognitive status. (**a**) Box plots showing relative fluorescent units (RFU) of the seven independent protein predictors (CCL25, CHCHD7, LRRN1, Otulin, S100A13, SPC25, and TBCA) showing differences between APOE4 carriers and non-carriers. (**b**) Principal component analysis (PCA) of all 7,000 CSF protein measured across the six APOE4 carrier and non-carrier groups before machine learning identification of the unique proteome signature. (**c**) Principal component analysis (PCA) of the 57 proteins (independent and interactive) identified by machine learning as being unique to APOE4 carriers relative to non-carriers irrespective of cognitive status. (**d**) Heat map of the relative levels of 57 proteins (independent and interactive) identified by machine learning as being unique to APOE4 carriers relative to non-carriers irrespective of cognitive status. Abbreviations: AD: Alzheimer's disease; MCI: mild cognitive impairment; NI: no impairment.

APOE4 carriers are more likely to progress to cognitive impairment and AD

To identify if the APOE4$^+$ CSF proteome signature is pathogenic or benign, we examined whether these patients were more likely to progress to clinical cognitive impairment over time. APOE4$^+$ NI or MCI patients were more likely to progress to either MCI or AD relative to APOE4$^-$ patients (14% vs. 5% progressed, respectively; $X^2 = 7.14$, df = 1, $p = 0.008$).

Characterization of the APOE4$^+$ proteome signature

To identify potential mechanisms underlying APOE4$^+$ vulnerability to progression, we performed enrichment analyses on our CSF proteome signatures. APOE4$^+$ independent proteins (n=7) were involved in T cell development, synapse assembly, innate immune responses, and cell division (Supplementary Table 3). A Steiner Forest Network PPI analysis of APOE4$^+$ interactive proteins (n=50) showed 41 of these formed a network (Figure 2a). Pathway enrichment analysis indicated biological and molecular involvement in catabolism, protein ubiquitination, mitosis, DNA damage response, ATP binding, and oxidative and cellular stress responses (adj. $p < 0.01$; Supplementary Tables 5 and 6). Enriched signaling pathways included regulation of mitosis, immune system, apoptosis, inflammation and RNA and DNA regulation (adj. $p < 0.001$; Supplementary Table 7).

Interestingly, our APOE4$^+$ proteome signature was enriched in the caudate and cerebral cortex, with less hippocampal involvement (Figure 2b). Single cell RNA-sequencing of brain tissue also found enrichment for both independent and interactive proteins in endothelial cells (Figure 2c), suggestive of blood-brain-barrier (BBB) dysregulation. Astrocytes and oligodendrocytes were also enriched for APOE4$^+$ proteins, further indicative

of immune and inflammatory pathway involvement. This immune dysregulation was also mirrored in the periphery. Single cell RNA-sequencing enrichment analysis for immune cell subtypes indicated enrichment for T cells, macrophages, and B cells for our APOE4+ proteome signature (Figure 2d).

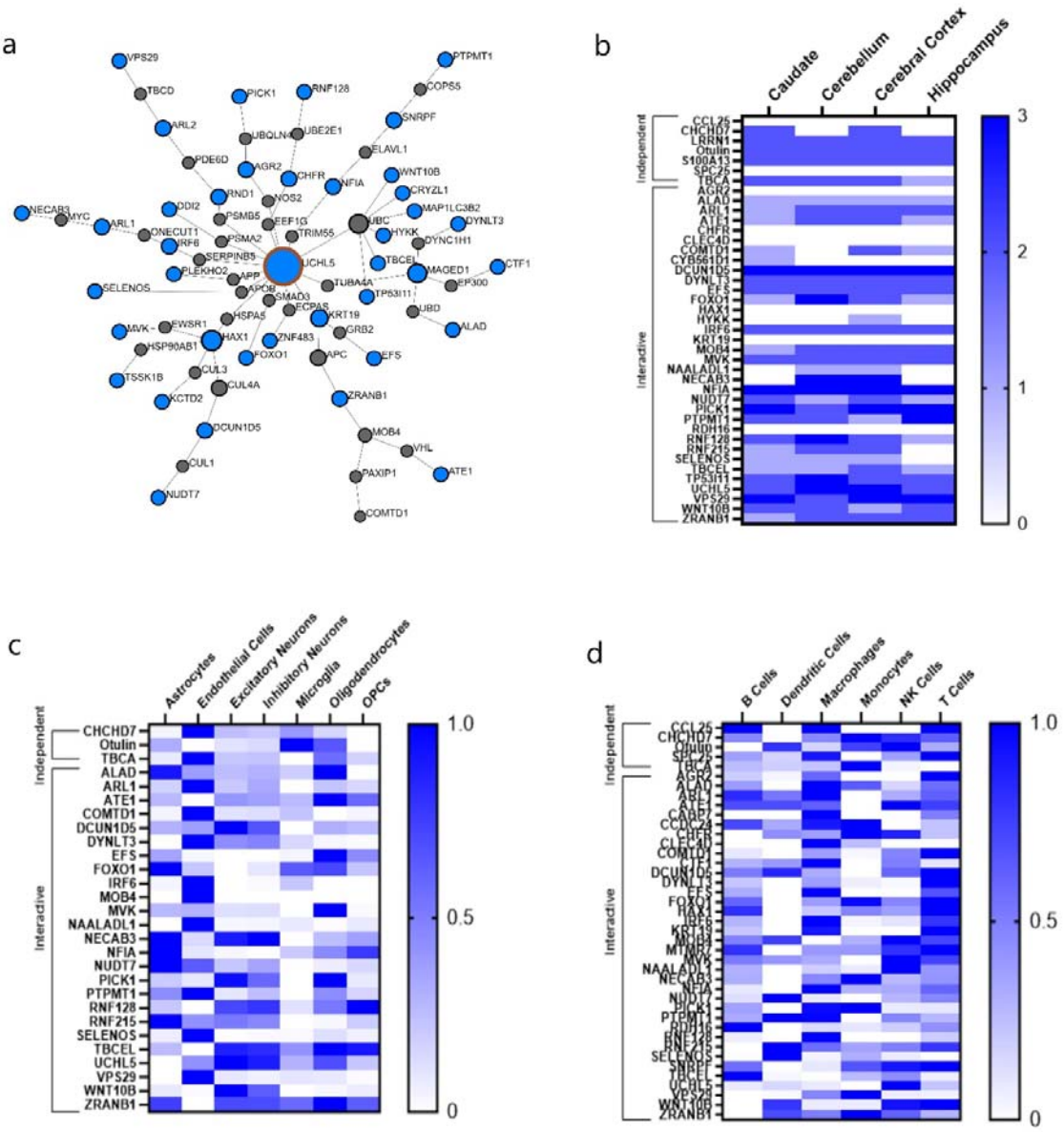


Figure 2. Characterization of the APOE4 carrier proteome signature identified by supervised machine learning models. (**a**) Steiner Forest Network protein-protein interaction analysis showing functional connections between the independent and interactive proteins (total n=41/50) in the APOE4+ proteome signature. Blue nodes indicate proteins identified. Grey nodes indicate inferred functionally connected proteins. (**b**) Enrichment across brain regions for the independent (n=7/7) and interactive (n=34/50) proteins within the APOE4 carrier proteome signature based on confirmed immunohistochemistry microarrays from normal tissue in the Human Protein Atlas. (**c**) Enrichment across brain cell subtypes for the

independent (n=3/7) and interactive (n=25/50) proteins within the APOE4 carrier proteome signature based on single cell RNA sequencing data from normal tissue in the Human Protein Atlas. (**d**) Enrichment across immune cell subtypes for the independent (n=5/7) and interactive (n=36/50) proteins within the APOE4$^+$ proteome signature based on single cell RNA sequencing data from normal tissue in the Human Protein Atlas.

Discussion

We identified an APOE4$^+$ specific CSF proteome signature of 57 proteins that was independent of cognitive status and was associated with an increased risk of progression to cognitive impairment (MCI or AD) over time. While unique CSF and plasma proteins have previously been associated with APOE4$^+$ AD patients [4-6], we expand on the existing literature by reporting on a wider proteome and identifying that this signature is independent of AD-specific brain pathology and cognitive status.

The APOE4$^+$ proteome signature was enriched in the caudate, cortex, and endothelial cells and was associated with a strong immune and inflammatory phenotype across the brain and periphery. These likely underly the vulnerability to cognitive decline and AD. This idea is strengthened by previous research showing that cognitively healthy APOE4 carriers have volumetric and functional changes in the caudate and cortex [12-14] and signs of BBB breakdown predictive of future cognitive decline [15]. APOE4 has also been linked to AD via its immunomodulatory functions [16] and a unique plasma complement cascade [5] relative to non-carriers. Importantly, our findings suggest that these changes are occurring independent of neurodegeneration and are not a consequence of tau and amyloid-β deposition, in line with other findings [15].

Limitations

Our study has limitations. First, the lower n and machine learning precluded examining differences between APOE4 homozygotes and heterozygotes. Although there may be differences this is unlikely because our models performed strongly irrespective of the relative distribution of APOE4 genotypes between groups (Supplementary Table 1). Similarly, we were unable to examine these signatures in males versus females. Our use of both sexes and strong performance metrics, however, highlights that the APOE4$^+$ proteome signature may be sex independent.

Conclusions

In our study, we found a CSF proteome signature that was unique to APOE4 carriers irrespective of cognitive status, AD pathology, and neurodegeneration and is associated with an increased risk of future cognitive impairment. In addition to being implicated in the BBB these proteins have strong immune enrichment, suggesting that immunomodulation or improving BBB function may be a precision medicine treatment strategy for APOE4 carriers at risk of AD.

Author Contributions: Dr. Finney gained full access to the ADNI cohort data used in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Concept and Design: Shvetcov, Finney

Acquisition of data: Finney

Statistical analysis: Shvetcov, Finney

Interpretation of data: Shvetcov, Thomson, Cho, Wilkins, Reed, Swerdlow, Brown, Finney

Drafting of the manuscript: Finney

Critical review of the manuscript for important intellectual content: Shvetcov, Thomson, Cho, Wilkins, Reed, Swerdlow, Brown, Finney

Supervision: Finney

Conflict of Interest Disclosures: Dr. Brown is a co-inventor of a licensed patent, owned by St. Vincent's Hospital, that pertains to the diagnostic and clinical use of GDF15 and receives royalties for its use. No other disclosures were reported.

Funding/Support: This study was not funded by any government, philanthropic, or commercial entity although the authors report receiving other funding during this study. Dr. Cho receives funding from the University of Sydney Horizon Fellowship program. Dr. Wilkins receives support from the Alzheimer's Association and the State of Kansas. Drs. Wilkins and Swerdlow receive are supported by funding from the National Institutes of Health (NIH; AG072973). Dr. Reed receives funding from the National Health and Medical Research Council (NHMRC), Medical Research Future Fund (MRFF), and philanthropic support from the Peter Tosi Family. Dr. Brown receives funding from the MRFF. Dr. Finney

receives funding from Dementia Australia and philanthropic support from the Neil & Norma Hill Foundation and the Leece Family.

Role of the Funder/Sponsor: The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Data Sharing Statement: Data for this study was provided by ADNI and is accessible at (https://ida.loni.usc.edu/). All data generated from our analyses is available in the Supplementary Tables. Code used for data analytics and supervised machine learning is available on Dr. Shvetcov's Github at https://github.com/Art83.

Acknowledgements: The authors are grateful to ADNI for providing the data. The authors are thank the Neurogenomics and Informatics Centre at Washington University (https://neurogenomics.wustl.edu/) and the Cruchaga Lab at Washington University School of Medicine for generating the CSF SomaScan 7k assay proteome data (https://cruchagalab.wustl.edu/) for the ADNI cohort. Finally, the authors thank all the patients and their families for their involvement in ADNI.

References

1. Safieh M, Korczyn AD, Michaelson DM. ApoE4: An emerging therapeutic target for Alzhiemer's disease. *BMC Medicine*. 2019;17
2. Emrani S, Arain HA, DeMarshall C, Nuriel T. APOE4 is associated with cognitive nad pathological heterogeneity in patients with Alzheimer's disease: A systematic review. *Alzheimer's Research & Therapy*. 2020;12(141)
3. Jansen WJ, Ossenkoppele R, Knol DL, et al. Prevalence of cerebral amyloid pathology in persons without dementia. *JAMA*. 2015;313(19):1924-1938.
4. Konijnenberg E, Tijms BM, Gobom J, et al. APOE ε4 genotype-dependent cerebrospinal fluid proteomic signatures in Alzheimer's disease. *Alzheimer's Research & Therapy*. 2020;12
5. Tao Q, Zhang C, Mercier G, et al. Identification of an APOE ε4-specific blood-based molecular pathway for Alzhiemer's disease risk. *Alzheimer's & Dementia*. 2023;15(4)
6. Ayton S, Faux NG, Bush AI. Association of cerebrospinal fluid ferritin level with preclinical cognitive decline in APOE-ε4 carriers. *JAMA Neurology*. 2017;74(1):122-125.
7. Petersen RC, Aisen PS, Beckett LA, et al. Alzheimer's Disease Neuroimaging Initiative (ADNI): Clinical characterization. *Neurology*. 2009;74(3):201-209.
8. Cruchaga C, Western D, Timsina J, et al. Proteogenomic analysis of human cerebrospinal fluid uses neurologically relevant regulation and informs causal proteins for Alzheimer's disease. *Research Square* 2023;
9. Zhou G, Soufan O, Ewald J, Hancock REW, Basu N, Xia J. NetworkAnalyst 3.0: A visual analytics platform for comprehensive gene expression profiling and meta-analysis. *Nucleic Acids Research*. 2019;47:W234-241.
10. Sjostedt E, Zhong W, Fagerberg L, et al. An atlas of the protein-coding genes in the human, pig, and mouse brain. *Science*. 2020;367
11. Karlsson M, Zhang C, Mear L, et al. A single-cell type transcriptomics map of human tissues. *Science Advances*. 2021;7(31)
12. Cacciaglia R, Molinuevo JL, Falcon C, et al. Effects of APOE-ε4 allele load on brain morphology in a cohort of middle-aged healthy individuals with enriched genetic risk for Alzheimer's disease. *Alzheimer's & Dementia*. 2018;14(7):902-912.
13. Matura S, Prvulovic D, Jurcoane A, et al. Differential effects of the ApoE4 genotype on brain structure and function. *NeuroImage*. 2014;89:81-91.
14. Novellino F, Loipez ME, Vaccaro MG, Miguel Y, Delgado ML, Maestu F. Association between hippocampus, thalamus, and caudate in mild cognitive impairment APOEε4 carriers: A structural covariance MRI study. *Frontiers in Neurology*. 2019;10
15. Montagne A, Nationa DA, Sagare AP, et al. APOE4 leads to blood-brain barrier dysfunction predicting cognitive decline. *Nature*. 2020;581:71-76.
16. Shi Y, Holtzman DM. Interplay between innate immunity and Alzheimer disease: APOE and TREM2 in the spotlight. *Nature Reviews Immunology*. 2018;18:759-772.