



# Mendelian randomization identifies proteins involved in neurodegenerative diseases

 Lazaros Belbasis, Sam Morris,  Cornelia van Duijn, Derrick Bennett and Robin Walters

Proteins are involved in multiple biological functions. High-throughput technologies have allowed the measurement of thousands of proteins in population biobanks. In this study, we aimed to identify proteins related to Alzheimer's disease, Parkinson's disease, multiple sclerosis and amyotrophic lateral sclerosis by leveraging large-scale genetic and proteomic data.

We performed a two-sample *cis* Mendelian randomization study by selecting instrumental variables for the abundance of >2700 proteins measured by either Olink or SomaScan platforms in plasma from the UK Biobank and the deCODE Health Study. We also used the latest publicly available genome-wide association studies for the neurodegenerative diseases of interest. The potentially causal effect of proteins on neurodegenerative diseases was estimated based on the Wald ratio.

We tested 13 377 protein–disease associations, identifying 169 associations that were statistically significant (5% false discovery rate). Evidence of co-localization between plasma protein abundance and disease risk (posterior probability > 0.80) was identified for 61 protein–disease pairs, leading to 50 unique protein–disease associations. Notably, 23 of 50 protein–disease associations corresponded to genetic loci not previously reported by genome-wide association studies. The two-sample Mendelian randomization and co-localization analysis also showed that APOE abundance in plasma was associated with three subcortical volumes (hippocampus, amygdala and nucleus accumbens) and white matter hyper-intensities, whereas PILRA and PILRB abundance in plasma was associated with caudate nucleus volume.

Our study provided a comprehensive assessment of the effect of the human proteome that is currently measurable through two different platforms on neurodegenerative diseases. The newly associated proteins indicated the involvement of complement (C1S and C1R), microglia (SIRPA, SIGLEC9 and PRSS8) and lysosomes (CLN5) in Alzheimer's disease; the interleukin-6 pathway (CTF1) in Parkinson's disease; lysosomes (TPP1), blood–brain barrier integrity (MFAP2) and astrocytes (TNFSF13) in amyotrophic lateral sclerosis; and blood–brain barrier integrity (VEGFB), oligodendrocytes (PARP1), node of Ranvier and dorsal root ganglion (NCS1, FLRT3 and CDH15) and the innate immune system (CR1, AHSG and WARS) in multiple sclerosis. Our study demonstrates how harnessing large-scale genomic and proteomic data can yield new insights into the role of the plasma proteome in the pathogenesis of neurodegenerative diseases.

Nuffield Department of Population Health, University of Oxford, Oxford OX3 7LF, UK

Correspondence to: Lazaros Belbasis  
Big Data Institute, University of Oxford  
Old Road Campus, Headington, Oxford OX3 7LF, UK  
E-mail: lazarus.belbasis@ndph.ox.ac.uk

**Keywords:** Alzheimer's disease; amyotrophic lateral sclerosis; Mendelian randomization; multiple sclerosis; Parkinson's disease; proteomics

## Introduction

Neurological diseases are the leading cause of disability and the second leading cause of death worldwide.<sup>1</sup> Neurodegenerative diseases constitute a distinct group of neurological diseases, which are characterized by progressive neuronal loss and formation of distinct pathological changes in the brain.<sup>2</sup> During the last three decades, there has been a substantial increase in the number of people living with neurodegenerative diseases such as Alzheimer’s disease (AD), Parkinson’s disease (PD), amyotrophic lateral sclerosis (ALS) and multiple sclerosis (MS).<sup>3–6</sup> Although MS is mainly an autoimmune disorder, early signs of neurodegeneration are evident in the disease course as a reactive response to the autoimmune process.<sup>7</sup> Genome-wide association studies (GWAS) have identified molecular pathways leading to neurodegenerative diseases and have increased our knowledge on causal pathways involved in these diseases.<sup>8–11</sup>

Proteins play a key role in a range of biological processes, hence their dysregulation can lead to the development of diseases, and even minor modulation of their levels or function can modify disease risk. They represent a major source of biomarkers for the diagnosis or prediction of disease and can also be crucial to improving our understanding of the pathogenesis of diseases.<sup>12</sup> About 75% of US Food and Drug Administration-approved medications were targeted at human proteins.<sup>13,14</sup> Therefore, by combining large-scale genomic and proteomic profiling, there is a potential to identify disease-causing pathways, uncover new drug targets, highlight new therapeutic indications and identify clinically relevant biomarkers.<sup>15,16</sup>

Recent technological advances have allowed the measurement of thousands of proteins in large population-based studies. To date, two different high-throughput techniques to measure the abundance of multiple proteins have been used in large population samples: an antibody-based proximity-extension assay (Olink platform) and an aptamer affinity-based assay (SomaScan platform).<sup>16</sup> GWAS of plasma protein abundance have identified protein quantitative trait loci (pQTLs), which can be used to examine the potentially causal effect of proteins on human diseases and traits using the Mendelian

randomization (MR) framework.<sup>12</sup> MR is an instrumental variable (IV) approach, which can be used to accelerate the discovery of biomarkers and the drug development pipeline.<sup>17</sup> MR studies have examined the potential role of proteins in the development of neurological diseases, mainly by adopting a transcriptome-wide MR approach, which uses expression quantitative trait loci (eQTLs) as IVs<sup>18–23</sup>; however, the value of such analyses is limited by the fact that eQTLs frequently do not reflect protein abundance accurately.<sup>24</sup>

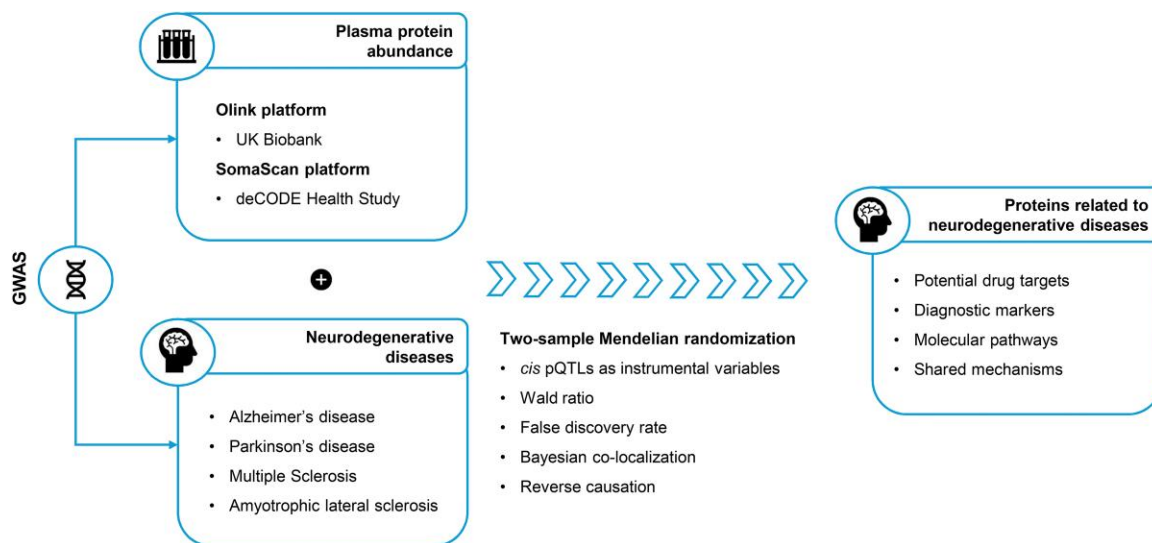
In the present study, we harnessed summary-level genetic data from two large proteo-genomic studies and from the largest GWAS for major neurodegenerative diseases, to identify proteins whose abundance in plasma is associated with these diseases (Fig. 1). We followed a two-sample cis MR approach, and we minimized the risk of confounding by linkage disequilibrium (LD) by performing a co-localization analysis.<sup>25</sup> We complemented our analysis by exploring the potential effects of these proteins on multiple brain imaging phenotypes.

## Materials and methods

### Data sources

#### Genome-side association study of human plasma proteome

We used summary-level data from the two largest proteogenomics studies conducted in populations of European ancestry using either Olink or SomaScan platforms, which were identified through a publicly available catalogue of proteogenomics studies (last updated on 29 January 2024).<sup>16</sup> The association between circulating protein levels and genetic variants was assessed in 35 571 participants in the UK Biobank using the Olink Explore 3072 platform, measuring 2941 protein analytes, capturing 2923 unique proteins,<sup>26</sup> and in 35 559 participants in the Icelandic Cancer Project and deCODE Health Study using the SomaScan version 4 platform, measuring 4907 aptamers, capturing 4719 proteins.<sup>27</sup>



**Figure 1 Schematic representation of the study.** Cis protein quantitative trait loci (pQTLs) from two large proteogenomic studies were used as instrumental variables for protein abundance in plasma. These were then integrated with genome-wide association studies (GWAS) for four neurodegenerative diseases. The association of plasma proteins with neurodegenerative diseases was assessed by estimating the Wald ratio within a two-sample Mendelian randomization framework. The analysis was complemented by Bayesian co-localization. The potentially causal proteins for neurodegenerative diseases might serve as drug targets and diagnostic markers or help to elucidate molecular pathways and shared biological mechanisms. The figure was created using Microsoft PowerPoint.

We do not assume equivalence of the two platforms, but we included pQTLs identified through both platforms to increase the completeness of our analysis.

### Genome-wide association study of neurological diseases

We searched GWAS Catalog for published GWAS on neurological diseases (Supplementary Table 1).<sup>28</sup> We selected the largest publicly available GWAS in population of European ancestry for AD (111 326 cases, 75 genome-wide significant variants),<sup>9</sup> PD (33 674 cases, 90 genome-wide significant variants),<sup>8</sup> MS (47 429 cases, 200 genome-wide significant variants)<sup>11</sup> and ALS (27 205 cases, 15 genome-wide significant variants).<sup>10</sup>

### Selection of instrumental variables for protein abundance

pQTLs are genetic variants with an effect on protein expression, and they can be either *cis* or *trans* based on their proximity to the gene encoding the protein of interest.<sup>29</sup> *trans* pQTLs map to genes that do not code directly for the targeted proteins or that correspond to intergenic regions, and it is difficult to distinguish between detected effects owing to vertical and horizontal pleiotropy.<sup>12,30</sup> For this reason, we restricted our MR analysis to *cis* pQTLs. Only autosomal genetic variants were included in the analyses, because summary-level data for the X chromosome are not available in some of the GWAS for neurological diseases.

We retrieved statistically significant *cis* pQTLs from each of the proteogenomics studies, applying the same level of statistical significance as used in those studies ( $P < 3.40 \times 10^{-11}$  for the study using the Olink platform, and  $P < 1.80 \times 10^{-9}$  for the study using the SomaScan platform). For the Olink platform, the *cis* region was defined as a distance of 1 Mb upstream or downstream from the end or start, respectively, of the gene encoding the protein of interest.<sup>26</sup> For the SomaScan platform, the *cis* region was defined as a distance of 1 Mb upstream or downstream from the transcription start site of the gene encoding the protein of interest.<sup>27</sup>

We used the following criteria to filter the list of statistically significant *cis* pQTLs in each study:

- (i) Owing to the complex LD structure of single nucleotide polymorphisms within the human major histocompatibility complex region, single nucleotide polymorphisms and proteins encoded by genes within the major histocompatibility complex region (Chromosome 6: from 26 to 34 Mb) were excluded.
- (ii) To reduce the risk for weak instrument bias, we calculated the *F*-statistic for each single nucleotide polymorphism, and we excluded genetic instruments with an *F*-statistic of  $< 10$ .<sup>31</sup>
- (iii) We obtained the genetic variants that were also tested in the GWAS for neurological diseases.
- (iv) For each protein, we selected only the *cis* pQTL with the lowest *P*-value, which we refer as the 'lead variant' or 'lead *cis* pQTL'.

There is substantial heterogeneity between Olink and Soma Scan platforms, which means that they might measure different proteoforms of the same protein.<sup>32</sup> For this reason, when an assay in both platforms targeted the same protein, we considered the lead *cis* pQTL from both platforms. When multiple assays in the same platform targeted the same protein (as defined by UniProt ID), we included only the instrument with the lowest *P*-value.

### Data harmonization

All GWAS summary statistics were lifted over to genomic build 38.<sup>33</sup> We followed the recommended harmonization framework for two-sample MR analyses.<sup>34,35</sup> Ambiguous palindromic single

nucleotide polymorphisms with an allele frequency between 0.42 and 0.58 were excluded to avoid potential allele mismatch across different GWAS.<sup>34</sup> Data harmonization was implemented using the *TwoSampleMR* package.<sup>36,37</sup>

### Statistical analysis

#### Association of protein abundance with neurological diseases

The Wald ratio, which is defined as the ratio of the gene-outcome effect divided by the gene-exposure effect, was calculated for all the protein–disease associations.<sup>38</sup> To identify the statistically significant associations, a multiplicity correction was applied using the Benjamini–Hochberg method.<sup>39</sup> Evidence of a statistically significant protein–disease association was based on 5% false discovery rate (FDR). MR analyses were performed using the *TwoSampleMR* package.<sup>36,37</sup> We prioritized proteins with a statistically significant association with a neurological disease for further analyses to: (i) assess reverse causality; and (ii) perform Bayesian co-localization.

#### Assessment of reverse causation

Reverse causation could be a potential explanation for positive findings in an MR analysis. We explored the potential for reverse causality by taking a bi-directional MR approach. We performed LD clumping to obtain approximately independent genetic variants to model the genetic liability to AD, PD, ALS and MS. Clumping was performed using the reference panel from 1000Genomes for population of European ancestry, setting a statistical significance threshold of  $P < 5 \times 10^{-8}$ , a genetic window of 1 Mb and an LD  $r^2 < 0.1\%$ . We used the *ld\_clump* function from the *ieugwas* package and PLINK version 1.90.<sup>40</sup> We derived four genetic instruments consisting of genome-wide significant genetic variants as reported in the relevant publications.<sup>8–11</sup> We used these genetic instruments to examine whether the genetically predicted liability to each one of the neurological diseases of interest was related to the proteins associated with each one of the diseases. We estimated the Wald ratio for each one of the genetic variants, and we combined them using a random-effects inverse-variance weighted model.<sup>41</sup> Evidence of statistically significant findings were based on 5% FDR.

#### Bayesian co-localization

Evidence of co-localization supports the validity of the IVs and strengthens the MR findings.<sup>25</sup> To assess potential confounding by LD, we examined whether the prioritized proteins share the genetic variant with the outcomes of interest by conducting a co-localization analysis assuming a single causal variant in each genetic locus.<sup>42</sup> We used the *coloc* package for the co-localization analysis. Variants within  $\pm 1$  Mb window around the *cis* pQTLs with the smallest *P*-value were included. We used a posterior probability of  $> 80\%$  as strong evidence of co-localization, and a posterior probability of  $> 60\%$  as moderate evidence of co-localization. However, we acknowledge that lack of co-localization does not invalidate the MR findings, because co-localization methods have a high false negative rate.<sup>13,43</sup> The GWAS by Bellenguez et al.<sup>9</sup> for AD does not provide the majority of the genetic variants in the APOE gene locus. For this reason, we repeated the co-localization analysis using the GWAS by Kunkle et al.<sup>44</sup> for the genetic loci located near APOE (i.e. APOE, APOC1 and NECTIN2).

**Association of mRNA abundance with neurological diseases**

GWAS of gene expression reported cis eQTLs that are genetic variants affecting the mRNA abundance.<sup>45</sup> The eQTLGen consortium examined eQTLs from blood-derived expression of 19 250 autosomal genes and reported at least one cis eQTL for 16 987 genes using a sample of 31 684 individuals.<sup>46</sup> The MetaBrain consortium provides cis eQTLs in five tissues [cortex (2683 individuals), cerebellum (492 individuals), basal ganglia (208 individuals), hippocampus (168 individuals) and spinal cord (108 individuals)].<sup>47</sup> For each one of the proteins measured through either Olink or SomaScan platforms, the relevant lead cis eQTL was selected as a genetic IV. We used the same statistical significance threshold as the GWAS on plasma and brain eQTLs to identify appropriate genetic instruments. We estimated the Wald ratio as the ratio of the genetic effect on disease risk divided by the genetic effect on mRNA abundance. A multiplicity correction was applied using the Benjamini–Hochberg method separately in plasma and brain tissues,<sup>39</sup> and statistically significant associations were assessed at 5% FDR.

**Association of protein abundance with brain imaging phenotypes**

The potentially causal effect of the prioritized proteins on brain imaging traits was examined using summary-level GWAS data for nine brain volumes, mean cortical thickness and surface, and white matter hyper-intensities.<sup>48–50</sup> The available brain volumes were intracranial volume,<sup>51</sup> hippocampal volume<sup>52</sup> and other subcortical structures volume<sup>53</sup> (nucleus accumbens, amygdala, brainstem, caudate nucleus, globus pallidus, putamen and thalamus). A

multiplicity correction was applied using the Benjamini–Hochberg method,<sup>39</sup> and statistically significant findings were assessed at 5% FDR. Additionally, for the statistically significant associations, we performed Bayesian co-localization analysis assuming a single causal variant and using the same specifications as described before.

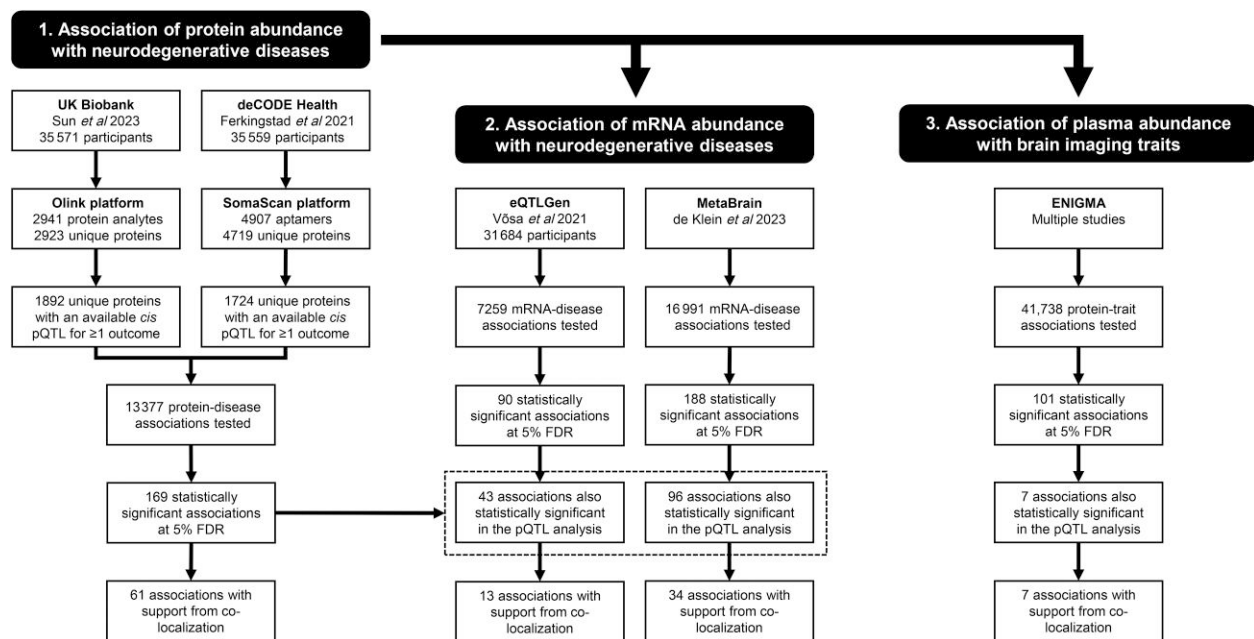
**Downstream analyses**

For the proteins that showed a statistically significant association with a neurodegenerative disease and had additional support from co-localization, we performed an enrichment analysis in Gene Ontology,<sup>54</sup> Reactome Pathway Database,<sup>55</sup> WikiPathways<sup>56</sup> and Kyoto Encyclopedia of Genes and Genomes.<sup>57</sup> A multiplicity correction was performed, and statistically significant enrichment was based on 5% FDR. We used the STRING database to perform the enrichment analysis.<sup>58</sup> We also examined whether the identified proteins show high specificity in particular tissues or single cells, and whether they belong to a particular gene expression cluster using the Human Protein Atlas.<sup>59,60</sup>

**Results**

**Association of plasma protein abundance with neurodegenerative diseases**

An overview of the study design is presented in Fig. 2. To evaluate systematically the evidence for a causal effect of 2738 proteins on four neurological diseases (AD, PD, ALS and MS), we undertook a proteome-wide two-sample MR. All the selected IVs had an F-statistic > 10, minimizing the influence of weak instrument bias



**Figure 2 Overview of the study design and main results.** The study consists of three key steps. First, the lead cis protein quantitative trait loci (pQTLs) from two proteogenomic genome-wide association studies (GWAS) using either the Olink or SomaScan platforms in European populations were used as instrumental variables (IVs) to perform a two-sample Mendelian randomization (MR) analysis for Alzheimer’s disease (AD), Parkinson’s disease (PD), multiple sclerosis (MS) and amyotrophic lateral sclerosis (ALS). A total of 13 377 protein–disease associations were tested, with 169 statistically significant associations at 5% false discovery rate (FDR) and 61 associations supported by co-localization. Second, MR analysis was repeated using plasma and brain cis eQTLs as IVs from the eQTLGen and MetaBrain consortia, yielding 43 (eQTLGen) and 96 (MetaBrain) overlapping associations with the pQTL analysis, with 13 and 34 supported by co-localization, respectively. Third, the association of plasma protein abundance with 12 brain-imaging traits was assessed using data from the ENIGMA consortium, identifying 101 statistically significant associations at 5% FDR, of which 7 were supported by co-localization.

**Table 1 Summary of the proteome-wide two-sample Mendelian randomization analysis on neurological diseases**

Disease	Number of associations tested	Number of statistically significant associations		Inflation factor <sup>b</sup>	Co-localization <sup>c</sup>
		P < 0.05	5% FDR <sup>a</sup>		
Alzheimer's disease	3610	420 (11.6%)	78 (2.2%)	1.36	27
Parkinson's disease	2757	204 (7.4%)	10 (0.4%)	1.07	2
Amyotrophic lateral sclerosis	3518	278 (7.9%)	17 (0.5%)	1.28	3
Multiple sclerosis	3492	377 (10.8%)	64 (1.8%)	1.41	29
Total	13 377	1279 (9.6%)	169 (1.3%)	1.28	61

<sup>a</sup>False discovery rate based on Benjamini–Hochberg correction, which corresponds to  $P < 6.3 \times 10^{-4}$ .

<sup>b</sup>Inflation factor estimated after removal of the statistically significant associations at 5% false discovery rate (FDR).

<sup>c</sup>Posterior probability >80%.

on the MR estimates. Overall, 13 377 protein–disease associations were tested (50.5% using *cis* pQTLs derived using Olink platform measurements and 49.5% using *cis* pQTLs from the SomaScan platform; [Table 1](#) and [Supplementary Table 2](#)). We observed 1279 (9.6%) nominally significant protein–disease associations at  $P < 0.05$ , constituting a substantial excess in comparison to the number expected under the null. Of these, 169 protein–disease associations (1.3%) remained statistically significant at 5% FDR, corresponding to  $P < 6.3 \times 10^{-4}$  ([Tables 2](#) and [3](#) and [Figs 3](#) and [4](#)). Even after exclusion of these associations, the remaining associations displayed substantial inflation in comparison to the null ([Table 1](#)).

In our study, 120 unique proteins were associated with at least one neurodegenerative disease. Ten of them were associated with more than one neurodegenerative disease ([Supplementary Figs 1](#) and [2](#)). Also, 59 unique proteins were associated with AD, and 10 of them (17%) showed an association with AD and an additional neurological disease (CTF1, NSF and PRSS53 with PD; SIGLEC9 with ALS; CR1, CTSH, PARP1 and PVR with MS; LRRC37A2 with both PD and ALS; and IDUA with PD, ALS and MS). Two proteins (LMAN2 and INHBC) showed an association with both MS and ALS. However, we should highlight that the proteoform of IDUA measured through the Olink platform (OID21468) was associated with both AD and MS, whereas the proteoform of IDUA measured through the SomaScan platform (3169\_70) was associated with both ALS and PD. The proteoform of CTSH measured through the Olink platform (OID20113) was associated with MS, and the proteoform of CTSH measured through the SomaScan platform (8465\_52) was associated with AD.

Associations between proteins and phenotypes in the MR framework might reflect causality, but potential alternative explanations are reverse causality, confounding by LD or horizontal pleiotropy.<sup>12</sup> We evaluate each of these explanations below.

### Cross-platform comparison of *cis* protein quantitative trait locus associations

We hypothesize that a consistent causal effect in both Olink and SomaScan platforms strengthens confidence in the instruments and in the robustness of the protein–disease associations, by showing that the inferred effect on disease risk is independent of the platform used for protein abundance measurement and pQTL identification. Indeed, 34 protein–disease associations were statistically significant across both platforms, but 5 of the 34 associations were not directionally consistent (BCAM, HDGF and PILRA for AD, and CD58 and KLRB1 for MS). Fourteen protein–disease associations were statistically significant only in the Olink platform (ARHGAP45, EPHA1, GC, IDUA, TREML2 and PRSS8 for AD; SERPINE2 for PD; EGF, TPP1 and ATXN3 for ALS; and CTSH, IL7R, PARP1 and IDUA for MS).

We also observed that 11 protein–disease associations reached statistical significance only when the lead *cis* pQTL from the SomaScan platform was used but not when the lead *cis* pQTL from the Olink platform was used (CTSH, LGALS3, PARP1 and LILRB1 for AD; IDUA for PD; CD200, INHBC, THSD1 and AHSG for MS; and IDUA and INHBC for ALS; [Supplementary Fig. 3](#)).

### Sensitivity analysis for Alzheimer's disease

The GWAS by Bellenguez *et al.*<sup>9</sup> used proxy cases for AD, where possible cases identified through self-reported family history of dementia were included; these contributed 42% of AD cases, with the balance being physician-diagnosed cases of AD.<sup>9</sup> As a sensitivity analysis, we used the GWAS by Kunkle *et al.*<sup>44</sup> to validate that the proteins associated with AD are specific for AD, because it is the latest GWAS for AD that does not include proxy cases ([Supplementary Table 3](#)).<sup>44</sup> Out of 78 protein–AD associations, 23 (28%) were associated with AD in the MR analysis using the GWAS by Kunkle *et al.*<sup>44</sup> at 5% FDR ([Supplementary Fig. 4](#)). All the validated associations were directionally consistent across the two GWAS for AD. Among the non-validated associations, four associations (ARHGAP45, EPHX2, PARP1 and PLOD2) were not directionally consistent. The non-validation of some of the proteins could be explained either by the smaller sample size of the study by Kunkle *et al.*<sup>44</sup> (i.e. loss of statistical power) or by the misclassification of other causes of dementia as AD in the study by Bellenguez *et al.*<sup>9</sup>

### Sensitivity analysis for reverse causation

To explore the potential for reverse causation, we performed a bi-directional MR analysis, which examines whether the genetic liability to the outcome is associated with the exposure of interest. For this reason, we performed clumping to identify independent genome-wide significant variants ( $P < 5.00 \times 10^{-8}$ ) from the GWAS for AD, PD, ALS and MS, for use as IVs modelling the genetic liability to these diseases. Using a random-effects inverse-variance weighted model, we found that genetically predicted risk for AD was associated with plasma protein abundance of APOE and CEACAM19, genetically predicted risk for ALS was associated with DTX3, and genetically predicted risk for MS was associated with plasma protein abundance of CD58, JUND, CD200, PGLYRP1, TNFRSF1A, LMAN2, INHBC, DKKL1, AHSG and KLRB1 (5% FDR; [Supplementary Table 4](#)).

### Co-localization between plasma protein abundance and neurodegenerative diseases

To examine whether confounding by LD can explain the observed associations, we performed co-localization of the association

**Table 2** Olink proteins associated with risk of neurodegenerative diseases identified using two-sample Mendelian randomization (5% FDR) and supported by co-localization (posterior probability > 0.80)

Outcome	Assay ID <sup>a</sup>	Protein	rsID	EA/OA <sup>b</sup>	EAF	OR (95% CI) <sup>c</sup>	P-value	Novel <sup>d</sup>
AD	OID30727	APOE	rs8106813	G/A	0.494	0.61 (0.57–0.64)	3.37 × 10 <sup>-65</sup>	No
	OID30697	CR1	rs679515	T/C	0.174	1.26 (1.21–1.31)	5.65 × 10 <sup>-33</sup>	No
	OID20197	PILRB	rs1859788	G/A	0.681	1.07 (1.05–1.08)	1.16 × 10 <sup>-17</sup>	No
	OID21129	PILRA	rs1859788	G/A	0.681	1.08 (1.06–1.10)	1.16 × 10 <sup>-17</sup>	No
	OID20177	CD2AP	rs1385742	A/T	0.355	1.24 (1.17–1.32)	3.99 × 10 <sup>-13</sup>	No
	OID21159	GRN	rs5848	C/T	0.726	0.77 (0.72–0.83)	2.19 × 10 <sup>-12</sup>	No
	OID20763	PRSS8	rs889555	T/C	0.282	0.62 (0.53–0.72)	1.07 × 10 <sup>-9</sup>	Yes
	OID30541	BLNK	rs55769428	A/C	0.962	2.06 (1.54–2.74)	8.55 × 10 <sup>-7</sup>	No
	OID20809	IL34	rs4985556	C/A	0.878	0.95 (0.93–0.97)	5.65 × 10 <sup>-6</sup>	No
	OID30731	C1S	rs12146727	G/A	0.865	1.09 (1.05–1.12)	6.41 × 10 <sup>-6</sup>	Yes
	OID30753	C1R	rs10849546	G/A	0.865	1.13 (1.07–1.19)	8.55 × 10 <sup>-6</sup>	Yes
	OID20304	SIRPA	rs6136377	A/G	0.618	1.03 (1.02–1.04)	1.37 × 10 <sup>-5</sup>	Yes
	OID21472	CD33	rs2455069	G/A	0.427	1.03 (1.02–1.05)	1.83 × 10 <sup>-5</sup>	No
	OID21390	SIGLEC9	rs2075803	A/G	0.447	1.06 (1.03–1.10)	4.30 × 10 <sup>-5</sup>	Yes
	OID21205	ZBTB16	rs73000929	A/G	0.037	1.11 (1.05–1.18)	1.37 × 10 <sup>-4</sup>	Yes
	OID21307	MME	rs79837905	A/G	0.920	1.68 (1.26–2.24)	4.54 × 10 <sup>-4</sup>	No
	PD	OID31141	HIP1R	rs10847864	T/G	0.359	4.01 (2.66–6.05)	3.64 × 10 <sup>-11</sup>
OID20061		CTF1	rs11150601	A/G	0.628	6.84 (3.23–14.46)	4.92 × 10 <sup>-7</sup>	Yes
ALS	OID20750	TPP1	rs11827437	C/T	0.370	0.72 (0.62–0.83)	1.61 × 10 <sup>-5</sup>	Yes
	OID20733	TNFRSF13	rs3803800	A/G	0.211	0.84 (0.77–0.91)	3.59 × 10 <sup>-5</sup>	Yes
MS	OID20716	CD58	rs10801908	C/T	0.880	2.30 (1.88–2.81)	3.54 × 10 <sup>-16</sup>	No
	OID21449	CD5	rs4939491	G/A	0.609	0.37 (0.29–0.47)	4.28 × 10 <sup>-15</sup>	No
	OID30519	EVI5	rs11808092	C/A	0.745	0.38 (0.30–0.49)	4.70 × 10 <sup>-14</sup>	No
	OID21155	TNFRSF1A	rs1800693	T/C	0.597	0.12 (0.07–0.21)	1.02 × 10 <sup>-13</sup>	No
	OID20724	CD40	rs4810485	G/T	0.752	0.74 (0.68–0.80)	1.41 × 10 <sup>-12</sup>	No
	OID21313	DKKL1	rs2303759	T/G	0.748	0.91 (0.88–0.94)	1.89 × 10 <sup>-10</sup>	No
	OID21136	IL7R	rs6897932	C/T	0.729	1.13 (1.09–1.18)	1.84 × 10 <sup>-9</sup>	No
	OID20234	TYMP	rs131805	C/T	0.783	0.61 (0.52–0.73)	2.61 × 10 <sup>-8</sup>	No
	OID20496	SLAMF1	rs7535367	G/T	0.861	2.22 (1.57–3.13)	6.24 × 10 <sup>-6</sup>	No
	OID21420	PVALB	rs4821544	T/C	0.697	1.10 (1.05–1.15)	1.31 × 10 <sup>-5</sup>	Yes
	OID20868	TST	rs4821544	T/C	0.697	2.08 (1.50–2.89)	1.31 × 10 <sup>-5</sup>	Yes
	OID21011	PVR	rs2301274	T/C	0.759	0.90 (0.86–0.95)	2.46 × 10 <sup>-5</sup>	No
	OID30423	SPRED2	rs7569084	C/T	0.414	0.39 (0.25–0.61)	3.76 × 10 <sup>-5</sup>	No
	OID20500	PARP1	rs1433574	A/C	0.839	1.70 (1.30–2.23)	1.04 × 10 <sup>-4</sup>	Yes
	OID21005	CDH15	rs11646135	A/G	0.142	1.22 (1.10–1.35)	1.19 × 10 <sup>-4</sup>	Yes
	OID30697	CR1	rs679515	T/C	0.174	1.16 (1.07–1.26)	1.81 × 10 <sup>-4</sup>	Yes
	OID30554	VEGFB	rs660442	A/G	0.200	0.51 (0.36–0.73)	1.98 × 10 <sup>-4</sup>	Yes
OID21084	WARS	rs12882934	C/A	0.745	1.31 (1.14–1.52)	2.36 × 10 <sup>-4</sup>	Yes	
OID21458	NCS1	rs1054879	A/G	0.508	1.99 (1.36–2.92)	4.06 × 10 <sup>-4</sup>	Yes	

AD = Alzheimer’s disease; ALS = amyotrophic lateral sclerosis; CI = confidence interval; EA = effect allele; FDR = false discovery rate; MS = multiple sclerosis; OA = other allele; OR = odds ratio; PD = Parkinson’s disease.

<sup>a</sup>Assay identifier as provided by the Olink platform.

<sup>b</sup>The alleles have been orientated to reflect an increase in plasma protein abundance.

<sup>c</sup>The odds ratio correspond to risk for neurological disease per one normalized protein expression risk unit increase in plasma protein abundance measured through the Olink platform and estimated using the Wald ratio method.

<sup>d</sup>A genetic locus was considered novel if it was not reported by a genome-wide association study in European ancestry in the GWAS Catalog.

signals for protein and disease, assuming a single causal variant at each gene locus (Supplementary Table 5).<sup>25</sup> Among the 169 protein–disease pairs, 61 (36%) had strong evidence of co-localization (posterior probability > 0.80), and 35 additional protein–disease pairs (21%) showed moderate evidence of co-localization (0.60 < posterior probability < 0.80). CR1 co-localized with both AD and MS. Eleven protein–disease pairs showed strong evidence of co-localization in both Olink and SomaScan platforms (C1R, C1S, CR1, GRN, PILRA, SIGLEC9, SIRPA for AD; and CR1, DKKL1, TNFRSF1A and WARS for MS).

The 61 protein–disease pairs with strong evidence of co-localization correspond to 50 unique protein–disease associations and 49 unique proteins, which were prioritized for further analyses (Tables 2 and 3). Twenty-three of these 50 protein–disease pairs

(48%) represent previously unreported genetic associations, because they did not reach genome-wide significance in the disease GWAS, nor have they been otherwise previously reported in the GWAS Catalog (Tables 2 and 3).<sup>28</sup> Although PRSS8 has not been previously reported as a genetic locus involved in AD, the cis-pQTL for PRSS8 is a previously identified genome-wide significant association for AD, which is located within a different nearby gene (BCKDK) related to AD.<sup>9</sup>

### Association of plasma mRNA abundance with neurodegenerative diseases

We performed an additional two-sample MR analysis to investigate the association of the plasma abundance of mRNA encoding the proteins associated with risk of neurodegenerative diseases using

**Table 3 SomaScan proteins associated with risk of neurodegenerative diseases identified using two-sample Mendelian randomization (5% FDR) and supported by co-localization (posterior probability > 0.80)**

Outcome	Assay ID <sup>a</sup>	Protein	rsID	EA/OA <sup>b</sup>	EAF	OR (95% CI) <sup>c</sup>	P-value	Novel <sup>d</sup>
AD	19556_12	CR1	rs679515	T/C	0.183	1.28 (1.23–1.33)	$5.65 \times 10^{-33}$	No
	10816_150	PILRA	rs1859788	A/G	0.304	0.94 (0.93–0.96)	$1.16 \times 10^{-17}$	No
	4992_49	GRN	rs5848	C/T	0.744	0.74 (0.68–0.81)	$2.19 \times 10^{-12}$	No
	8687_26	TMEM106B	rs3173615	C/G	0.598	1.44 (1.25–1.66)	$6.93 \times 10^{-7}$	No
	8465_52	CTSH	rs2289702	C/T	0.894	1.04 (1.03–1.06)	$3.10 \times 10^{-6}$	No
	3285_23	C1R	rs12146727	G/A	0.881	1.11 (1.06–1.17)	$6.41 \times 10^{-6}$	Yes
	8840_61	C1S	rs12146727	G/A	0.881	1.07 (1.04–1.10)	$6.41 \times 10^{-6}$	Yes
	5430_66	SIRPA	rs6136376	A/G	0.640	1.03 (1.02–1.04)	$1.30 \times 10^{-5}$	Yes
	3007_7	SIGLEC9	rs2075803	A/G	0.479	1.03 (1.02–1.04)	$4.30 \times 10^{-5}$	Yes
	6923_1	PLOD2	rs148118826	G/A	0.997	0.59 (0.46–0.76)	$6.18 \times 10^{-5}$	Yes
	8874_53	CLN5	rs700363	G/A	0.910	0.70 (0.59–0.84)	$6.51 \times 10^{-5}$	Yes
	9294_45	MFAP2	rs3738814	A/G	0.584	0.83 (0.75–0.91)	$5.03 \times 10^{-5}$	Yes
	ALS	2654_19	TNFRSF1A	rs1800693	T/C	0.597	0.19 (0.13–0.30)	$1.02 \times 10^{-13}$
10346_5		STAT3	rs4796791	C/T	0.589	0.43 (0.35–0.54)	$1.17 \times 10^{-13}$	No
MS	16309_30	DKKL1	rs2288481	G/A	0.779	0.13 (0.07–0.26)	$9.33 \times 10^{-10}$	No
	4440_15	FCRL3	rs7528684	G/A	0.416	0.84 (0.80–0.89)	$8.93 \times 10^{-9}$	No
	9468_8	LMAN2	rs4131289	A/G	0.379	2.30 (1.61–3.30)	$4.85 \times 10^{-6}$	No
	3581_53	AHSG	rs4686790	T/G	0.707	0.80 (0.73–0.89)	$2.74 \times 10^{-5}$	Yes
	18172_71	ASF1A	rs4946366	T/C	0.164	0.49 (0.34–0.71)	$1.53 \times 10^{-4}$	Yes
	19556_12	CR1	rs679515	T/C	0.183	1.17 (1.08–1.27)	$1.81 \times 10^{-4}$	Yes
	9870_17	WARS	rs4905957	T/C	0.764	1.32 (1.14–1.53)	$2.21 \times 10^{-4}$	Yes
	13123_3	FLRT3	rs1932953	T/G	0.270	1.09 (1.04–1.14)	$2.60 \times 10^{-4}$	Yes

AD = Alzheimer's disease; ALS = amyotrophic lateral sclerosis; CI = confidence interval; EA = effect allele; FDR = false discovery rate; MS = multiple sclerosis; OA = other allele; OR = odds ratio.

<sup>a</sup>Assay identifier as provided by the SomaScan platform.

<sup>b</sup>The alleles have been orientated to reflect an increase in plasma protein abundance.

<sup>c</sup>The odds ratio corresponds to the risk for neurological disease per one standard deviation increase in plasma protein abundance measured through SomaScan platform and estimated using the Wald ratio method.

<sup>d</sup>A genetic locus was considered novel if it was not reported by a genome-wide association study in European ancestry in the GWAS Catalog.

the lead cis eQTL reported by the eQTLGen consortium.<sup>46</sup> We tested 7259 plasma mRNA abundance–disease associations, and 90 (1.2%) were statistically significant at 5% FDR, which corresponds to  $P < 6.2 \times 10^{-4}$ . Among these 90 associations, 43 (32.5%) were also statistically significant using the lead plasma cis pQTL (Supplementary Table 6 and Supplementary Fig. 5). When we compared the associations using plasma pQTLs and plasma eQTLs (Supplementary Fig. 6), we found inconsistencies in the direction of effect in six associations using the Olink platform (CEACAM19 and TREML2 with AD; TPP1 with ALS; and CD58, IL2RA and PTPRC with MS) and two associations using the SomaScan platform (PILRA with AD; and IL2RA with MS).

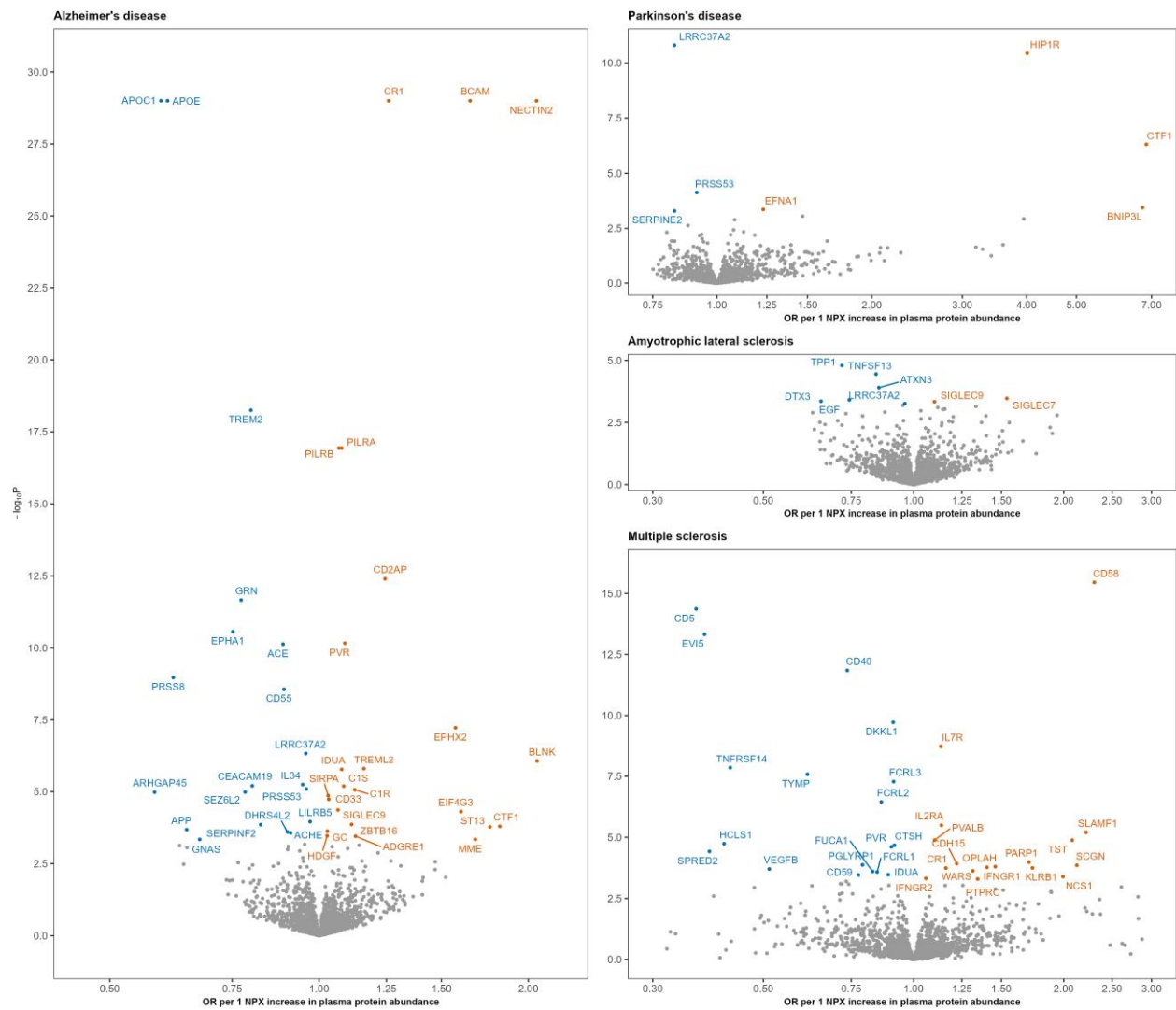
Thirteen of 43 associations (30%) also showed evidence of co-localization across the locus between mRNA abundance and disease risk (Supplementary Table 7). Ten proteins showed co-localization with a neurological disease using both plasma protein abundance and plasma mRNA abundance (SIRPA, CTSH and CD33 with AD; and ASF1A, FCRL3, VEGFB, TYMP, PVALB, LMAN2 and CD5 with MS). The remaining three proteins (ACE with AD; SIGLEC9 with ALS; and PLEK with MS) showed evidence of co-localization with plasma mRNA abundance but not with plasma protein abundance. There is strong evidence for distinct causal variants between ACE protein abundance and AD (posterior probability = 1.00), and moderate evidence for shared causal variants of PLEK protein abundance with MS and SIGLEC9 with ALS (posterior probability = 0.74 and 0.70, respectively). Discrepancies in the MR and co-localization analyses are expected because of the uncoupling of gene and protein expression, which is a frequently described phenomenon attributed to differential translation, protein

degradation, contextual confounders such as time and developmental state, or protein-level buffering.<sup>61,62</sup>

### Association of brain mRNA abundance with neurodegenerative diseases

As an additional step to aid interpretation of our findings, we used data on genetic associations with mRNA abundance in four brain regions (i.e. cortex, basal ganglia, hippocampus and cerebellum) and the spinal cord, as provided by the MetaBrain consortium.<sup>47</sup> For each protein measured through Olink or SomaScan platforms, we selected the lead cis eQTL per tissue as an IV, testing a total of 16 991 mRNA abundance–disease associations (8803 in cortex, 5881 in cerebellum, 1013 in basal ganglia, 710 in hippocampus and 584 in spinal cord), of which 188 (1.1%) were statistically significant at 5% FDR, which corresponds to  $P < 5.5 \times 10^{-4}$ . Of 188 associations, 96 (51%) were also statistically significant using the plasma lead cis pQTL (Supplementary Table 8). When we compared these associations with the associations using the lead plasma cis pQTL, we found that 74 (56 with Olink and 18 with SomaScan) were consistent in the direction of effect, whereas 22 (15 with Olink and 7 with SomaScan) were directionally inconsistent (Supplementary Figs 7–12).

Thirty-four of these 96 associations (35%) showed strong evidence of co-localization (Supplementary Table 9). Two proteins (CR1 with AD in basal ganglia, cortex and hippocampus; and GRN with AD in cerebellum, cortex and hippocampus) were found to have evidence for co-localization with disease risk from mRNA abundance in three brain regions, and a further six proteins (ACE, CD33, CTSH and SIRPA with AD in cerebellum and cortex;



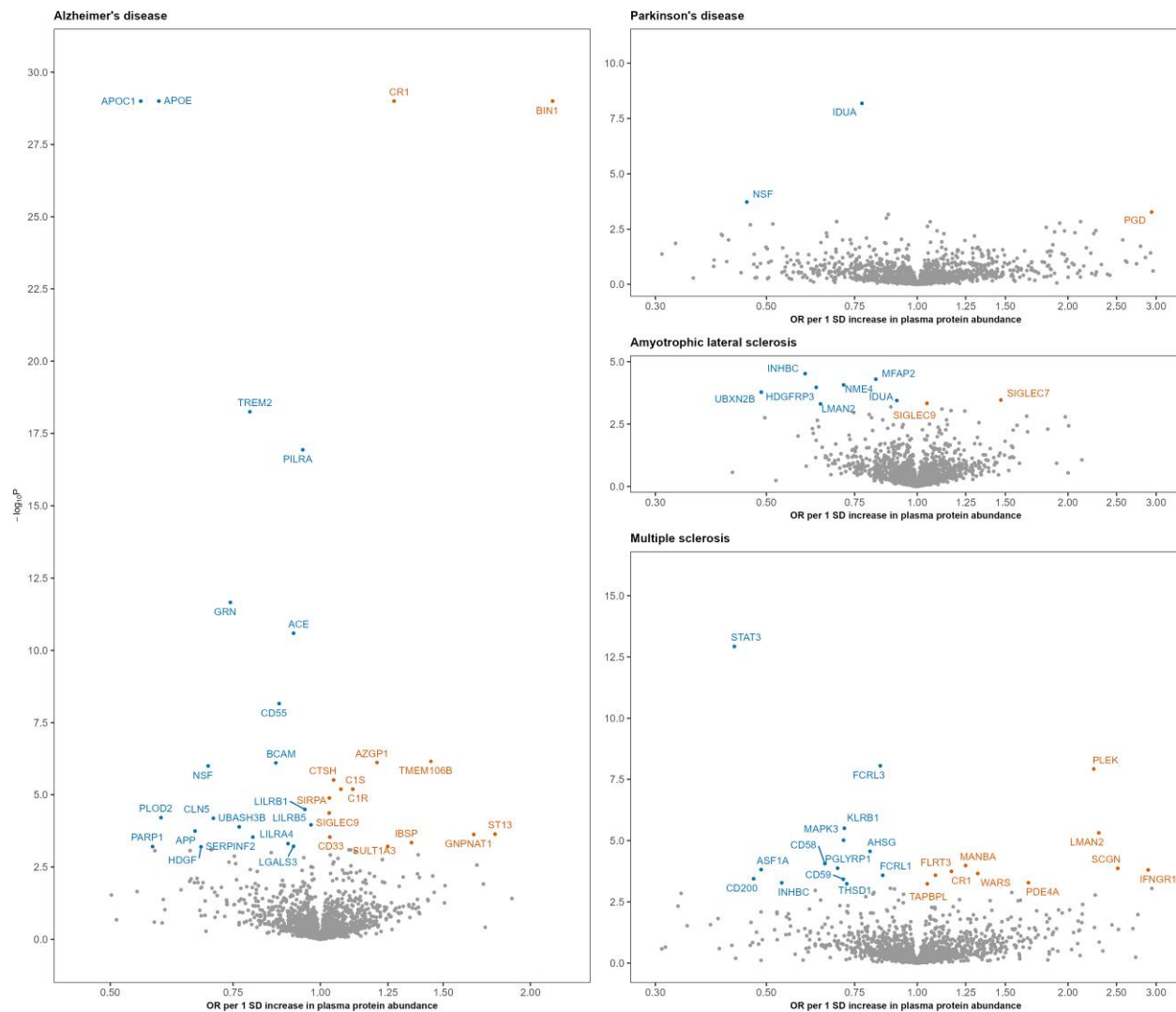
**Figure 3 Association of Olink proteins with neurodegenerative diseases using two-sample Mendelian randomization.** The lead *cis* protein quantitative trait locus was used as an instrumental variable (IV) for proteins measured through the Olink platform. In total, we tested 6762 protein–disease associations, and 95 of these were statistically significant (at 5% false discovery rate) and are annotated in this figure. The odds ratio corresponds to the Wald ratio, which is calculated by dividing the genetic effect of the IV on the disease by the genetic effect of the IV on the plasma protein abundance. To improve figure readability, one statistically significant association (TNFRSF1A with multiple sclerosis) with an extreme odds ratio (OR) is not shown. Additionally, in Alzheimer’s disease, five associations (APOE, APOC1, NECTIN2, BCAM and CR1) were capped at a  $-\log_{10} P$  of 30 to enhance visual clarity.

LRR37A2 with AD in cortex and hippocampus; and PVR with MS in cerebellum and cortex) have support for co-localization with mRNA abundance in two brain regions. Of note is that six proteins did not show evidence of co-localization with plasma protein abundance, but they co-localized with mRNA abundance in one or more brain regions (UBASH3B, NSF, LRR37A2, BIN1 and ACE with AD; and PTPRJ with MS). Discrepancies in the MR and co-localization analyses are expected, because plasma pQTLs show the least concordance and co-localization with brain eQTLs, which could be attributed to factors affecting access to the circulation, such as the blood–brain barrier (BBB).<sup>26</sup>

### Association of plasma protein abundance with brain imaging traits

To assess whether the identified proteins might have an impact on brain structure, we examined the association of *cis* pQTLs for the

proteins measured through either Olink or SomaScan platforms with 12 brain-imaging traits (i.e. intracranial brain volume, mean cortical thickness and surface area, eight subcortical brain volumes, and white matter hyper-intensities). Out of 41 738 tested associations, 3272 (7.8%) were nominally significant at  $P < 0.05$ , but only 101 (0.2%) were significant at 5% FDR (equivalent to  $P < 1.2 \times 10^{-4}$ ). Out of 101 statistically significant associations, we focused on seven associations that represented proteins associated with at least one neurodegenerative disease with additional support from co-localization (Supplementary Table 10). All these latter associations were supported by co-localization with a brain imaging trait (Supplementary Table 11). Plasma APOE abundance was associated with hippocampal volume, amygdala volume, nucleus accumbens volume and white matter hyper-intensities. Plasma PILRA and PILRB abundance was associated with caudate nucleus volume. PILRA co-localized with caudate nucleus volume using a *cis* pQTL from either Olink or SomaScan platforms.



**Figure 4** Association of SomaScan proteins with neurodegenerative diseases using two-sample Mendelian randomization. The lead *cis* protein quantitative trait locus was used as an instrumental variable (IV) for proteins measured through the SomaScan platform. In total, we tested 6615 protein–disease associations, and 74 of these were statistically significant (at 5% false discovery rate) and are annotated. The odds ratio (OR) corresponds to the Wald ratio, which is calculated by dividing the genetic effect of the IV on the disease by the genetic effect of the IV on the plasma protein abundance. To improve figure readability, five statistically significant associations (TNFRSF1A, DKKL1, JUNN, IL2RA and PTPRJ with multiple sclerosis) with extreme ORs are not shown. Additionally, in Alzheimer’s disease, four associations from the SomaScan platform (APOE, APOC1, CR1 and BIN1) were capped at a  $-\log_{10} P$  of 30 to enhance visual clarity.

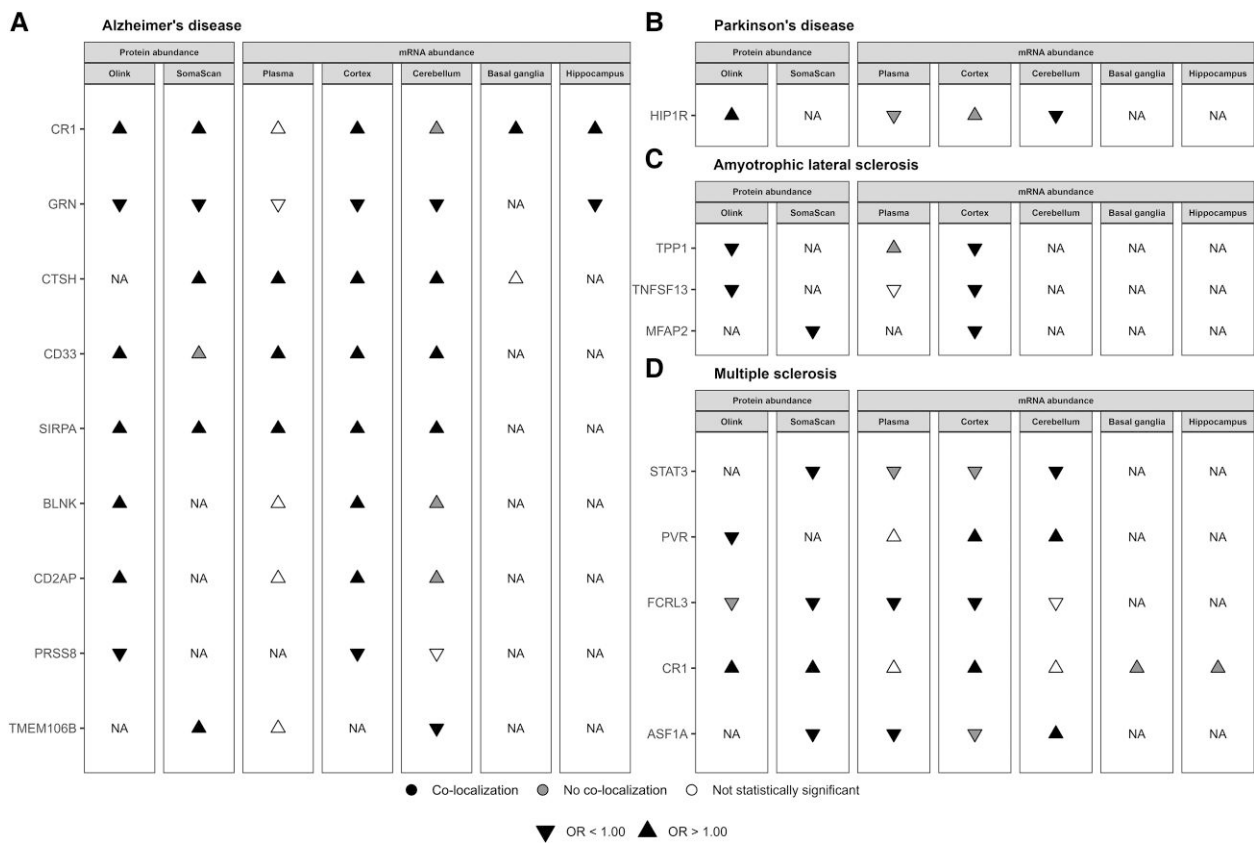
### Cumulative evidence from *cis* protein and expression quantitative trait loci

Our study has combined evidence from MR and co-localization to identify potentially causal relationships between proteins and neurodegenerative diseases using genetic associations with plasma protein abundance, plasma mRNA abundance, and brain and spinal cord mRNA abundance. On the basis of the overall evidence for association with neurological diseases, we defined three categories of evidence (Fig. 5 and Supplementary Fig. 13). In Category 1, we identified 18 proteins that showed evidence of association and co-localization when we used a plasma pQTL and an eQTL in at least one brain region. These proteins are TMEM106B, SIRPA, PRSS8, GRN, CTSH, CR1, CD33, CD2AP and BLNK for AD; HIP1R for PD; TPP1, MFAP2 and TNFSF13 for ALS; and STAT3, PVR, FCRL3, CR1 and ASF1A for MS (Fig. 5). In Category 2, we identified five proteins that showed evidence of association and co-localization when we used a plasma

pQTL and plasma eQTL (but not studied or detected in any brain region). In Category 3, we identified 26 proteins that showed association and co-localization only when we considered a plasma pQTL but not a plasma or brain eQTL.

### Downstream analyses

To gain a better understanding of the biological implications of our findings, we examined whether the prioritized proteins show high specificity in particular tissues and single cells, and we assessed their presence in distinct gene expression clusters, using data derived from the Human Protein Atlas (Supplementary Table 12 and Supplementary Fig. 14).<sup>59</sup> Although the majority of the proteins presented low tissue specificity, multiple proteins were present in the tissue expression clusters related to liver, lymphoid tissue and spleen. Likewise, although most proteins showed low brain regional specificity, several proteins were present in the brain expression



**Figure 5 Summary of cumulative evidence using plasma cis protein quantitative trait loci and plasma or brain cis expression quantitative trait loci as instrumental variables for proteins associated with neurodegenerative diseases.** This figure shows the 18 proteins that demonstrated evidence of association and co-localization when we used plasma protein quantitative trait loci (pQTLs) and expression quantitative trait loci (eQTLs) in at least one brain region. The remaining proteins are presented in [Supplementary Fig. 13](#). A–D each correspond to one neurodegenerative disease: (A) Alzheimer’s disease; (B) Parkinson’s disease; (C) amyotrophic lateral sclerosis; and (D) multiple sclerosis. When the Mendelian randomization analysis did not show a statistically significant effect, co-localization was not performed. NA (Not Available) indicates that no cis pQTL or cis eQTL was available for a particular protein.

clusters related to astrocytes, macrophages and microglia. At the single-cell level, the majority of the proteins showed high specificity in cells related to the immune system.

To assess whether the prioritized proteins indicate specific biological mechanisms underlying neurodegenerative diseases, we performed an enrichment analysis. Our analysis highlighted the involvement of the innate and adaptive immune system, and the role of lysosomes in AD. Likewise, proteins associated with MS were enriched for immune-related processes ([Supplementary Table 13](#)). In contrast, no enrichment was observed for the proteins associated with PD and ALS.

## Discussion

This study has systematically assessed the associations of >2700 proteins with four neurodegenerative diseases using summary statistics from large-scale proteogenomic data and the latest GWAS for disease risk. We identified 50 unique associations between plasma protein abundance and neurodegenerative diseases with support from MR and evidence of co-localization. Twenty-seven of these associations are known disease loci reported in GWAS, including *APOE*, *MME*, *CD2AP*, *CD33* and *IL34* for AD, and *CD40*, *CD58*, *EVIS*, *IL7R* and *STAT3* for MS, and 23 associations represent previously unreported genetic associations with neurodegenerative diseases.

## Proteins related to Alzheimer’s disease

Accumulation of  $\beta$ -amyloid is a key pathological feature for AD. *APOE* is a protein directly involved in the regulation of the  $\beta$ -amyloid aggregation and clearance in the brain.<sup>63</sup> *CD2AP* actively participates in the metabolism of  $\beta$ -amyloid, and knockout of *CD2AP* results in endosomal accumulation of  $\beta$ -amyloid in animal models.<sup>64</sup> *MME* is also another important enzyme of  $\beta$ -amyloid degradation.<sup>65</sup>

Our findings highlight the role of microglia, which is a cell type equivalent to peripheral macrophages in the brain responsible for the clearance of  $\beta$ -amyloid peptides.<sup>66</sup> Of note is that our analysis contributes three newly reported AD loci related to microglial function (*SIRPA*, *SIGLEC9* and *PRSS8*). *CD33* and *IL34* are expressed in microglia and inhibit the microglial uptake of  $\beta$ -amyloid and, therefore, influence the accumulation of amyloid plaque.<sup>67,68</sup> Anti-*CD33* antibodies are used for the treatment of acute myeloid leukaemia and have previously been suggested for drug repurposing for AD.<sup>69</sup> *SIGLEC9* participates in the immune response to several bacterial pathogens by reducing bacterial dissemination into the brain and exerts neuro-protective effects by suppressing inflammatory responses to the brain.<sup>70</sup> *SIRPA* regulates microglial phagocytosis and the transmigration of monocytes across the BBB and participates in the pathogenesis of neurodegeneration in preclinical models.<sup>71,72</sup> *PRSS8* modulates Toll-like receptor 4, which is a receptor in the membrane of microglia and contributes to microglial activation and

phagocytosis of  $\beta$ -amyloid.<sup>73,74</sup> The complement system regulates microglial function and neuro-inflammation,<sup>75</sup> and we identified one known locus (CR1) and two newly identified loci (C1R and C1S) for AD.

Herpes simplex virus-1 (HSV-1) has been linked with neurodegeneration and cognitive defects in mouse animal models.<sup>76,77</sup> HSV-1 binds to PILRA, a protein associated with AD, to infect cells. PILRA is a cell surface inhibitory receptor expressed on innate immune cells, including microglia.<sup>78,79</sup> In our analysis, PILRA was associated with caudate nucleus atrophy, which has been previously observed in other neurodegenerative diseases, including frontotemporal dementia, PD and Huntington's disease.<sup>80–82</sup> This finding could indicate that HSV-1 participates in the pathogenesis of AD by affecting caudate nucleus.

There is an increasing amount of evidence supporting the role in neurodegenerative diseases of lysosomes,<sup>83</sup> which play an important role in phagocytic cells, including microglia.<sup>84</sup> Our analysis identified three known (GRN, TMEM106B and CTSH) and one novel (CLN5) AD-associated loci related to lysosomal functions. GRN protects against  $\beta$ -amyloid deposition and toxicity in AD mouse models, and its deficiency has been linked to neural circuit development and maintenance, stress response and innate immunity.<sup>84,85</sup> TMEM106B has previously been linked to frontotemporal dementia, and there is evidence of its interaction with GRN; both of them are considered crucial markers of brain ageing.<sup>86</sup> Moreover, genetic deficiency of either CLN5 or GRN is responsible for an inherited lysosomal disease.<sup>87,88</sup> Loss of CLN5 leads to deficits in neurodevelopment in mice models.<sup>87</sup> CTSH belongs to the cathepsin superfamily, which is a large group of proteases located in the lysosomes.<sup>89,90</sup> Knockout of CTSH affects the phagocytosis of amyloid- $\beta$  in microglial cells.<sup>90</sup>

Our analysis found one known protein (BLNK) and two additional newly identified proteins (PLOC2 and ZBTB18) potentially participating in the pathogenesis of AD through other pathways. PLOC2 is overexpressed in fibroblasts, strengthening the current evidence for a potential role for fibroblasts in the pathogenesis of AD through remodelling of the extracellular matrix alongside amyloid plaques.<sup>91</sup> ZBTB18 is an essential transcription factor for embryonic cerebral cortex development<sup>92</sup>; it has been identified as a contributing factor to the 1q43q44 microdeletion syndrome, which is characterized by variable intellectual disability and brain malformations.<sup>93</sup> BLNK is involved in B-cell receptor signalling; although the role of B cells in AD is not well understood, targeting B cells has been suggested to be beneficial for AD patients by delaying disease progression.<sup>94</sup>

### Proteins related to Parkinson's disease and amyotrophic lateral sclerosis

Although the GWAS for PD and ALS have relatively small sample sizes, our study was nevertheless able to identify one novel locus for PD (CTF1) and three novel loci for ALS (TPP1, TNFSF13 and MFAP2). CTF1 is a neurotrophic factor in the interleukin-6 cytokine family. Pro-inflammatory cytokines, including interleukin-6, have previously been associated with PD.<sup>95</sup> It has also been shown in a mouse model that CTF1 transfection and expression is neuroprotective and slows progression of spinal muscular atrophy.<sup>96</sup> One of the aetiologies of ALS is mis-localization of TDP-43 to mitochondria, causing neurotoxicity.<sup>97</sup> TPP1 is a lysosomal enzyme, and loss-of-function mutations in the gene are causally linked to a familial lysosomal disorder, in which TPP1 loss affects regulation of axonal mitochondrial transport.<sup>98</sup> Also, loss of TPP1 activity results in progressive neurological phenotypes, including ataxia and increased motor deficiency.<sup>99</sup> TNFSF13 is expressed in astrocytes and regulates neuro-inflammatory

responses.<sup>100</sup> Reactive astrocytes have neurotoxic properties and are involved in the pathogenesis of ALS.<sup>101</sup> Although MFAP2 has not previously been linked with ALS, there are potential mechanisms that can explain this association. MFAP2 is essential in maintaining vessel wall integrity, and its dysfunction leads to BBB disruption.<sup>102</sup> MFAP2 is also secreted to the extracellular matrix, and its composition in the brain affects the integrity of neurons.<sup>103,104</sup>

### Proteins related to multiple sclerosis

Experimental autoimmune encephalomyelitis is an animal model for MS. Nine of the MS-associated proteins, six known (CD5, CD40, IL7R, STAT3, TNFRSF1A and TYMP) and three newly associated with MS (PARP1, PVALB and VEGFB), are involved in the pathogenesis of experimental autoimmune encephalomyelitis.<sup>105–113</sup> This observation strengthens the validity of our findings.

The innate immune system participates in pathogen removal and regulates the response of the adaptive immune system,<sup>114</sup> including the response to Epstein–Barr virus infection, which is a pathogen associated with MS.<sup>115</sup> PVR, a known locus for MS, encodes the polio virus receptor, which is involved in the immune response to Epstein–Barr virus. Increased expression of PVR downregulates the expression of microRNAs produced by Epstein–Barr virus,<sup>116</sup> which potentially explains the apparent protective effect of higher plasma levels of PVR in our analysis. WARS, a protein newly associated with MS, is an aminoacyl-tRNA synthetase with a role as an innate immune activator in the extracellular space, acting as a primary defence system against infections, especially antiviral immunity.<sup>117,118</sup> Moreover, a newly identified MS protein, PARP1, is involved in the nuclear factor- $\kappa$ B signalling pathway,<sup>119</sup> which is activated as a response to infectious antigens, including Epstein–Barr virus,<sup>120</sup> and is an important pathway for the activation of macrophages and other innate immune cells.<sup>119</sup> The newly identified association of CR1 with MS risk indicates a role of complement, which is an important innate immune defence against infection, as has been suggested recently.<sup>121</sup> SLAMF1 and AHSG participate in Toll-like receptor 4 signalling, which activates macrophages against bacterial pathogens.<sup>122,123</sup> This observation potentially provides support to the hygiene hypothesis for the development of MS.<sup>124</sup>

The adaptive immune system consists of B cells and T cells, which are activated by innate immune cells. IL7R has a role in T- and B-cell differentiation, and its plasma levels are associated with elevated risk of MS,<sup>125</sup> but experimental IL7R inhibitors have not been successful in treating MS.<sup>126</sup> CD5 and CD58 are also involved in B- and T-cell differentiation, whose activation has a role in autoimmunity.<sup>127,128</sup> CD40 and its ligand form a complex that has a central role in the regulation of both humoral and cell-mediated immunity. Blockade of CD40L is effective in ameliorating experimental autoimmune conditions, and it has also been suggested as a potential therapeutic strategy for MS.<sup>129</sup> Also, FCRL3 is mainly expressed in B cells, and it has been linked to multiple autoimmune conditions.<sup>130</sup>

Demyelinating lesions in white and grey matter are the histopathological landmarks of MS, which are infiltrated by cells of the innate and adaptive immune system,<sup>114</sup> whereas oligodendrocytes are responsible for the myelination process.<sup>131</sup> PARP1, a newly associated protein for MS, is a driver for oligodendroglial development and myelination.<sup>132</sup> and PARP1 inhibitors have been suggested as a potential therapy for MS,<sup>133</sup> in line with our finding that elevated plasma PARP1 is associated with increased MS risk. STAT3, a known protein for MS, is important for myelin repair, and pharmacological blockade of STAT3 activation with JAK2 inhibitors inhibits survival

and differentiation of oligodendrocyte precursor cells.<sup>134</sup> Another known protein for MS, TNFRSF1A, is involved in the TNF receptor-associated periodic syndrome, which is characterized by inflammatory demyelination. There is evidence that anti-TNF $\alpha$  therapies can result in new episodes of inflammation in MS patients.<sup>135</sup>

The BBB protects the CNS parenchyma from harmful circulating molecules and pathogens,<sup>136</sup> and altered BBB function is believed to be an important early stage in MS pathology. Several identified proteins have potential roles in the BBB, including TYMP, a key astrocyte-derived permeability factor promoting BBB breakdown,<sup>106</sup> CD40, which influences the permeability of the BBB,<sup>137</sup> and VEGFB, a newly identified MS-associated protein that is a member of the vascular growth factor family, again involved in the permeability of the BBB.<sup>138</sup>

The node of Ranvier on white matter demyelinated axons is profoundly altered or disrupted in patients with MS,<sup>139,140</sup> and two newly identified proteins (NCS1 and CDH15) are involved in its function. NCS1 is involved in the regulation of intracellular calcium signalling and is identified in the nodes of Ranvier. NCS1 also participates in the pathogenesis of chemotherapy-induced peripheral neuropathy.<sup>141</sup> A member of the cadherin protein family, CDH15, participates in the function of the node of Ranvier and has previously been associated with chronic inflammatory demyelinating polyneuropathy, a demyelinating disease of the peripheral nervous system.<sup>142</sup> Moreover, lesions in the dorsal root ganglion are identified in experimental autoimmune encephalomyelitis,<sup>143,144</sup> and FLRT3, a newly identified protein related to MS, is overexpressed in the dorsal root ganglion and has been associated with neuropathic pain in animal models.<sup>145</sup>

Three further newly identified MS-associated loci (PVALB, TST and ASF1A) potentially indicate additional molecular pathways contributing to MS. PVALB is specifically expressed by GABAergic interneurons and has been suggested as a potential MS-specific marker of grey matter neurodegeneration.<sup>146</sup> TST is an enzyme involved in mitochondrial sulphur and selenium metabolism,<sup>147</sup> and it has been shown that exposure to oxidative stress owing to mitochondrial dysfunction contributes to the chronic demyelination.<sup>148</sup> ASF1A is a histone chaperone that has been implicated in neuro-inflammation and neurodegeneration processes through activation of microglia.<sup>149,150</sup>

### Comparison of Mendelian randomization studies and observational studies

Observational studies have examined the association between the plasma proteome and risk of AD.<sup>151–153</sup> We observed that there is no agreement between the proteins in our study and the proteins identified from the observational analyses. Some proteins do not have a cis pQTL, and they could not be tested in our analysis. There are also additional reasons that could lead to differences between observational studies and MR studies that should be acknowledged. First, the observational analyses adjusted for APOE  $\epsilon$ 4 status, and there is evidence that APOE  $\epsilon$ 4 status modifies proteomic signatures in AD.<sup>154</sup> Second, effect estimates based on MR assume a lifelong exposure to altered protein levels from birth.<sup>155</sup> In contrast, observational analyses are based on the human proteome measured at a single time point in adult life. Third, a cis pQTL for a particular protein can have pleiotropic effects, acting as *trans* pQTL for other proteins. Fourth, statistical power of an MR analysis depends on the explained variability of the protein abundance from the cis pQTL.<sup>156</sup> Therefore, using a lead cis pQTL with a low explained variability might lead to an underpowered MR analysis. Fifth, results from observational analyses are sensitive to the selection of covariates for adjustment.<sup>157</sup> For example, by adjusting for a covariate that mediates the effect of

a protein on risk of AD, the causal pathway is blocked, hindering the identification of an association in an observational analysis.<sup>158</sup> Also, adjusting for a covariate that is a collider introduces collider bias that leads to spurious associations. Currently, there are no large prospective studies examining the role of the human proteome in the development of PD, ALS or MS.

### Plasma and tissue-specific proteomic effects

The identification of proteins with roles in many of the biological processes relevant to neurodegenerative diseases supports the idea that targeting such proteins might form the basis of future drug development. However, it seems likely that abundance of these proteins in plasma is not directly relevant to disease pathology, and that therapies will need to be targeted to the relevant tissue or cell type. Although a drug might modify levels of the identified proteins in plasma, we cannot assume that it would cross the BBB for brain-targeting drugs.<sup>159</sup> Nevertheless, although our results are based primarily on proteins measured in plasma, it is plausible that the same genetic factors have similar effects on protein levels in more relevant tissues, and that our results reflect similarities in processes such as macrophage activity, lysosomal activity and  $\beta$ -amyloid metabolism in blood and brain. For example, in the pathogenesis of MS, the activation of the innate and adaptive immune system occurs initially in the periphery and is then transferred to the CNS.<sup>114</sup> In particular, cis pQTLs, particularly those directly impacting protein-coding sequences, will frequently have similar effects across diverse tissues.<sup>160</sup>

### Limitations

There are several limitations to the present study. First, our analyses are underpowered for PD, because full summary statistics including 23andMe are not publicly available, greatly reducing the sample size in the available data. Second, GWAS for the neurological diseases primarily assess risk for disease and not disease progression. Therefore, the identified proteins can be considered as potential biomarkers for prediction or diagnosis of the diseases or drug targets for disease prevention but not necessarily for disease progression.<sup>161</sup> Third, our study is constrained by the number of proteins that can be analysed using Olink and SomaScan platforms, which cover only a fraction of the entire proteome. Fourth, our colocalization analysis assumed the presence of a single causal genetic variant per genetic locus. Absence of co-localization could be observed in the case of violation of this assumption, where more than one causal genetic variant exists in a particular genetic locus for a particular trait.<sup>25</sup> Fifth, sample overlap can bias the results of an MR analysis.<sup>162</sup> For the MR effect estimates using pQTLs from the SomaScan platform, there is no sample overlap for any of the traits of interest. For the MR analyses on ALS and MS, there is no inclusion of participants from the UK Biobank or deCODE Health Study in the respective GWAS. For the MR analyses on AD and PD using cis pQTLs from Olink platform, there is a sample overlap of <5%. Owing to the minimal sample overlap, we expect that our results will not be impacted substantially by this.

### Conclusion

We have presented a comprehensive analysis of associations of the plasma proteome with neurodegenerative diseases by considering proteins measured through either Olink or SomaScan platforms. We identified multiple proteins with a potential causal role in neurodegenerative diseases. The newly identified proteins for AD are

involved in the immune response to bacterial pathogens, complement system, transmigration of monocytes across the BBB, Toll-like receptor 4 signalling, lysosomal function and fibroblasts. The newly identified proteins for MS are involved in the innate immune system, complement, microglia, oligodendrocytes, permeability of the BBB, GABAergic interneurons and the function of the node of Ranvier and dorsal root ganglion. Our analysis covered only a modest proportion of the human proteome and was limited to proteins measured in plasma; therefore, further expansion of the multiplexed antibody-based and aptamer-based assays and conducting large-scale assays in more directly relevant tissues will offer additional insights into the role of protein abundance in the development of neurodegenerative diseases. Moreover, better characterization of the protein isoforms targeted by these complementary proteomics platforms will offer additional insights into the biological interpretation of the findings.

## Data availability

Descriptive characteristics of the data sources used in this study are shown in [Supplementary Table 1](#). The cis pQTLs that were used as IVs are publicly available in the relevant publications.<sup>26,27</sup> Summary statistics for the GWAS on AD, PD, ALS and MS are available through GWAS Catalog (<https://www.ebi.ac.uk/gwas/>). Summary statistics for the GWAS on brain volume traits are available upon request from the ENIGMA consortium (<https://enigma.ini.usc.edu/>). Summary statistics for the GWAS on white matter hyper-intensities are publicly available through the Cerebrovascular Disease Knowledge Portal (<https://cd.hugeamp.org/>). Summary statistics for the cis region of gene expression in plasma and brain regions are publicly available through the eQTLGen (<https://www.eqtlgen.org/>) and the MetaBrain (<https://www.metabrain.nl/>) consortia, respectively.

## Acknowledgements

A preliminary version of this research article was presented as a poster at the 2023 Annual Meeting of the American Society of Human Genetics.

## Funding

Lazaros Belbasis is supported by an Oxford Population Health Early Career Research Fellowship. Cornelia van Duijn is supported by the US National Institute on Aging, the NovoNordisk - Oxford Fellowship Programme, the Oxford - GlaxoSmithKline Institute of Molecular and Computational Medicine, the Centre for Artificial Intelligence for Precision Medicine at the University of Oxford and King Abdul Aziz University, the Alzheimer's Research UK, the UK National Institute for Health and Care Research (NIHR) Oxford Biomedical Research Centre (BRC), the ZonMW, and the Alzheimer Nederland. The computational aspects of this research were supported by the Wellcome Trust Core Award Grant Number 203141/Z/16/Z and the NIHR Oxford BRC. The views expressed are those of the author(s) and not necessarily those of the UK National Health Service, the NIHR or the UK Department of Health and Social Care.

## Competing interests

The authors report no competing interests.

## Supplementary material

Supplementary material is available at *Brain* online.

## References

1. Feigin VL, Nichols E, Alam T, et al. Global, regional, and national burden of neurological disorders, 1990–2016: A systematic analysis for the Global Burden of Disease Study 2016. *Lancet Neurol.* 2019;18:459–480.
2. Dugger BN, Dickson DW. Pathology of neurodegenerative diseases. *Cold Spring Harb Perspect Biol.* 2017;9:a028035.
3. Nichols E, Szeke CEI, Vollset SE, et al. Global, regional, and national burden of Alzheimer's disease and other dementias, 1990–2016: A systematic analysis for the Global Burden of Disease Study 2016. *Lancet Neurol.* 2019;18:88–106.
4. Dorsey ER, Elbaz A, Nichols E, et al. Global, regional, and national burden of Parkinson's disease, 1990–2016: A systematic analysis for the Global Burden of Disease Study 2016. *Lancet Neurol.* 2018;17:939–953.
5. Wallin MT, Culpepper WJ, Nichols E, et al. Global, regional, and national burden of multiple sclerosis 1990–2016: A systematic analysis for the Global Burden of Disease Study 2016. *Lancet Neurol.* 2019;18:269–285.
6. Logroscino G, Piccininni M, Marin B, et al. Global, regional, and national burden of motor neuron diseases 1990–2016: A systematic analysis for the Global Burden of Disease Study 2016. *Lancet Neurol.* 2018;17:1083–1097.
7. Friese MA, Schattling B, Fugger L. Mechanisms of neurodegeneration and axonal dysfunction in multiple sclerosis. *Nat Rev Neurol.* 2014;10:225–238.
8. Nalls MA, Blauwendraat C, Vallerga CL, et al. Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: A meta-analysis of genome-wide association studies. *Lancet Neurol.* 2019;18:1091–1102.
9. Bellenguez C, Küçükali F, Jansen IE, et al. New insights into the genetic etiology of Alzheimer's disease and related dementias. *Nat Genet.* 2022;54:412–436.
10. van Rheenen W, van der Spek RAA, Bakker MK, et al. Common and rare variant association analyses in amyotrophic lateral sclerosis identify 15 risk loci with distinct genetic architectures and neuron-specific biology. *Nat Genet.* 2021;53:1636–1648.
11. International Multiple Sclerosis Genetics Consortium. Multiple sclerosis genomic map implicates peripheral immune cells and microglia in susceptibility. *Science.* 2019;365:eaav7188.
12. Zheng J, Haberland V, Baird D, et al. Phenome-wide Mendelian randomization mapping the influence of the plasma proteome on complex diseases. *Nat Genet.* 2020;52:1122–1131.
13. Chen L, Peters JE, Prins B, et al. Systematic Mendelian randomization using the human plasma proteome to discover potential therapeutic targets for stroke. *Nat Commun.* 2022;13:6143.
14. Santos R, Ursu O, Gaulton A, et al. A comprehensive map of molecular drug targets. *Nat Rev Drug Discov.* 2017;16:19–34.
15. Plenge RM, Scolnick EM, Altshuler D. Validating therapeutic targets through human genetics. *Nat Rev Drug Discov.* 2013;12:581–594.
16. Suhre K, McCarthy MI, Schwenk JM. Genetics meets proteomics: Perspectives for large population-based studies. *Nat Rev Genet.* 2021;22:19–37.
17. Holmes MV, Richardson TG, Ference BA, Davies NM, Davey Smith G. Integrating genomics with biomarkers and therapeutic targets to invigorate cardiovascular drug development. *Nat Rev Cardiol.* 2021;18:435–453.

18. Kibinge NK, Relton CL, Gaunt TR, Richardson TG. Characterizing the causal pathway for genetic variants associated with neurological phenotypes using human brain-derived proteome data. *Am J Hum Genet.* 2020;106:885–892.
19. Wingo AP, Liu Y, Gerasimov ES, et al. Integrating human brain proteomes with genome-wide association data implicates new proteins in Alzheimer's disease pathogenesis. *Nat Genet.* 2021;53:143–146.
20. Baird DA, Liu JZ, Zheng J, et al. Identifying drug targets for neurological and psychiatric disease via genetics and the brain transcriptome. *PLoS Genet.* 2021;17:e1009224.
21. Storm CS, Kia DA, Almramhi MM, et al. Finding genetically-supported drug targets for Parkinson's disease using Mendelian randomization of the druggable genome. *Nat Commun.* 2021;12:7342.
22. Ou YN, Yang YX, Deng YT, et al. Identification of novel drug targets for Alzheimer's disease by integrating genetics and proteomes from brain and blood. *Mol Psychiatry.* 2021;26:6065–6073.
23. Jacobs BM, Taylor T, Awad A, et al. Summary-data-based Mendelian randomization prioritizes potential druggable targets for multiple sclerosis. *Brain Commun.* 2020;2:fcaa119.
24. Stunnenberg HG, Hubner NC. Genomics meets proteomics: Identifying the culprits in disease. *Hum Genet.* 2014;133:689–700.
25. Zuber V, Grinberg NF, Gill D, et al. Combining evidence from Mendelian randomization and colocalization: Review and comparison of approaches. *Am J Hum Genet.* 2022;109:767–782.
26. Sun BB, Chiou J, Traylor M, et al. Plasma proteomic associations with genetics and health in the UK Biobank. *Nature.* 2023;622:329–338.
27. Ferkingstad E, Sulem P, Atlason BA, et al. Large-scale integration of the plasma proteome with genetics and disease. *Nat Genet.* 2021;53:1712–1721.
28. Welter D, MacArthur J, Morales J, et al. The NHGRI GWAS catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.* 2014;42(Database issue):D1001–D1006.
29. Aguet F, Alasoo K, Li YI, et al. Molecular quantitative trait loci. *Nat Rev Methods Primers.* 2023;3:4.
30. Gkatzionis A, Burgess S, Newcombe PJ. Statistical methods for cis-Mendelian randomization with two-sample summary-level data. *Genet Epidemiol.* 2023;47:3–25.
31. Burgess S, Thompson SG. Bias in causal estimates from Mendelian randomization studies with weak instruments. *Stat Med.* 2011;30:1312–1323.
32. Pietzner M, Wheeler E, Carrasco-Zanini J, et al. Synergistic insights into human health from aptamer- and antibody-based proteomic profiling. *Nat Commun.* 2021;12:6822.
33. Kuhn RM, Haussler D, Kent WJ. The UCSC genome browser and associated tools. *Brief Bioinform.* 2013;14:144–161.
34. Hartwig FP, Davies NM, Hemani G, Davey Smith G. Two-sample Mendelian randomization: Avoiding the downsides of a powerful, widely applicable but potentially fallible technique. *Int J Epidemiol.* 2016;45:1717–1726.
35. Burgess S, Scott RA, Timpson NJ, Davey Smith G, Thompson SG, EPIC-InterAct Consortium. Using published data in Mendelian randomization: A blueprint for efficient identification of causal risk factors. *Eur J Epidemiol.* 2015;30:543–552.
36. Hemani G, Zheng J, Elsworth B, et al. The MR-base platform supports systematic causal inference across the human phenotype. *Elife.* 2018;7:e34408.
37. Hemani G, Tilling K, Davey Smith G. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. *PLoS Genet.* 2017;13:e1007081.
38. Burgess S, Small DS, Thompson SG. A review of instrumental variable estimators for Mendelian randomization. *Stat Methods Med Res.* 2017;26:2333–2355.
39. Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J Royal Stat Soc Ser B.* 1995;57:289–300.
40. Purcell S, Neale B, Todd-Brown K, et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007;81:559–575.
41. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet Epidemiol.* 2013;37:658–665.
42. Giambartolomei C, Vukcevic D, Schadt EE, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* 2014;10:e1004383.
43. Hukku A, Pividori M, Luca F, Pique-Regi R, Im HK, Wen X. Probabilistic colocalization of genetic variants from complex and molecular traits: Promise and limitations. *Am J Hum Genet.* 2021;108:25–35.
44. Kunkle BW, Grenier-Boley B, Sims R, et al. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates A $\beta$ , tau, immunity and lipid processing. *Nat Genet.* 2019;51:414–430.
45. Albert FW, Kruglyak L. The role of regulatory variation in complex traits and disease. *Nat Rev Genet.* 2015;16:197–212.
46. Vösa U, Claringbould A, Westra HJ, et al. Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nat Genet.* 2021;53:1300–1310.
47. de Klein N, Tsai EA, Vochteloo M, et al. Brain expression quantitative trait locus and network analyses reveal downstream effects and putative drivers for brain-related diseases. *Nat Genet.* 2023;55:377–388.
48. Thompson PM, Stein JL, Medland SE, et al. The ENIGMA consortium: Large-scale collaborative analyses of neuroimaging and genetic data. *Brain Imaging Behav.* 2014;8:153–182.
49. Persyn E, Hanscombe KB, Howson JMM, Lewis CM, Traylor M, Markus HS. Genome-wide association study of MRI markers of cerebral small vessel disease in 42,310 participants. *Nat Commun.* 2020;11:2175.
50. Grasby KL, Jahanshad N, Painter JN, et al. The genetic architecture of the human cerebral cortex. *Science.* 2020;367:eaay6690.
51. Adams HHH, Hibar DP, Chouraki V, et al. Novel genetic loci underlying human intracranial volume identified through genome-wide association. *Nat Neurosci.* 2016;19:1569–1582.
52. Hibar DP, Adams HHH, Jahanshad N, et al. Novel genetic loci associated with hippocampal volume. *Nat Commun.* 2017;8:13624.
53. Satizabal CL, Adams HHH, Hibar DP, et al. Genetic architecture of subcortical brain structures in 38,851 individuals. *Nat Genet.* 2019;51:1624–1636.
54. Ashburner M, Ball CA, Blake JA, et al. Gene ontology: Tool for the unification of biology. *Nat Genet.* 2000;25:25–29.
55. Jassal B, Matthews L, Viteri G, et al. The reactome pathway knowledgebase. *Nucleic Acids Res.* 2020;48(D1):D498–D503.
56. Martens M, Ammar A, Riutta A, et al. WikiPathways: Connecting communities. *Nucleic Acids Res.* 2021;49(D1):D613–D621.
57. Kanehisa M. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 2000;28:27–30.
58. Szklarczyk D, Gable AL, Nastou KC, et al. The STRING database in 2021: Customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* 2021;49(D1):D605–D612.

59. Sjöstedt E, Zhong W, Fagerberg L, et al. An atlas of the protein-coding genes in the human, pig, and mouse brain. *Science*. 2020; 367:eay5947.
60. Digre A, Lindskog C. The human protein atlas—Integrated omics for single cell mapping of the human proteome. *Protein Sci*. 2023;32:e4562.
61. Pietzner M, Wheeler E, Carrasco-Zanini J, et al. Mapping the proteo-genomic convergence of human diseases. *Science*. 2021;374:eabj1541.
62. Buccitelli C, Selbach M. mRNAs, proteins and the emerging principles of gene expression control. *Nat Rev Genet*. 2020;21: 630–644.
63. Liu CC, Kanekiyo T, Xu H, Bu G. Apolipoprotein E and Alzheimer disease: Risk, mechanisms and therapy. *Nat Rev Neurol*. 2013;9:106–118.
64. Ubelmann F, Burrinha T, Salavessa L, et al. Bin1 and CD2AP polarise the endocytic generation of beta-amyloid. *EMBO Rep*. 2017;18:102–122.
65. Marr RA, Rockenstein E, Mukherjee A, et al. Neprilysin gene transfer reduces human amyloid pathology in transgenic mice. *J Neurosci*. 2003;23:1992–1996.
66. Perry VH, Nicoll JAR, Holmes C. Microglia in neurodegenerative disease. *Nat Rev Neurol*. 2010;6:193–201.
67. Griuciu A, Serrano-Pozo A, Parrado AR, et al. Alzheimer's disease risk gene CD33 inhibits microglial uptake of amyloid beta. *Neuron*. 2013;78:631–643.
68. Baghdadi M, Umeyama Y, Hama N, et al. Interleukin-34, a comprehensive review. *J Leukoc Biol*. 2018;104:931–951.
69. Malik M, Chiles J, Xi HS, et al. Genetics of CD33 in Alzheimer's disease and acute myeloid leukemia. *Hum Mol Genet*. 2015;24: 3557–3570.
70. Macauley MS, Crocker PR, Paulson JC. Siglec-mediated regulation of immune cell function in disease. *Nat Rev Immunol*. 2014; 14:653–666.
71. Ding X, Wang J, Huang M, et al. Loss of microglial SIRP $\alpha$  promotes synaptic pruning in preclinical models of neurodegeneration. *Nat Commun*. 2021;12:2030.
72. Zhang H, Li F, Yang Y, Chen J, Hu X. SIRP/CD47 signaling in neurological disorders. *Brain Res*. 2015;1623:74–80.
73. Uchimura K, Hayata M, Mizumoto T, et al. The serine protease prostaticin regulates hepatic insulin sensitivity by modulating TLR4 signalling. *Nat Commun*. 2014;5:3428.
74. Chen GF, Xu TH, Yan Y, et al. Amyloid beta: Structure, biology and structure-based therapeutic development. *Acta Pharmacol Sin*. 2017;38:1205–1235.
75. Morgan BP. Complement in the pathogenesis of Alzheimer's disease. *Semin Immunopathol*. 2018;40:113–124.
76. Lopatko Lindman K, Jonsson C, Weidung B, et al. PILRA polymorphism modifies the effect of APOE4 and GM17 on Alzheimer's disease risk. *Sci Rep*. 2022;12:13264.
77. De Chiara G, Piacentini R, Fabiani M, et al. Recurrent herpes simplex virus-1 infection induces hallmarks of neurodegeneration and cognitive deficits in mice. *PLoS Pathog*. 2019;15:e1007617.
78. Agostini S, Costa AS, Mancuso R, Guerini FR, Nemni R, Clerici M. The PILRA G78R variant correlates with higher HSV-1-specific IgG titers in Alzheimer's disease. *Cell Mol Neurobiol*. 2019;39:1217–1221.
79. Itzhaki RF, Lathé R, Balin BJ, et al. Microbes and Alzheimer's disease. *J Alzheimers Dis*. 2016;51:979–984.
80. Bogaard SJA, Dumas EM, Acharya TP, et al. Early atrophy of pallidum and accumbens nucleus in Huntington's disease. *J Neurol*. 2011;258:412–420.
81. Möller C, Dieleman N, van der Flier WM, et al. More atrophy of deep gray matter structures in frontotemporal dementia compared to Alzheimer's disease. *J Alzheimers Dis*. 2015;44: 635–647.
82. Carriere N, Besson P, Dujardin K, et al. Apathy in Parkinson's disease is associated with nucleus accumbens atrophy: A magnetic resonance imaging shape analysis. *Mov Disord*. 2014;29: 897–903.
83. Bonam SR, Wang F, Muller S. Lysosomes as a therapeutic target. *Nat Rev Drug Discov*. 2019;18:923–948.
84. Kao AW, McKay A, Singh PP, Brunet A, Huang EJ. Progranulin, lysosomal regulation and neurodegenerative disease. *Nat Rev Neurosci*. 2017;18:325–333.
85. Minami SS, Min SW, Krabbe G, et al. Progranulin protects against amyloid  $\beta$  deposition and toxicity in Alzheimer's disease mouse models. *Nat Med*. 2014;20:1157–1164.
86. Feng T, Mai S, Roscoe JM, et al. Loss of TMEM106B and PGRN leads to severe lysosomal abnormalities and neurodegeneration in mice. *EMBO Rep*. 2020;21:e50219.
87. Singh Y, Leinonen H, Fazaludeen F, et al. Loss of Cln5 leads to altered *Gad1* expression and deficits in interneuron development in mice. *Hum Mol Genet*. 2019;28:3309–3322.
88. Petkau TL, Leavitt BR. Progranulin in neurodegenerative disease. *Trends Neurosci*. 2014;37:388–398.
89. Saftig P, Klumperman J. Lysosome biogenesis and lysosomal membrane proteins: Trafficking meets function. *Nat Rev Mol Cell Biol*. 2009;10:623–635.
90. Li Y, Xu M, Xiang BL, et al. Functional genomics identify causal variant underlying the protective CTSH locus for Alzheimer's disease. *Neuropsychopharmacology*. 2023;48:1555–1566.
91. Dorrier CE, Jones HE, Pintarić L, Siegenthaler JA, Daneman R. Emerging roles for CNS fibroblasts in health, injury and disease. *Nat Rev Neurosci*. 2022;23:23–34.
92. Heng JI-T, Viti L, Pugh K, Marshall OJ, Agostino M. Understanding the impact of ZBTB18 missense variation on transcription factor function in neurodevelopment and disease. *J Neurochem*. 2022;161:219–235.
93. van der Schoot V, de Munnik S, Venselaar H, et al. Toward clinical and molecular understanding of pathogenic variants in the ZBTB18 gene. *Mol Genet Genomic Med*. 2018;6:393–400.
94. Kim K, Wang X, Ragonnaud E, et al. Therapeutic B-cell depletion reverses progression of Alzheimer's disease. *Nat Commun*. 2021;12:2185.
95. Tan EK, Chao YX, West A, Chan LL, Poewe W, Jankovic J. Parkinson disease and the immune system—Associations, mechanisms and therapeutics. *Nat Rev Neurol*. 2020;16:303–318.
96. Lesbordes JC, Cifuentes-Diaz C, Miroglio A, et al. Therapeutic benefits of cardiotrophin-1 gene transfer in a mouse model of spinal muscular atrophy. *Hum Mol Genet*. 2003;12: 1233–1239.
97. Yu CH, Davidson S, Harapas CR, et al. TDP-43 triggers mitochondrial DNA release via mPTP to activate cGAS/STING in ALS. *Cell*. 2020;183:636–649.e18.
98. Shlevkov E, Basu H, Bray MA, et al. A high-content screen identifies TPP1 and Aurora B as regulators of axonal mitochondrial transport. *Cell Rep*. 2019;28:3224–3237.e5.
99. Domowicz MS, Chan WC, Claudio-Vázquez P, et al. Global brain transcriptome analysis of a *Tpp1* neuronal ceroid lipofuscinoses mouse model. *ASN Neuro*. 2019;11:1759091419843393.
100. Lee WH, Seo D, Lim SG, Suk K. Reverse signaling of tumor necrosis factor superfamily proteins in macrophages and microglia: Superfamily portrait in the neuroimmune interface. *Front Immunol*. 2019;10:262.
101. Vahsen BF, Gray E, Thompson AG, et al. Non-neuronal cells in amyotrophic lateral sclerosis—From pathogenesis to biomarkers. *Nat Rev Neurol*. 2021;17:333–348.

102. Craft CS, Broekelmann TJ, Mecham RP. Microfibril-associated glycoproteins MAGP-1 and MAGP-2 in disease. *Matrix Biol.* 2018;71–72:100–111.
103. Fawcett JW, Oohashi T, Pizzorusso T. The roles of perineuronal nets and the perinodal extracellular matrix in neuronal function. *Nat Rev Neurosci.* 2019;20:451–465.
104. Dityatev A, Schachner M. Extracellular matrix molecules and synaptic plasticity. *Nat Rev Neurosci.* 2003;4:456–468.
105. Tham E, Gielen AW, Khademi M, Martin C, Piehl F. Decreased expression of VEGF-A in rat experimental autoimmune encephalomyelitis and in cerebrospinal fluid mononuclear cells from patients with multiple sclerosis. *Scand J Immunol.* 2006;64:609–622.
106. Chapouly C, Tadesse Argaw A, Horng S, et al. Astrocytic TYMP and VEGFA drive blood–brain barrier opening in inflammatory central nervous system lesions. *Brain.* 2015;138:1548–1567.
107. Walline CC, Kanakasabai S, Bright JJ. IL-7R $\alpha$  confers susceptibility to experimental autoimmune encephalomyelitis. *Genes Immun.* 2011;12:1–14.
108. Gerritse K, Laman JD, Noelle RJ, et al. CD40-CD40 ligand interactions in experimental allergic encephalomyelitis and multiple sclerosis. *Proc Natl Acad Sci U S A.* 1996;93:2499–2504.
109. Cavone L, Aldinucci A, Ballerini C, Biagioli T, Moroni F, Chiarugi A. PARP-1 inhibition prevents CNS migration of dendritic cells during EAE, suppressing the encephalitogenic response and relapse severity. *Mult Scler J.* 2011;17:794–807.
110. Falco A, Pennucci R, Brambilla E, de Curtis I. Reduction in parvalbumin-positive interneurons and inhibitory input in the cortex of mice with experimental autoimmune encephalomyelitis. *Exp Brain Res.* 2014;232:2439–2449.
111. Axtell RC, Webb MS, Barnum SR, Raman C. Cutting edge: Critical role for CD5 in experimental autoimmune encephalomyelitis: Inhibition of engagement reverses disease in mice. *J Immunol.* 2004;173:2928–2932.
112. Selmaj K, Papierz W, Glabiński A, Kohno T. Prevention of chronic relapsing experimental autoimmune encephalomyelitis by soluble tumor necrosis factor receptor I. *J Neuroimmunol.* 1995;56:135–141.
113. Liu Y, Holdbrooks AT, De Sarno P, et al. Therapeutic efficacy of suppressing the JAK/STAT pathway in multiple models of experimental autoimmune encephalomyelitis. *J Immunol.* 2014;192:59–72.
114. Hemmer B, Kerschensteiner M, Korn T. Role of the innate and adaptive immune responses in the course of multiple sclerosis. *Lancet Neurol.* 2015;14:406–419.
115. Soldan SS, Lieberman PM. Epstein–Barr virus and multiple sclerosis. *Nat Rev Microbiol.* 2023;21:51–64.
116. Afrasiabi A, Parnell GP, Swaminathan S, Stewart GJ, Booth DR. The interaction of multiple sclerosis risk loci with Epstein–Barr virus phenotypes implicates the virus in pathogenesis. *Sci Rep.* 2020;10:193.
117. Ahn YH, Park S, Choi JJ, et al. Secreted tryptophanyl-tRNA synthetase as a primary defence system against infection. *Nat Microbiol.* 2016;2:16191.
118. Nie A, Sun B, Fu Z, Yu D. Roles of aminoacyl-tRNA synthetases in immune regulation and immune diseases. *Cell Death Dis.* 2019;10:901.
119. Liu T, Zhang L, Joo D, Sun SC. NF- $\kappa$ B signaling in inflammation. *Signal Transduct Target Ther.* 2017;2:17023.
120. Perkins ND. Integrating cell-signalling pathways with NF- $\kappa$ B and IKK function. *Nat Rev Mol Cell Biol.* 2007;8:49–62.
121. Höftberger R, Lassmann H, Berger T, Reindl M. Pathogenic autoantibodies in multiple sclerosis—From a simple idea to a complex concept. *Nat Rev Neurol.* 2022;18:681–688.
122. Yurchenko M, Skjesol A, Ryan L, et al. SLAMF1 is required for TLR4-mediated TRAM-TRIF-dependent signaling in human macrophages. *J Cell Biol.* 2018;217:1411–1429.
123. Pal D, Dasgupta S, Kundu R, et al. Fetuin-A acts as an endogenous ligand of TLR4 to promote lipid-induced insulin resistance. *Nat Med.* 2012;18:1279–1285.
124. Wasko NJ, Nichols F, Clark RB. Multiple sclerosis, the microbiome, TLR2, and the hygiene hypothesis. *Autoimmun Rev.* 2020;19:102430.
125. Gregory SG, Schmidt S, Seth P, et al. Interleukin 7 receptor  $\alpha$  chain (IL7R) shows allelic and functional association with multiple sclerosis. *Nat Genet.* 2007;39:1083–1091.
126. Barata JT, Durum SK, Seddon B. Flip the coin: IL-7 and IL-7R in health and disease. *Nat Immunol.* 2019;20:1584–1593.
127. Dalloul A. CD5: A safeguard against autoimmunity and a shield for cancer cells. *Autoimmun Rev.* 2009;8:349–353.
128. De Jager PL, Baecher-Allan C, Maier LM, et al. The role of the CD58 locus in multiple sclerosis. *Proc Natl Acad Sci U S A.* 2009;106:5264–5269.
129. Fadul CE, Mao-Draayer Y, Ryan KA, et al. Safety and immune effects of blocking CD40 ligand in multiple sclerosis. *Neurol Neuroimmunol Neuroinflamm.* 2021;8:e1096.
130. Davis RS. Fc receptor-like molecules. *Annu Rev Immunol.* 2007;25:525–560.
131. Traka M, Podojil JR, McCarthy DP, Miller SD, Popko B. Oligodendrocyte death results in immune-mediated CNS demyelination. *Nat Neurosci.* 2016;19:65–74.
132. Wang Y, Zhang Y, Zhang S, et al. PARP1-mediated PARylation activity is essential for oligodendroglial differentiation and CNS myelination. *Cell Rep.* 2021;37:109695.
133. Cavone L, Chiarugi A. Targeting poly(ADP-ribose) polymerase-1 as a promising approach for immunomodulation in multiple sclerosis? *Trends Mol Med.* 2012;18:92–100.
134. Steelman AJ, Zhou Y, Koito H, et al. Activation of oligodendroglial Stat3 is required for efficient remyelination. *Neurobiol Dis.* 2016;91:336–346.
135. Ottoboni L, Frohlich IY, Lee M, et al. Clinical relevance and functional consequences of the TNFRSF1A multiple sclerosis locus. *Neurology.* 2013;81:1891–1899.
136. Zierfuss B, Larochele C, Prat A. Blood–brain barrier dysfunction in multiple sclerosis: Causes, consequences, and potential effects of therapies. *Lancet Neurol.* 2024;23:95–109.
137. Masuda H, Mori M, Umehara K, et al. Soluble CD40 ligand disrupts the blood–brain barrier and exacerbates inflammation in experimental autoimmune encephalomyelitis. *J Neuroimmunol.* 2018;316:117–120.
138. Lange C, Storkebaum E, de Almodóvar CR, Dewerchin M, Carmeliet P. Vascular endothelial growth factor: A neurovascular target in neurological diseases. *Nat Rev Neurol.* 2016;12:439–454.
139. Arancibia-Carcamo IL, Attwell D. The node of Ranvier in CNS pathology. *Acta Neuropathol.* 2014;128:161–175.
140. Lubetzki C, Sol-Foulon N, Desmazières A. Nodes of Ranvier during development and repair in the CNS. *Nat Rev Neurol.* 2020;16:426–439.
141. Mo M, Erdelyi I, Szigeti-Buck K, Benbow JH, Ehrlich BE. Prevention of paclitaxel-induced peripheral neuropathy by lithium pretreatment. *FASEB J.* 2012;26:4696–4709.
142. Moritz CP, Tholance Y, Stoesesandt O, Ferraud K, Camdessanché JP, Antoine JC. CIDP antibodies target junction proteins and identify patient subgroups: An autoantigenomic approach. *Neurol Neuroimmunol Neuroinflamm.* 2021;8:e944.
143. Warwick RA, Ledgerwood CJ, Brenner T, Hanani M. Satellite glial cells in dorsal root ganglia are activated in

- experimental autoimmune encephalomyelitis. *Neurosci Lett.* 2014;569:59-62.
144. Yousuf MS, Noh M-C, Friedman TN, et al. Sensory neurons of the dorsal root ganglia become hyperexcitable in a T-cell-mediated MOG-EAE model of multiple sclerosis. *eNeuro.* 2019;6:ENEURO.0024-19.2019.
  145. Yamada M, Fujita Y, Hayano Y, et al. Increased expression of fibronectin leucine-rich transmembrane protein 3 in the dorsal root ganglion induces neuropathic pain in rats. *J Neurosci.* 2019;39:7615-7627.
  146. Magliozzi R, Pitteri M, Ziccardi S, et al. CSF parvalbumin levels reflect interneuron loss linked with cortical pathology in multiple sclerosis. *Ann Clin Transl Neurol.* 2021;8:534-547.
  147. Kruithof PD, Lunev S, Aguilar Lozano SP, et al. Unraveling the role of thiosulfate sulfurtransferase in metabolic diseases. *Biochim Biophys Acta Mol Basis Dis.* 2020;1866:165716.
  148. Franklin RJM, ffrench-Constant C, Edgar JM, Smith KJ. Neuroprotection and repair in multiple sclerosis. *Nat Rev Neurol.* 2012;8:624-634.
  149. Martire S, Banaszynski LA. The roles of histone variants in fine-tuning chromatin organization and function. *Nat Rev Mol Cell Biol.* 2020;21:522-541.
  150. Zhang XW, Feng N, Wang LC, et al. Small-molecule arone protects from neuroinflammation in LPS-activated microglia BV-2 cells by targeting histone-remodeling chaperone ASF1a. *Biochem Pharmacol.* 2020;177:113932.
  151. Walker KA, Chen J, Zhang J, et al. Large-scale plasma proteomic analysis identifies proteins and pathways associated with dementia risk. *Nat Aging.* 2021;1:473-489.
  152. Guo Y, You J, Zhang Y, et al. Plasma proteomic profiles predict future dementia in healthy adults. *Nat Aging.* 2024;4:247-260.
  153. Walker KA, Chen J, Shi L, et al. Proteomics analysis of plasma from middle-aged adults identifies protein markers of dementia risk in later life. *Sci Transl Med.* 2023;15:eadf5681.
  154. Konijnenberg E, Tijms BM, Gobom J, et al. APOE  $\epsilon$ 4 genotype-dependent cerebrospinal fluid proteomic signatures in Alzheimer's disease. *Alzheimers Res Ther.* 2020;12:65.
  155. Holmes MV, Ala-Korpela M, Smith GD. Mendelian randomization in cardiometabolic disease: Challenges in evaluating causality. *Nat Rev Cardiol.* 2017;14:577-590.
  156. Brion MJA, Shakhbazov K, Visscher PM. Calculating statistical power in Mendelian randomization studies. *Int J Epidemiol.* 2013;42:1497-1501.
  157. Etminan M, Brophy JM, Collins G, Nazemipour M, Mansournia MA. To adjust or not to adjust: The role of different covariates in cardiovascular observational studies. *Am Heart J.* 2021;237:62-67.
  158. Richiardi L, Bellocchio R, Zugna D. Mediation analysis in epidemiology: Methods, interpretation and bias. *Int J Epidemiol.* 2013;42:1511-1519.
  159. Pardridge WM. Blood-brain barrier and delivery of protein and gene therapeutics to brain. *Front Aging Neurosci.* 2020;11:373.
  160. Hansson O, Kumar A, Janelidze S, et al. The genetic regulation of protein expression in cerebrospinal fluid. *EMBO Mol Med.* 2023;15:e16359.
  161. Paternoster L, Tilling K, Davey Smith G. Genetic epidemiology and Mendelian randomization for informing disease therapeutics: Conceptual and methodological challenges. *PLoS Genet.* 2017;13:e1006944.
  162. Burgess S, Davies NM, Thompson SG. Bias due to participant overlap in two-sample Mendelian randomization. *Genet Epidemiol.* 2016;40:597-608.

# Efficacy made Convenient



TYSABRI SC injection with the potential to administer **AT HOME** for eligible patients\*

Efficacy and safety profile comparable between TYSABRI IV and SC<sup>†1,2</sup>

<sup>†</sup>Comparable PK, PD, efficacy, and safety profile of SC to IV except for injection site pain.<sup>1,2</sup>

**CLICK HERE TO DISCOVER MORE ABOUT  
TYSABRI SC AND THE DIFFERENCE IT MAY  
MAKE TO YOUR ELIGIBLE PATIENTS**

Supported by



A Biogen developed and funded JCV antibody index PML risk stratification service, validated and available exclusively for patients on or considering TYSABRI.



\*As of April 2024, TYSABRI SC can be administered outside a clinical setting (e.g. at home) by a HCP for patients who have tolerated at least 6 doses of TYSABRI well in a clinical setting. Please refer to section 4.2 of the SmPC.<sup>1</sup>

TYSABRI is indicated as single DMT in adults with highly active RRMS for the following patient groups:<sup>1,2</sup>

- Patients with highly active disease despite a full and adequate course of treatment with at least one DMT
- Patients with rapidly evolving severe RRMS defined by 2 or more disabling relapses in one year, and with 1 or more Gd+ lesions on brain MRI or a significant increase in T2 lesion load as compared to a previous recent MRI

Very common AEs include nasopharyngitis and urinary tract infection. Please refer to the SmPC for further safety information, including the risk of the uncommon but serious AE, PML.<sup>1,2</sup>

**Abbreviations:** **AE:** Adverse Event; **DMT:** Disease-Modifying Therapy; **Gd+:** Gadolinium-Enhancing; **HCP:** Healthcare Professional; **IV:** Intravenous; **JCV:** John Cunningham Virus; **MRI:** Magnetic Resonance Imaging; **PD:** Pharmacodynamic; **PK:** Pharmacokinetic; **PML:** Progressive Multifocal Leukoencephalopathy; **RRMS:** Relapsing-Remitting Multiple Sclerosis; **SC:** Subcutaneous.

**References:** 1. TYSABRI SC (natalizumab) Summary of Product Characteristics. 2. TYSABRI IV (natalizumab) Summary of Product Characteristics.

Adverse events should be reported. For Ireland, reporting forms and information can be found at [www.hpra.ie](http://www.hpra.ie). For the UK, reporting forms and information can be found at <https://yellowcard.mhra.gov.uk/> or via the Yellow Card app available from the Apple App Store or Google Play Store. Adverse events should also be reported to Biogen Idec on [MedInfoUKI@biogen.com](mailto:MedInfoUKI@biogen.com) 1800 812 719 in Ireland and 0800 008 7401 in the UK.