

Proteogenomics in cerebrospinal fluid and plasma reveals new biological fingerprint of cerebral small vessel disease

Received: 16 September 2024

Accepted: 3 October 2025

Published online: 20 November 2025

 Check for updates

A list of authors and their affiliations appears at the end of the paper

Cerebral small vessel disease (cSVD) is a leading cause of stroke and dementia with no specific treatment, of which molecular mechanisms remain poorly understood. To identify potential biomarkers and therapeutic targets, we applied Mendelian randomization to examine over 2,500 proteins measured in plasma and, uniquely, cerebrospinal fluid, in relation to magnetic resonance imaging (MRI) markers of cSVD in more than 40,000 individuals. Here we show that 49 proteins are associated with MRI markers of cSVD, most prominently in cerebrospinal fluid. We highlight associations that are consistent across platforms and ancestries, and supported by complementary observational analyses, and we explore differences between fluids. The proteins are enriched in pathways related to the extracellular matrix, immune response and microglial activity. Many also associate with stroke and dementia, and several correspond to existing drug targets. Together, these findings reveal a robust biological fingerprint of cSVD and highlight opportunities for biomarker and drug discovery and repositioning.

Characterized by changes in the structure and function of small brain vessels, cSVD is a leading cause of ischemic and hemorrhagic stroke, cognitive decline and dementia. cSVD is extremely common with increasing age and most often ‘covert’, namely detectable on brain magnetic resonance imaging (MRI) in the absence of clinical stroke. Covert cSVD is associated with changes in cognitive performance, gait, balance and mood disturbances and portends a considerably increased risk of developing stroke and dementia, thus representing a major target to prevent these disabling conditions and promote healthier brain aging¹. The most common and heritable MRI markers of covert cerebral small vessel disease (MRI-cSVD) are white matter hyperintensities of presumed vascular origin (WMHs) and perivascular spaces (PVSs)².

While hypertension is the strongest known risk factor for cSVD¹, vascular risk factors explain only a limited fraction of MRI-cSVD variability in older age³, and drugs specifically targeting pathological processes underlying cSVD are lacking. Genomics can provide a strong foundation for mechanistic studies and drug target discovery⁴. Recent genetic studies identified >70 genetic risk loci for cSVD^{5,6}, but causal

genes and pathways remain poorly understood. As disease occurrence reflects the complex interplay of factors beyond DNA sequence, there is growing interest in identifying circulating biomarkers for clinical use, such as proteins, capturing these downstream factors, to unravel the underlying biology and accelerate omics-driven drug discovery⁷. While large-scale proteomic investigations have recently been conducted for stroke and dementia^{7–13}, so far cSVD proteome studies were performed on limited sets of proteins, in small studies of European ancestry ($N < 5,000$)^{14–18}, and only associations with plasma proteins were explored. We hypothesize that, while plasma may provide easy-access biomarkers, cerebrospinal fluid (CSF), the fluid circulating in PVSs, could reveal a more accurate biological fingerprint of cSVD.

We used two-sample Mendelian randomization (2SMR), leveraging large proteomic and genomic resources, to investigate the relation of circulating protein levels in CSF and plasma with WMHs and PVS burden. We then conducted a multipronged follow-up of identified associations in independent samples, across fluids, proteomic platforms, ancestries and the lifespan, using both 2SMR and

✉ e-mail: cruchagac@wustl.edu; stephanie.debette@inserm.fr

individual-level data from observational studies. We also tested the relation of cSVD-associated proteins with risk of stroke and dementia, and deciphered cell types and pathways involved using single-cell sequencing resources. Finally, we combined our results with pharmacological databases for proteomics-driven drug discovery.

Results

The study design is summarized in Fig. 1.

Discovery of protein–cSVD associations

We tested genetic associations of CSF and plasma protein levels with MRI-cSVD using 2SMR. We leveraged summary statistics for protein quantitative trait loci (pQTLs) in European-ancestry participants from CSF¹² ($N = 3,107$; aptamer-based SomaScan 7K assay) and plasma¹⁹ ($N = 35,559$; SomaScan 5K). *Cis*-genetic instruments were derived for 1,121 CSF and 1,731 plasma proteins, with 721 overlapping proteins (Methods and Supplementary Table 1). For MRI-cSVD, we used the largest published genome-wide association study (GWAS) in European-ancestry participants for WMH volume ($N = 48,454$, mean age 66.0 years)²⁰ and PVS burden ($N = 38,903$, 66.3 years)⁵. PVSs were studied separately in three sublocations, white matter (WM), basal ganglia (BG) and hippocampus (HIP), for which association profiles with risk factors and clinical outcomes were previously shown to differ⁵; and which are associated with other MRI-cSVD-related and cSVD-related clinical outcomes^{21–27}.

We identified 46 CSF proteins associated with at least one MRI-cSVD (false discovery rate (FDR)-corrected P value ($P_{\text{FDR}} < 0.05$): 24 with WMHs, and 25 with PVSs, predominantly in WM (Fig. 2, Extended Data Fig. 1 and Supplementary Tables 2 and 3). Nine plasma proteins were associated with MRI-cSVD ($P_{\text{FDR}} < 0.05$; WMH: 6, WM-PVS: 2, HIP-PVS: 1) of which four were also significantly associated in CSF (AMD, erythropoietin (EPO; WMH), paired immunoglobulin-like type 2 receptor alpha (PILRA)-M14 and PILRA-deltaTM (WM-PVS); Fig. 2c,d and Supplementary Tables 4 and 5). In both tissues, associations remained robust after sensitivity analyses, except for ACOX1 and WBP2, which showed no evidence of single-variant and fine-mapping-based colocalization and were excluded from further analyses. For three proteins (TFPI, NMT1 and FBLN3), standard single-variant colocalization analyses indicated evidence of colocalization ($PP4 \geq 0.5$). However, the fine-mapping-based colocalization using the sum of single effects (SuSiE) yielded more moderate probabilities for TFPI ($PP4 = 0.15$) and NMT1 ($PP4 = 0.48$), and a null probability for FBLN3 ($PP4 \approx 0$), suggesting a lack of strong support for colocalization in the latter case, warranting cautious interpretation of association results for these proteins (Supplementary Tables 3 and 5). In total, 49 proteins were associated robustly with MRI-cSVD, in CSF (41), plasma (4) or both (4). Three of these were associated with both WMHs and PVSs: cathepsin B (CTSB) and two soluble isoforms of PILRA—deltaTM and M14. Extending these analyses to cerebral microbleeds, we found that higher genetically determined CSF levels of APOE and APOE2 and lower plasma levels of APOE were associated with increased risk of microbleeds ($P_{\text{FDR}} < 0.05$). CSF and plasma levels of AMD and CSF levels of cystatin M showed nominally significant associations ($P < 0.05$) with microbleeds in plasma (Supplementary Table 6), in the same direction as significant associations with WMHs and WM-PVSs, respectively.

None of the single-variant pQTLs were nonsynonymous variants, which could have resulted in structural changes at the aptamer protein binding site and biased its measurement (Supplementary Table 7). Bidirectional MR ruled out reverse causation, except for an association of larger WM-PVS burden with higher PCSK9 CSF levels ($P_{\text{FDR}} = 0.011$; Supplementary Table 3). Given the known impact of hypertension on cSVD risk, we conducted sensitivity analyses using multivariable Mendelian randomization (MVMR) accounting for systolic blood pressure (SBP; Methods). Associations with MRI-cSVD were mostly unchanged, but appeared weakened and no longer significant for seven proteins

(APOE, CTSB (WMH), AP4A, BCL7A (WM-PVS), COL6A1, CTSB (BG-PVS) in CSF and PILRA-M14 (WM-PVS) in plasma), while maintaining consistent effect directions (Extended Data Fig. 2 and Supplementary Table 8).

The vast majority of protein–cSVD associations revealed previously unreported pathways. A few relate to previous cSVD GWAS findings. Two *cis*-pQTLs were associated with WMH volume at genome-wide significance²⁰: for FBLN3 (encoded by *EFEMP1*, chr2p16.1) and NMT1 (chr17q21.31), lower genetically determined plasma levels were associated with larger WMH volume. High-temperature requirement A serine peptidase 1 (HTRA1), of which genetically predicted lower plasma levels were associated with extensive HIP-PVSs, is encoded by a gene harboring both rare mutations causing monogenic cSVD²⁸ and common variants associated with ischemic stroke and WMHs^{20,29,30}.

Genetically determined levels of cSVD-associated proteins were mostly independent of each other using linkage disequilibrium score regression³¹, except EPHB4 with PILRA-M14 in plasma (genetic correlation ($r_g = 0.296$; $P < 5 \times 10^{-5}$; Methods and Extended Data Fig. 3).

Follow-up and extensions of significant protein–cSVD associations in complementary 2SMR and individual-level data settings

We implemented a multipronged follow-up approach across fluids, platforms, ancestries and lifespan (Figs. 1 and 3). For follow-up analyses, significant associations were defined as $P_{\text{FDR}} < 0.05$. For cross-ancestry and lifespan approaches, a nominal threshold of $P < 0.05$ was used due to smaller sample sizes and reduced statistical power. First, using 2SMR leveraging summary-based data, we tested whether cSVD–protein associations in CSF showed notable association in plasma, and vice versa, with less stringent multiple-testing correction than in discovery analyses, considering significant associations in the original fluid only. Thirty-seven cSVD-associated CSF proteins had plasma pQTLs available. Nine (24%) were associated with the same MRI-cSVD in plasma ($P_{\text{FDR}} < 0.05$): APOE, arylsulphatase B (ARSB), EPO, AMD, CTSS, PSMP (WMHs) and PILRA-M14, PILRA-deltaTM, KTEL1 (WM-PVS; Figs. 3a and 6 and Supplementary Table 9). Six cSVD-associated plasma proteins had CSF *cis*-pQTLs available. Four (67%) were associated with the same MRI-cSVD in CSF ($P_{\text{FDR}} < 0.05$): AMD and EPO (WMHs), and PILRA-M14 and PILRA-deltaTM (WM-PVS; Figs. 3b and 6 and Supplementary Table 10). Directions of association were mostly concordant except for EPO, APOE and PSMP.

Second, we performed a cross-platform follow-up, testing associations of plasma protein measurements (Olink Explore 3072) with WMH and PVS burden in two independent population-based studies, 3C-Dijon ($N = 1,087$; mean age 72.5 years) and the UK Biobank ($N = 5,494$; 63.5 years), using linear regression of individual-level data (Supplementary Table 11). Twenty-six of 49 cSVD-associated proteins were available and included in meta-analyses of both cohorts ($N = 6,881$; Methods). Seven proteins (27%, all identified in CSF 2SMR) showed associations with the same MRI-cSVD at $P_{\text{FDR}} < 0.05$ (ARSB, serine protease 8 (PRSS8), CTSS, CTSB, TFPI and BT3A2 (WMH), IL-6 (HIP-PVS); Figs. 3 and 6 and Supplementary Tables 12 and 13). Directionality was consistent in CSF and plasma for ARSB, CTSB and BT3A2, but not for PRSS8, TFPI, IL-6 and CTSS, all four presenting low correlation ($r < 0.025$, $P > 0.05$) and nonsignificant negative genetic correlation (for IL-6 and TFPI) between tissues (Extended Data Fig. 4 and Supplementary Tables 14 and 15). Inter-platform correlations between SomaScan and Olink were moderate to good in plasma and CSF, respectively (Methods and Supplementary Table 16)³². For proteins showing associations with MRI-cSVD at $P_{\text{FDR}} < 0.05$ using direct Olink measurements, we ran secondary analyses adjusting for SBP and stratifying on hypertension status (Methods), a key risk factor for MRI-cSVD, thus complementing the aforementioned MVMR analyses. Most associations remained nominally significant after SBP adjustment except for CTSS and TFPI with WMHs. No significant interaction with hypertension status was observed (Supplementary Table 17 and Extended Data Fig. 5).

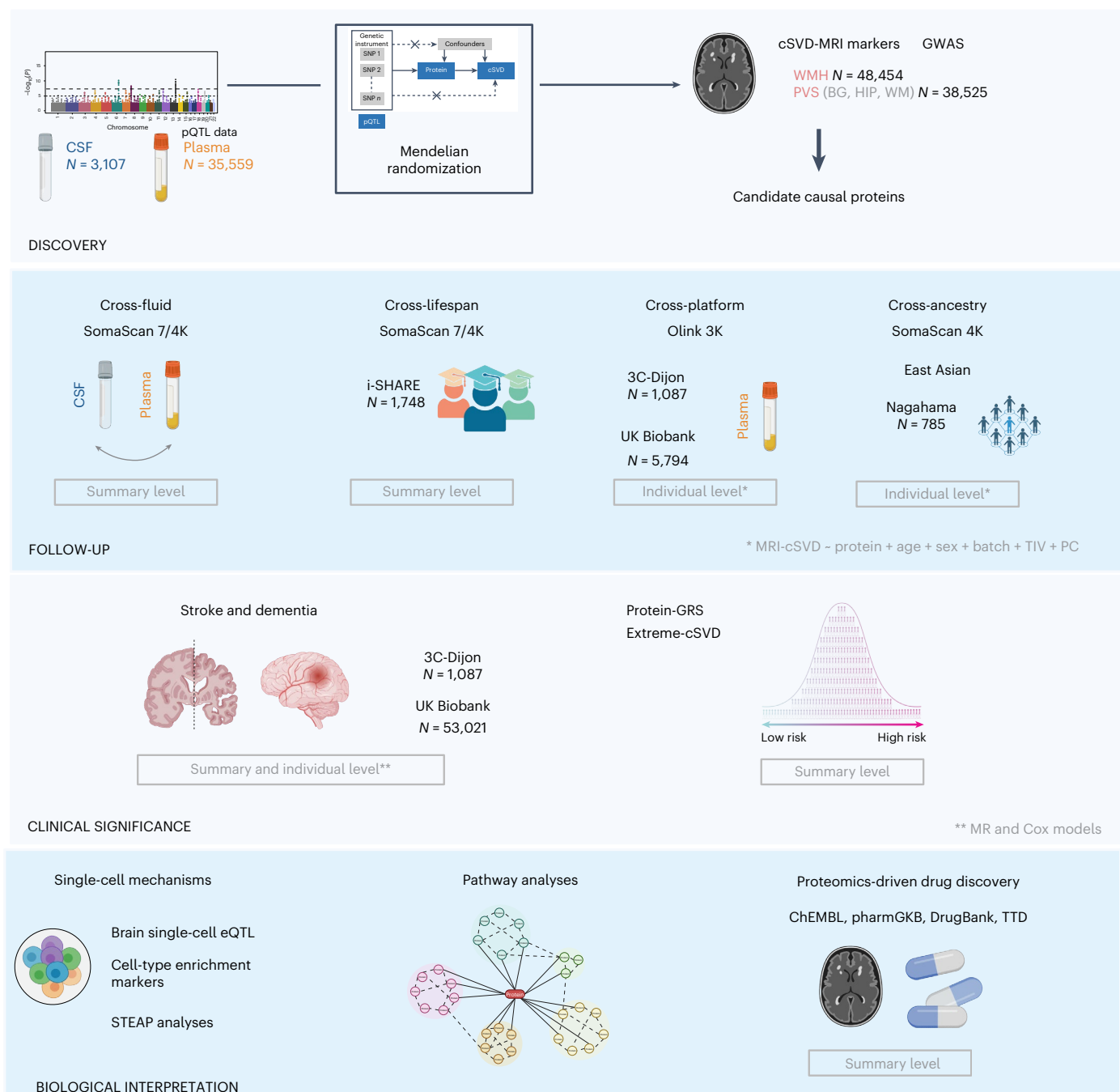


Fig. 1 | Summary of the analysis plan. ‘Summary level’ and ‘individual level’ correspond to analyses conducted using summary-level-based and individual-level-based datasets, respectively. Created with [BioRender.com](https://www.biorender.com).

Third, we conducted a cross-ancestry exploratory follow-up, testing associations of SomaScan plasma protein measurements with WMHs and PVSs using individual-level data from the Japanese population-based Nagahama study ($N = 785$; mean age 68 years). Thirty-eight of the 49 cSVD-associated proteins were available for analyses. Two proteins (identified in CSF in Europeans) were associated with the same MRI-cSVD with consistent directionality at $P_{\text{FDR}} < 0.05$ (ERO1B and PCSK9 (WM-PVS)). Nominal associations were observed for four additional proteins (BT2A1, CTSS, TNC and PSMP (WMH)); Figs. 3 and 6 and Supplementary Table 18).

Fourth, we took an exploratory lifespan approach by testing the relation of cSVD-associated proteins measured in older adults with MRI-cSVD in young adults (Internet-based Students Health Research Enterprise (i-Share) study, $N = 1,748$; mean age 22.1 years). Using the

same *cis*-pQTLs as in the discovery phase, we investigated how genetically regulated protein levels might influence early phenotypic manifestations of cSVD. Consistent with findings in older adults, higher genetically determined CSF levels of PILRA-M14 and PILRA-deltaTM were associated with larger WMH volume at $P_{\text{FDR}} < 0.05$. In addition, higher CSF GPNMB:CD and GPNMB:ECD levels were associated with extensive BG-PVS and higher Toll-like receptor 1 (TLR1):ECD CSF levels with larger WMH volume, in a direction consistent with older adults (Fig. 3 and Supplementary Table 19).

Overall, of 49 cSVD-associated proteins identified using 2SMR, (i) 16 CSF proteins showed associations with the same MRI-cSVD in plasma in at least one follow-up analysis at $P_{\text{FDR}} < 0.05$, with consistent directionality across fluids for 10, of which 3 also showed lifespan effects (CTSS, PILRA-deltaTM, PILRA-M14); (ii) 24 CSF proteins were

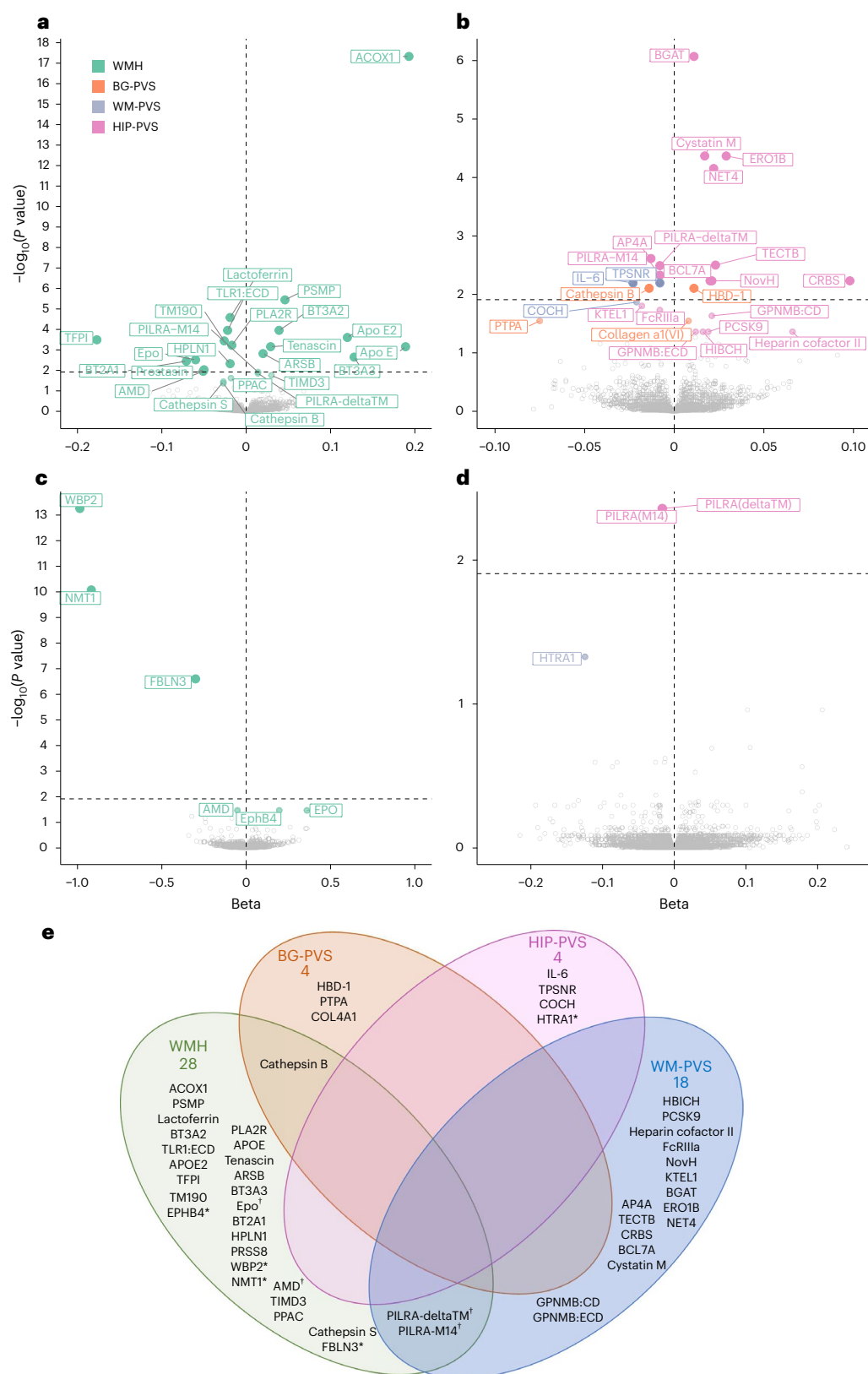


Fig. 2 | Discovery protein-cSVD associations in CSF and plasma using *cis*-pQTL MR. **a**, Volcano plots of proteins associated with WMHs using *cis*-pQTL MR in CSF. **b**, Volcano plots of proteins associated with PVS burden using *cis*-pQTL MR in CSF. **c**, Volcano plots of proteins associated with WMHs using *cis*-pQTL MR in plasma. **d**, Volcano plots of proteins associated with PVSs using *cis*-pQTL MR in plasma. Each dot represents the MR results for proteins using inverse-variance weighted (IVW) analysis when multiple instrumental variables available, or the Wald ratio when only one instrument was available. Benjamini–Hochberg

P_{FDR} values are represented in this graph. Represented proteins are significantly associated with MRI marker at P_{FDR} (Benjamini–Hochberg FDR threshold) < 0.05 . The dashed line in each volcano plot represents the corrected threshold after additionally correcting for the number of phenotypes tested ($P < 0.0125$). **e**, Venn diagram of identified causal proteins associated with MRI-cSVD. An asterisk denotes proteins identified in plasma; a dagger symbol denotes proteins associated in both plasma and CSF; other proteins are associated in CSF.

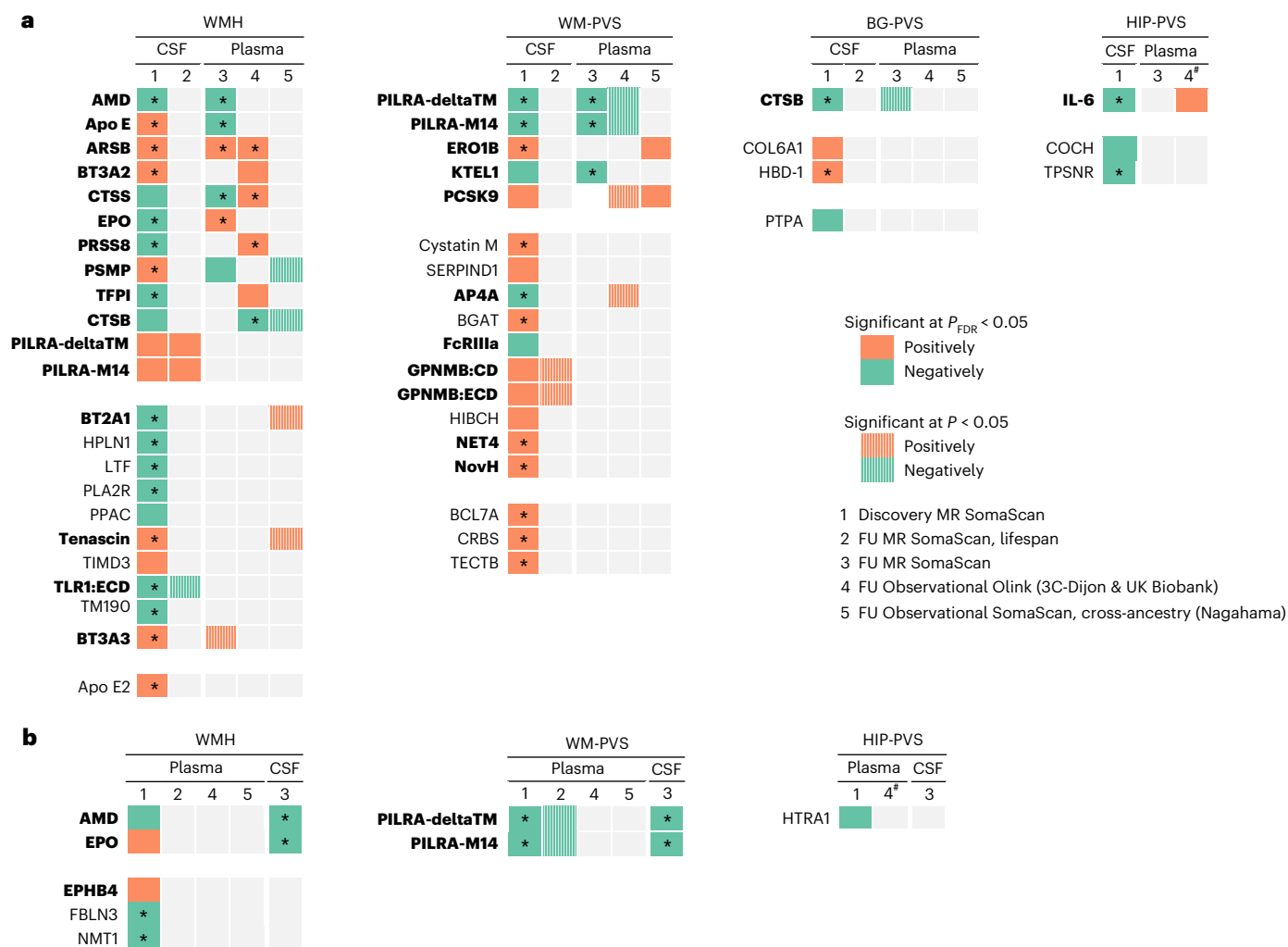


Fig. 3 | Summary of protein–cSVD associations in discovery and follow-up (lifespan, cross-fluid, cross-platform and cross-ancestry) analyses. a, Heatmap of protein–cSVD associations using CSF-based MR analyses as discovery. **b**, Heatmap of protein–cSVD associations using plasma-based MR analyses as discovery. 1. Discovery MR using *cis*-pQTLs from CSF (**a**) and plasma (**b**) and testing their association with WMH volume and PVS burden in the largest meta-analysis of GWAS. 2. Lifespan follow-up MR using *cis*-pQTLs from CSF (**a**) and plasma (**b**) and testing their association with WMHs and PVSs in young adults (i-Share study). 3. Cross-fluid follow-up MR using *cis*-pQTLs from plasma (**a**) and CSF (**b**) and testing their association with WMHs and PVSs in the other fluid than the discovery findings. 4. Cross-platform follow-up using plasma individual-level proteomic data measured with the Olink platform in independent European-ancestry samples (3C-Dijon and UK Biobank studies). 5. Cross-ancestry follow-up using plasma individual-level proteomic data measured with the SomaScan platform in an independent East Asian-ancestry sample (Nagahama study).

Columns 1, 2 and 3 represent the direction of effect size from IVW or Wald ratio MR depending on the number of instrumental variables. Columns 4 and 5 represent the direction of effect size from two-sided linear regression analyses adjusted for age, sex, batch, total intracranial volume for WMHs and the first four principal components of population stratification. Dark squares correspond to significant results after FDR correction ($P_{FDR} < 0.05$). The asterisk corresponds to significant associations after additional correction for the four phenotypes tested ($*P_{FDR} < 0.0125$). Hatched squares correspond to nominal associations ($P < 0.05$). Orange squares correspond to a positive association (higher protein levels being associated with higher cSVD burden) and green squares correspond to a negative association. FU, follow-up. The hash symbol denotes results of the 3C-Dijon analysis only. Proteins in bold are those showing at least one nominal association ($P < 0.05$) in the same direction in follow-up analyses. The exact P values are reported in Supplementary Tables 2 and 9 (**a**) and 4 and 10 (**b**).

not associated with the same MRI-cSVD marker in plasma ($P \geq 0.05$) and may be considered CSF specific, with 3 of these showing lifespan effects (GPNMB:CD, GPNMB:ECD, TLR1:ECD); (iii) 3 plasma proteins were not associated with the same MRI marker in CSF, and may be considered plasma specific; and (iv) 5 CSF proteins and 1 plasma protein had no follow-up data available^{30,33,34} (Supplementary Table 20 and Figs. 3 and 6).

Relation of cSVD-associated proteins with stroke and dementia
To assess the clinical implications of the 49 cSVD-associated proteins, we explored their relation with stroke and dementia, using 2SMR and observational survival analyses with individual-level plasma protein measurements (Methods).

For 2SMR, we leveraged the aforementioned CSF and plasma pQTLs and European-ancestry summary statistics of GWAS for stroke and its subtypes ($N \leq 73,652$ cases) and dementia ($N = 71,880$ cases; Methods). Twenty-four proteins (49%) showed associations with at least one clinical outcome at $P < 0.05$ (Figs. 4 and 6). At $P_{FDR} < 0.05$, eight CSF proteins (APOE, PILRA-M14, PILRA-deltaTM, FcRIIIa, BGAT, PLA2R, TIMD3, TPSNR) and four plasma proteins (EPHB4, HTRA1, PILRA-M14, PILRA-deltaTM) were associated with dementia, while one CSF protein (BGAT) and one plasma protein (FBLN3) were associated with any stroke and ischemic stroke (Supplementary Tables 20–22).

For observational survival analyses, 1,087 and 53,021 participants with 40/84 and 1,400/1,471 incident stroke (any)/dementia (all-cause) cases were available in the 3C-Dijon and UK Biobank population-based

cohorts, with measurements for 24 of 49 cSVD-associated proteins. Association statistics from Cox cause-specific models were meta-analyzed (Methods; $N = 54,108; 1,440/1,555$ incident stroke and dementia cases). Fourteen proteins showed an association with risk of stroke or dementia at $P < 0.05$, of which 11 were at $P_{\text{FDR}} < 0.05$. Four of the latter also reached $P < 0.05$ in 2SMR analyses (PILRA, for stroke and dementia, and BT2A1, BT3A2 and IL-6 for stroke), while 7 were significant only in observational analyses (5 for stroke: ARSB, COCH, EPO, PPAC and PRSS8; 3 for dementia: AP4A, PRSS8 and tenascin; Fig. 4c and Supplementary Tables 20 and 23).

Overall, 30 cSVD-associated proteins (61%) were associated at $P < 0.05$ and 18 at $P_{\text{FDR}} < 0.05$ with stroke or dementia (both for PILRA and PRSS8). Fine-Gray sensitivity analyses were consistent with the main findings, except for APOE, of which the observed effect was attenuated and no longer significant ($P > 0.05$; Supplementary Table 24). All proteins associated with dementia were significant only for vascular or mixed dementia subtypes in secondary analyses (Methods), except for PILRA, whose association with dementia appeared to be driven by Alzheimer's disease (Supplementary Table 25 and Extended Data Fig. 6).

Nineteen of 49 cSVD-associated proteins were available for follow-up in East Asian participants in relation with stroke using 2SMR, leveraging plasma pQTLs from BioBank Japan (BBJ; $N = 2,886$) and an East Asian stroke GWAS ($N \leq 17,493$). Despite a substantially smaller sample size for exposure and outcome than in the European ancestry, correlation of effect sizes was moderate to high (Extended Data Fig. 7). Higher plasma levels of NovH, an extracellular matrix (ECM)-associated protein involved in cardiovascular development, were associated with increased risk of small vessel stroke at $P_{\text{FDR}} < 0.05$ (Supplementary Table 26).

Biological interpretation

In pathway enrichment analyses using FUMA, cSVD-associated proteins were enriched in proteoglycan binding and ECM proteins ($P_{\text{FDR}} < 0.05$; Supplementary Table 27a). Among cSVD-associated proteins in CSF, proteins involved in immune response activation and signaling regulation were overrepresented ($P_{\text{FDR}} < 0.05$; Supplementary Table 27a).

To explore cell specificity of protein-cSVD associations, we conducted single-cell enrichment analyses using the single-cell-type enrichment analysis for phenotypes (STEAP) tool, leveraging publicly available single-cell sequencing resources (Methods and Supplementary Table 28). Genes encoding cSVD-associated proteins were significantly enriched in microglia for several CSF proteins (BT2A1, BT3A2, BT3A3, CTSS, HIBCH) and in immune cells for plasma protein EPO (Supplementary Table 29 and Extended Data Fig. 8). Next, we used single-nucleus RNA sequencing (RNA-seq) in up to 443 postmortem brain samples (dorsolateral prefrontal cortex) from the ROSMAP population-based cohort^{26–29}. Cell-type-specific brain expression quantitative trait loci (eQTLs) could be derived for 19 and 10 genes encoding cSVD-associated proteins in nonvascular and vascular brain cells, respectively (Methods and Supplementary Tables 30 and 31). Using 2SMR, three associations were observed ($P_{\text{FDR}} < 0.05$) with evidence for colocalization: lower expression of *TLRI* in oligodendrocytes and *CTSS* in smooth muscle cells was associated with larger WMH volume (same direction as in CSF); higher expression of *ABO*, encoding BGAT, in pericytes was protective for extensive WM-PVSS (opposite direction

compared to CSF). Genes encoding cSVD-associated proteins showed a nonsignificant trend toward enrichment in pericytes and significant enrichment in a microglial state type overrepresented in processes such as ribosome biogenesis, amyloid fibril formation and regulation of T cell-mediated immunity (Extended Data Fig. 9)³⁵.

Proteomics-driven drug discovery

We used MR estimates from the 49 cSVD-associated proteins to support drug discovery. Using public drug databases (Methods), we curated drugs (commercialized for other indications or under investigation in clinical trials) targeting these proteins in a direction compatible with beneficial therapeutic effects against cSVD. We identified such drugs for EPO, lactoferrin (LTF), TFPI and EPHB4 for WMHs; GPNMB and PCSK9 for WM-PVSS and COL6A1 for BG-PVSS (Figs. 5 and 6 and Supplementary Tables 20 and 32). Some of these proteins have predicted or experimentally demonstrated interactions with each other (Fig. 2e), suggesting that identified drugs may impact related pathways. Of note, although this may not necessarily be required to treat cSVD, drugs targeting EPO and LTF as agonists and EPHB4 as inhibitors cross the blood-brain barrier (BBB; Supplementary Table 32). The association of plasma PCSK9 with WM-PVSS was independent of low-density lipoprotein cholesterol and triglyceride levels in UK Biobank (Supplementary Table 12b).

Results of protein-cSVD associations along with clinical significance, enrichment analyses and drug target identification are summarized in Supplementary Table 20 and Fig. 6.

Discussion

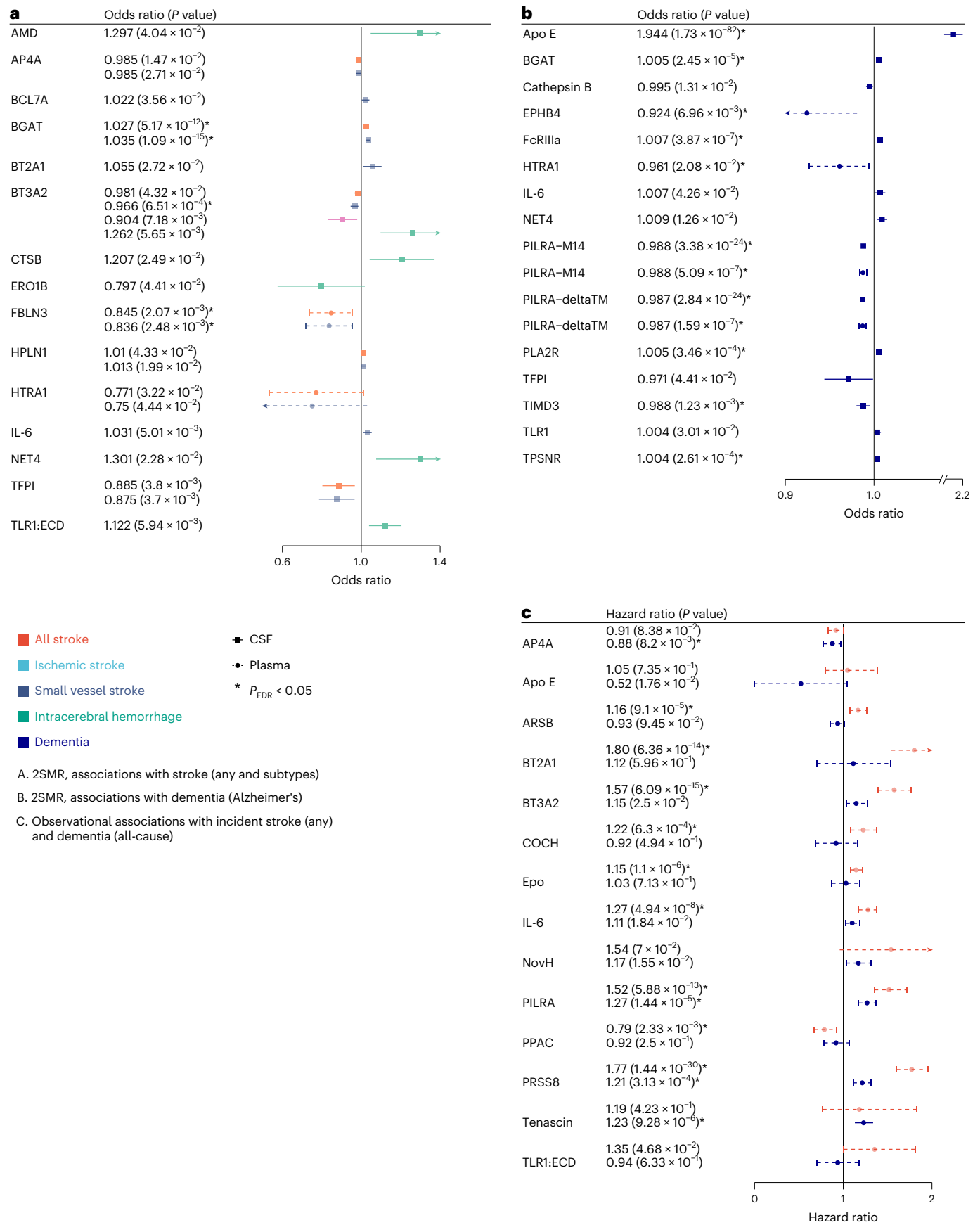
We describe a comprehensive biological fingerprint of cSVD, comprising 49 protein-cSVD associations, predominantly in the CSF, by integrating unique CSF and plasma pQTL resources with the largest GWAS of MRI-cSVD in an MR framework. We implemented a multipronged follow-up approach, across fluids, proteomic platforms and ancestries, using both MR and observational analyses. Sixteen proteins were associated with MRI-cSVD in both CSF and plasma, several of which were both in Europeans and East Asians, while 24 and 3 proteins were associated in CSF or plasma only. Pathway and cell-type enrichment analyses suggest an important role of ECM and immune response pathways, with single-cell RNA-seq analyses showing enrichment in microglia and a specific microglial state. Strikingly, several cSVD-associated proteins involved in immune response regulation already showed associations with MRI-cSVD at age 20 with consistent directionality. Over half of cSVD-associated proteins showed at least nominal associations with stroke, dementia or both, highlighting their clinical relevance. Importantly, our findings also provide genetic support for repositioning of drugs targeting seven cSVD-associated proteins in a direction compatible with beneficial therapeutic effects.

While a few proteins are encoded by genes within cSVD GWAS loci (NMT1, FBLN3, HTRA1, APOE, TFPI), shedding new light into underlying biological mechanisms^{20,33}, most associations are distinct from those previously reported. IL-6 was previously associated with WMH volume¹⁷, but its association with HIP-PVSS was not shown before. Earlier cSVD proteomic studies focused on limited protein panels^{36,37}, mostly in plasma^{14–18} and on smaller cohorts ($N < 1,000$)³⁸, with one recent study on 16 CSF proteins³⁹. Here, we analyzed >2,000 plasma and CSF proteins

Fig. 4 | Clinical significance of protein-cSVD findings in CSF and plasma.

a, Forest plot of protein-cSVD associations with stroke ($N = 73,652/1,234,808$) and its subtypes (ischemic stroke, $N = 62,100/1,234,808$; small vessel stroke, $N = 6,811/1,234,808$; and intracerebral hemorrhage, $N = 1,545/1,481$) using IVW or Wald ratio MR. **b**, Forest plot of protein-cSVD association with Alzheimer's disease ($N = 71,880/383,378$) using IVW or Wald ratio MR. **c**, Forest plot of protein-cSVD association with stroke and dementia using IVW meta-analysis of two-sided cause-specific Cox models adjusting for age, sex, self-reported ancestry and educational attainment (for incident dementia) of 3C-Dijon and

UK Biobank studies ($N = 54,108; 1,440/1,555$ incident stroke and dementia cases). All proteins associated with MRI-cSVD identified in the discovery analysis in CSF and plasma were used for this analysis. Full lines represent proteins measured in CSF. Dashed lines represent proteins measured in plasma. Proteins associated at least at $P < 0.05$ for at least one of the outcomes tested are represented (for stroke, associations with all (sub)types are represented when one or more was significant). Asterisks denote results that are significant after multiple-testing correction ($P_{\text{FDR}} < 0.05$). Dots correspond to effect estimates (odds ratio (**a** and **b**) and hazard ratio (**c**)), and errors bars correspond to 95% confidence intervals.



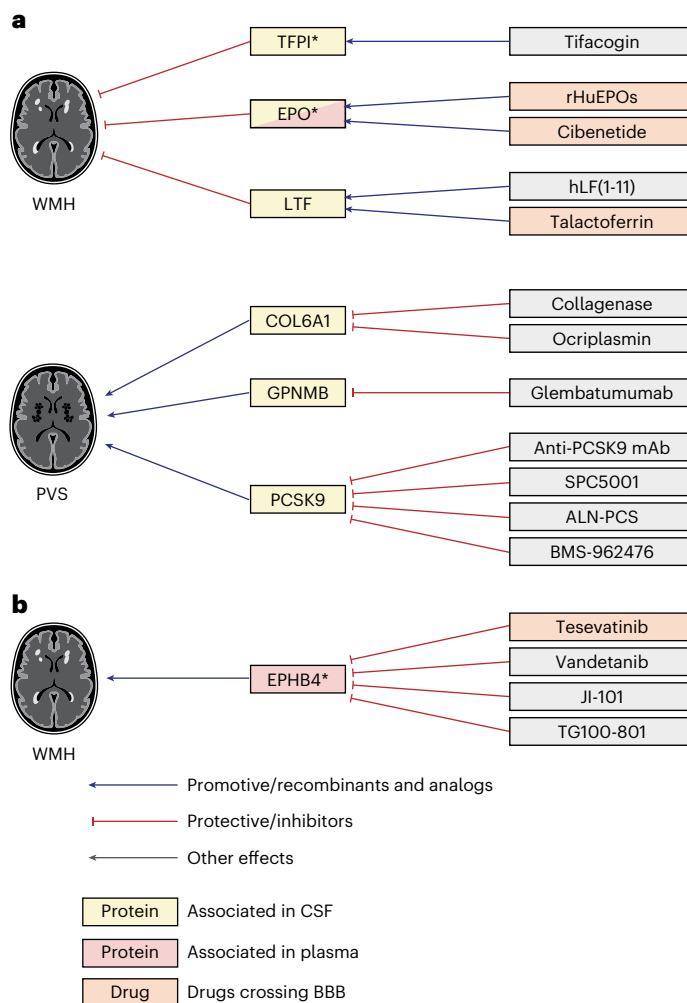


Fig. 5 | Proteomics-driven drug discovery. a, Drug-discovery analysis conducted using CSF protein-cSVD MR IVW or Wald ratio estimates for WMH and PVS findings. **b**, Drug-discovery analysis conducted using plasma protein-cSVD MR IVW or Wald ratio estimates for WMHs. Proteins in yellow correspond to proteins associated with the MRI-cSVD marker in CSF and in red in plasma, in discovery analyses. An asterisk denotes proteins with associations in at least one of the follow-up modalities (at $P < 0.05$). Red arrows correspond to a protective effect of a protein on MRI-cSVD (reducing cSVD burden) or an inhibitory effect of a drug on the cSVD-associated protein; blue arrows correspond to a deleterious effect of a protein on MRI-cSVD (promoting cSVD burden) or an analog effect of a drug on the cSVD-associated protein. Drugs in orange cross the BBB. mAb, monoclonal antibody.

with *cis*-pQTLs in >40,000 participants. CSF biomarkers have become crucial for understanding neurodegenerative and neuroinflammatory mechanisms given their proximity to the central nervous system^{40–42}, and our findings suggest this also applies to neurovascular diseases like cSVD. CSF-based MR revealed five times more protein-cSVD associations than plasma-based MR, despite a tenfold smaller sample size. Among proteins with pQTLs in both compartments, 67% of plasma cSVD-associated proteins were also associated in CSF, whereas only 24% of CSF-associated proteins showed plasma associations; adding follow-up with individual-level plasma measurements yielded 43% concordance between protein-cSVD associations in CSF and plasma, suggesting some protein-cSVD associations are CSF specific, as seen in other neurological disorders^{12,13}.

Some protein-cSVD associations were particularly robust and consistent across fluids, platforms, 2SMR and observational analyses, especially PILRA-deltaTM, PILRA-M14 and CTSB, associated with WMHs

and PVSs, and ARSB, associated with WMHs. Notably, the association of one PILRA isoform with WM-PVS and of CTSB with WMHs and BG-PVS lost significance in MVMR adjusting for SBP. CTSB, whose gene is located at a known SBP risk locus⁴³, remained associated with WMHs in observational analyses after blood pressure adjustment, suggesting additional mechanisms.

PILRA is a microglial immunoreceptor involved in amyloid- β (A β) uptake and herpes simplex virus 1 infection⁴⁴. SomaScan measures soluble PILRA isoforms lacking the transmembrane domain⁴⁵ (PILRA-deltaTM and PILRA-M14), while Olink detects the full protein. Higher CSF levels of PILRA-M14 and PILRA-deltaTM (2SMR, SomaScan) were associated with larger WMH volume, whereas higher CSF and plasma PILRA-M14/deltaTM (2SMR, SomaScan) and plasma PILRA (observational, Olink) were associated with smaller WM-PVS burden, potentially reflecting a protective effect on the cerebral amyloid angiopathy subtype of cSVD, of which WM-PVS is a marker⁴⁶. Consistent with WMH associations, higher plasma PILRA was associated with increased risk of incident stroke and dementia (observational), but in 2SMR higher plasma and CSF PILRA-M14/deltaTM were associated with lower dementia risk. Possible explanations include isoform-specific effects or differences in dementia definitions (2SMR: Alzheimer's disease; observational: all-cause dementia). Previous work supported PILRA as the likely causal gene at the chr7q21 Alzheimer's disease risk locus, suggesting a common missense variant (*rs1859788*, $r^2 = 0.3$ with PILRA pQTLs) protects against Alzheimer's disease via reduced inhibitory signaling in microglia and lower herpes simplex virus 1 infection during recurrence⁴⁷.

CTSB is a cerebrovascular matrisome protein identified in brain microvessels³⁷. This lysosomal cysteine protease is involved in proteolysis of ECM components and enhanced vessel wall permeability⁴⁸, and in proteolysis of amyloid precursor protein, implicated in Alzheimer's disease⁴⁹. Using 2SMR, lower CTSB levels in CSF were associated with larger WMHs and BG-PVSs, replicating in plasma and importantly also across platforms (2SMR and observational) and ancestries (Europeans and East Asians). Lower CTSB levels were also significantly associated with higher risk of dementia (2SMR). Mutations in *CTSA* (encoding cathepsin A) cause a rare monogenic autosomal recessive cSVD known as CARASAL⁵⁰. Our findings now expand the involvement of cathepsins to complex cSVD.

ARSB plays an important role in ECM degradation, regulation of neurite outgrowth and neuronal adaptability in the brain⁵¹, where it is expressed predominantly in microglia^{52,53}. While ARSB deficiency causes a lysosomal storage disorder⁵⁴, here higher CSF and plasma ARSB levels were associated with greater WMH volume (2SMR and observational), making ARSB a compelling candidate to explore as a potential biomarker of cSVD. This will require absolute quantification assays and testing of the predictive and prognostic potential of circulating ARSB levels.

Our proteogenomic analyses thus lend support to a prominent role of the matrisome in cSVD, corroborating and expanding findings from large genomic studies^{5,6} and preclinical work on monogenic cSVD³⁷, by revealing numerous matrisome proteins not previously implicated. PRSS8, associated with WMH, also shows a highly significant association with risk of incident stroke and dementia in observational analyses. Moreover, while matrisome protein HTRA1 is known to play a central role in cSVD, both monogenic⁵⁵ and multifactorial^{30,34}, our results reveal an association of lower HTRA1 plasma levels with extensive HIP-PVS, and, at nominal significance, with dementia. This expands recent descriptions of loss-of-function mechanisms for HTRA1 associations with ischemic stroke and coronary artery disease risk^{30,34}.

Our findings reveal associations of immune response pathways with MRI-cSVD. Enrichment in immune regulation was prominent for CSF proteins, and integration with single-cell data highlighted microglial cells, the brain's primary resident immune cells, particularly a microglial state overrepresented in T cell-mediated immunity and amyloid fibril formation. This provides molecular insights into immune

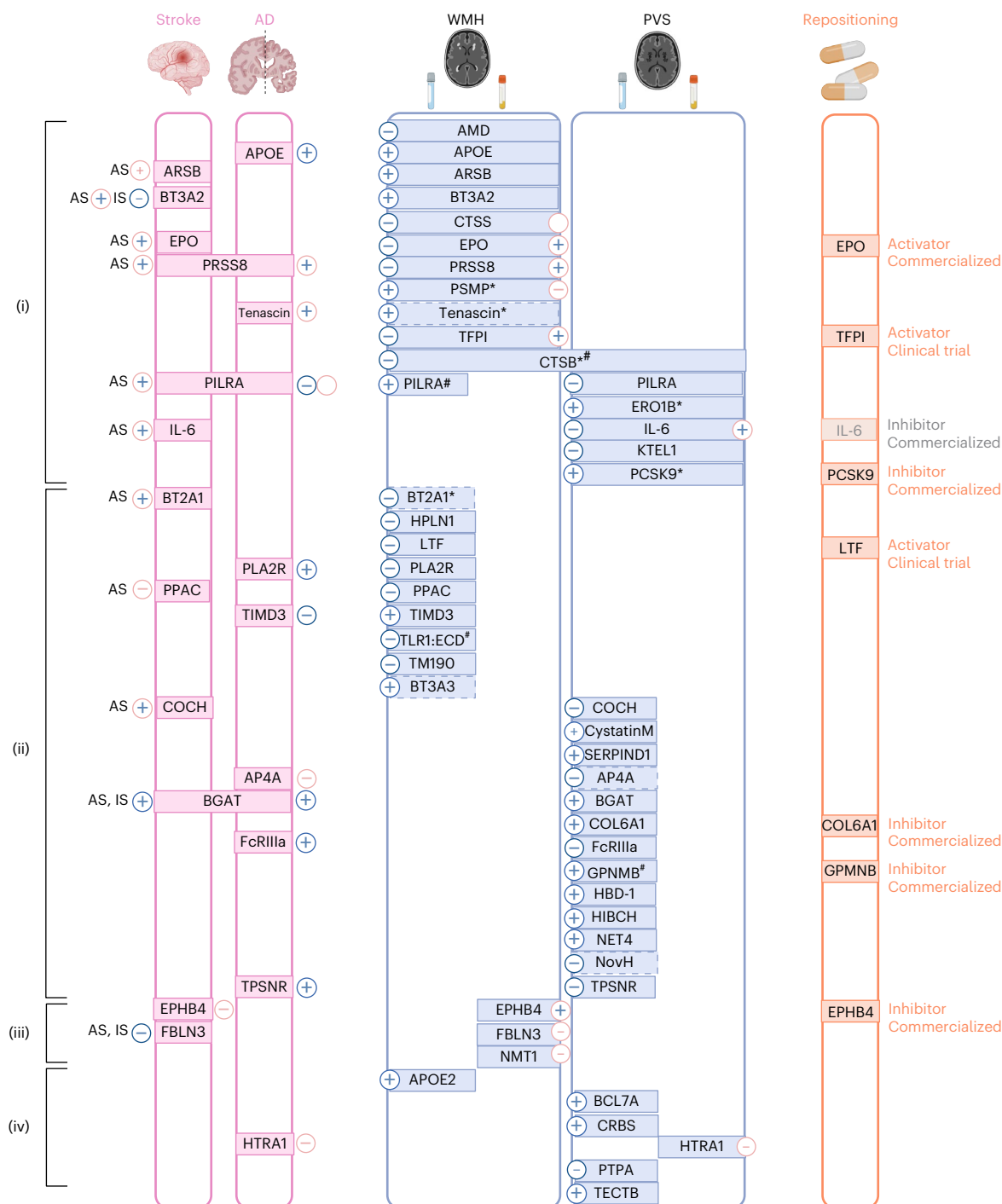


Fig. 6 | Integrated summary of our findings. Proteins associations with WMH, PVS or both are represented in the middle. For each MRI marker, the left side corresponds to CSF findings and the right side to plasma findings. An asterisk denotes proteins with cross-ancestry association. A hash symbol denotes proteins with lifespan association. Associations with stroke, dementia or both ($P_{FDR} < 0.05$) in either MR or observational analysis are represented on the left of the figure. Subtypes of stroke are as follows: AS, any stroke; IS, ischemic stroke. Minus and plus signs correspond to the direction of association referring to higher level of the protein. Blue plus or minus signs correspond to findings in CSF and pink in plasma. Empty plus or minus signs correspond to a situation where

opposite directions were observed in the same tissue using MR and observational study. Drug repositioning is represented on the right of the figure. (i) Proteins associated with the same MRI-cSVD marker in cross-fluid follow-up (for cSVD-associated proteins identified in CSF discovery: showing significant association in plasma follow-up; for cSVD-associated proteins identified in plasma discovery: showing significant association in CSF follow-up); (ii) CSF-specific proteins (showing no significant association in plasma follow-up); (iii) plasma-specific proteins (showing a nonsignificant association in CSF follow-up); (iv) no follow-up available. AD, Alzheimer's disease. Created with [BioRender.com](https://www.biorender.com).

activation in cSVD, complementing smaller biomarker studies and MRI–histopathology data linking WMHs to microglial activation^{56–61}. Notably, our results suggest that immune regulation could be one of the earliest processes involved in cSVD, as demonstrated for

Alzheimer's disease⁶². Of all cSVD-associated proteins, PILRA isoforms (microglial immunoreceptor), TLR1 (involved in activation of innate immunity) and GPNMB (transmembrane glycoprotein upregulated upon tissue damage and inflammation) associations were already

detectable in young adults in their twenties, in directions consistent with older adults.

This work unveiled emerging prospects for drug repositioning for cSVD, with the identification of multiple drugs targeting seven cSVD-associated proteins (EPO, LTF, TFPI, EPHB4, COL6A1, GPNMB and PCSK9). Based on the protective effect of higher genetically determined CSF EPO levels on WMH volume, potential repositioning of EPO analogs crossing the BBB and studied in phase II clinical trials for other indications (depression, neuropathy) is a compelling example. EPO, besides stimulating erythropoiesis, is a neuroprotective protein acting through the Keap1–Nrf2 pathway to protect from ischemia–reperfusion injury⁶³ and by stimulating neurogenesis. Neuroprotective effects are conserved for EPO derivatives lacking erythropoiesis-stimulating activity⁶⁴. In contrast with CSF EPO, high plasma EPO levels were associated with larger WMH volume. This likely reflects distinct sources and regulation of EPO in CSF and plasma. While the major EPO-producing cells in the brain are pericytes^{65,66}, most EPO is produced in the kidney. As EPO does not cross the BBB, plasma EPO likely mostly reflects the kidney production. LTF, of which higher genetically determined CSF levels were associated with smaller WMH volume, also has neuroprotective and anti-inflammatory properties^{67,68} and, interestingly, shows strong protein–protein interactions and collaborative anti-inflammatory properties with EPO⁶³. LTF agonists are currently tested in phase I and III trials for sepsis and cancer. Optimized versions of EPO and LTF have shown experimental evidence of neuroprotective effects in ischemic stroke and intracerebral hemorrhage^{69–71}. Other repositioning candidates for cSVD include drugs inhibiting PCSK9. In addition to significant 2SMR associations between higher CSF PCSK9 levels and larger WM-PVS burden, higher plasma PCSK9 levels were also associated with extensive WM-PVS in observational analyses in the East Asian Nagahama cohort and the UK Biobank, independent of lipid levels in the latter. A protective effect of PCSK9 inhibitors on any ischemic stroke has been demonstrated⁷² but has not been shown specifically for cSVD. Experimental work has shown PCSK9 to regulate A β clearance from the brain⁷³, and peripheral PCSK9 inhibition to reduce A β pathology in the prefrontal cortex and HIP in mice⁷⁴. Intriguingly, bidirectional MR also showed an association of genetically determined larger WM-PVS burden with higher CSF PCSK9 levels. Although speculative, this could potentially reflect an influence of glymphatic dysfunction, of which PVSs are thought to be a marker, on PCSK9 clearance from the brain⁷⁴.

As reported by others^{11,12}, for six proteins we observed opposite directionality of associations for plasma versus CSF protein levels (APOE, EPO, PSMP, PRSS8 and TFPI with WMHs, IL-6 with HIP-PVS, and BT3A2 with stroke). As for EPO, differences in protein production and regulation between CSF and blood, or exchanges from CSF to plasma, may explain distinct effects on cSVD¹¹. This underscores the value of multi-compartment proteomic analyses and warrants further study of underlying mechanisms.

We acknowledge limitations. Discovery was restricted to proteins quantified by SomaScan, for which *cis*-pQTLs could be derived, representing less than 10% of known proteins. Moreover, we were underpowered to conduct discovery association analyses with cerebral microbleeds, but performed exploratory analyses with WMH- and PVS-associated proteins. pQTLs were derived from a population enriched in neurologically impaired individuals (especially with Alzheimer's disease); however, we previously showed that pQTLs are only marginally influenced by disease status¹²; moreover, follow-up samples were not enriched in individuals with Alzheimer's disease. The unique CSF proteomics resource we used was crucial to derive a biological fingerprint of cSVD; however, no additional CSF proteogenomics resource was available for follow-up. Nevertheless, the multipronged follow-up and extension across fluids, platforms, lifespan and ancestries enhances the robustness of our findings and their transportability to East Asian populations where cSVD is particularly prevalent⁷⁵. Four plasma protein–cSVD associations discovered using 2SMR either were

not significant in follow-up observational analyses or had no follow-up available. While in some instances this may reflect spurious associations, it could also be explained by lack of power or modest correlation across platforms as previously reported^{32,76,77}. In fact, two of these proteins showed significant associations with stroke or dementia (FBLN3, HTRA1) in consistent direction and three were located within cSVD GWAS loci (FBLN3, HTRA1, NMT1), supporting the robustness of these findings. Inconsistent directionality of significant associations between pQTL analyses (SomaScan) and direct measurements (Olink) for two proteins (CTSS with WMH and PILRA with dementia) requires further exploration but could reflect that distinct isoforms are being captured as suggested by others^{32,76,77}. Importantly, associations of genetically proxied CTSS and PILRA levels with MRI-cSVD were consistently observed in plasma and CSF, and across the lifespan, in independent datasets, supporting the robustness of these results. We also acknowledge that our lifespan approach is limited by the use of pQTL data derived from older populations due to the current lack of genetic instruments available for younger individuals, although international efforts to address this gap are ongoing. Single-cell analyses were conducted independently on each cell-type and protein, as we are not equipped to assess how these different cell types and proteins may interact with each other. When larger single-cell eQTL datasets become available, future studies should address these questions.

In conclusion, our large-scale proteogenomic study provides a comprehensive *in vivo* biological fingerprint of cSVD, with 49 protein–cSVD associations, mostly in the CSF. The results highlight important biological processes underlying cSVD at the molecular and cellular levels and point to early life mechanisms involving immunity and inflammation. They also pave the way for deriving circulating biomarkers and drug repositioning for cSVD with concrete development opportunities, an important step forward for a highly prevalent condition with no specific biomarker and treatment to date.

Methods

This study complies with all relevant ethical regulations, and all participants gave written, informed consent (Supplementary Methods).

Discovery of protein–cSVD associations

Deriving genetic instruments for circulating protein levels (instrumental variables for the exposure) using pQTLs. pQTLs were generated from GWAS of circulating protein levels. CSF pQTL summary statistics were obtained from 7,028 proteins (SomaScan 7K platform; $N = 3,107$, European ancestry); 1,076 participants were cognitively normal, 1,001 had clinically determined late-onset Alzheimer's disease, 118 had early-onset Alzheimer's disease, 281 non-Alzheimer's disease dementia and 631 had Parkinson's disease¹². Plasma pQTL summary statistics were obtained from 4,907 proteins (SomaScan 5K platform; $N = 35,559$ European ancestry, cognitively normal) from either the Icelandic cancer project (52%) or deCODE genetics (48%)¹⁹. *cis*-pQTLs were defined as genetic variants within 1 Mb of the corresponding protein-coding gene. Genetic variants were selected based on genome-wide significant associations ($P < 5 \times 10^{-8}$) with protein abundance after clumping using PLINK2 (ref. 78) for linkage disequilibrium at $r^2 < 0.01$, within 1 Mb using the European 1000 Genome reference panel. Genetic variants included in the major histocompatibility complex region (chromosome 6: 26–34 Mb) were removed considering the complex linkage disequilibrium structure of the region. The strength of the instrumental variables was measured using the *F*-statistic (*F*-statistic > 10 was considered strong). Following these steps, we selected up to 1,121 CSF and 1,732 plasma proteins with *cis*-acting pQTLs for MR analyses.

Genetic associations with MRI-cSVD (outcome). We used summary statistics from the latest GWAS meta-analyses of WMH volume, in 48,454 participants (mean age 66.0 years), and of extensive PVS

burden in WM, BG and HIP, in up to 38,903 participants (mean age 68.3 years), from the general population, of European ancestry, and free of stroke^{5,20}. In exploratory analyses, we examined the relation of identified WMH- or PVS-associated proteins with cerebral microbleeds leveraging the only available GWAS, with limited statistical power ($N = 25,862$; 3,556 cases)⁷⁹. Cohorts from which the pQTLs were derived were not included in WMH, PVS or microbleed GWAS meta-analyses.

Analytical steps for MR analyses. We applied 2SMR analyses using the ‘TwoSampleMR’ package (v.0.5.6)⁸⁰ to assess the causal association between genetically predicted CSF and plasma protein levels and MRI-cSVD. pQTLs obtained after instrument selection for each protein were used as instrumental variables. We extracted the association estimates between the variants and the exposures or the outcomes and aligned the effect alleles. For proteins with multiple instrumental variables, we computed MR estimates with random-effect IVW analysis⁸¹ relying on distinct assumptions for validity: (i) Heterogeneity across the MR estimates was assessed for each instrument using Cochran’s Q statistic ($P < 0.05$ was considered significant)⁸¹; (ii) Horizontal pleiotropy was assessed using the MR-Egger intercept as a measure of directional pleiotropy ($P < 0.05$ was considered significant)⁸². We further conducted various sensitivity analyses⁸³:

- (1) The identification of outlier instrumental variables and their removal from analyses was conducted using MR pleiotropy residual sum and outlier (MR-PRESSO)⁸⁴ ($P < 0.05$ was considered significant).
- (2) Reverse MR was run by reversing the direction of inference, using the MRI-cSVD markers as the exposure and proteins as the outcome, to formally rule out reverse causation.
- (3) MR-Egger regression⁸⁵ and weighted median, which are more robust to the use of pleiotropic instruments, were used. When pleiotropy was observed, we retained results when at least two of the three sensitivity methods (MR-Egger, weighted median, MR-PRESSO) were concordant with each other and $P < 0.05$.
- (4) MVMR^{86,87} estimating the direct effect of multiple exposures, that is, simultaneously including in the same model genetic instruments for protein levels and for SBP to rule out confounding of protein-cSVD associations in SBP GWAS in European ancestry participants ($N = 757,601$)⁴³.

For proteins with a single instrumental variable, we computed MR estimates using the Wald ratio, followed by colocalization analyses using coloc⁸⁸, under the assumption of a single causal variant per trait, including variants ± 1 Mb surrounding the pQTL of interest. Associations were considered significant when the posterior probability H_4 (PPH4; shared association with single causal variant) was ≥ 0.70 and suggestive for PPH4 > 0.50 (ref. 89). Associations with PPH4 < 0.50 were removed from further analyses. As a complementary approach, we used SuSiE colocalization⁹⁰ (susieR package) to perform fine-mapping based on z-scores and linkage disequilibrium matrices using pQTLs in sample linkage disequilibrium matrix, allowing for the presence of multiple signals.

Discovery MR results were considered significant when passing the FDR Benjamini–Hochberg-corrected significance threshold ($P_{\text{FDR}} < 0.05$). In sensitivity analyses, we additionally corrected for the number of independent phenotypes tested, estimated using correlations between traits in the 3C-Dijon study applying the matrix spectral decomposition (matSpDLite⁹¹) method for WMH volume and each PVS location ($P_{\text{FDR}} < 1.2 \times 10^{-2}$; 0.05/4).

Correlation of identified cSVD-associated proteins in plasma and CSF. Genetic correlation. Genetic correlation analyses were conducted using linkage disequilibrium score regression based on pQTL summary statistics. This approach aimed to (i) differentiate between shared genetic regulation and independent signals, (ii) assess biological

coherence and clustering, and (iii) support the interpretation of 2SMR results. All proteins were tested; however, only those with sufficient single nucleotide polymorphism (SNP)-heritability estimates were retained for display, as low heritability values led to convergence issues. Genetic correlation could be reliably estimated for 24 proteins in CSF and 9 in plasma. $P < 4 \times 10^{-5}$ was used, correcting for the number of proteins tested and three situations: CSF-CSF, CSF-plasma and plasma-plasma; $0.05/(24 \times 24) \times 2 + (9 \times 9)$.

Inter-platform and intra-platform correlation. A subset of 259 ACE participants (Supplementary Methods) was analyzed in two independent experiments in CSF using the aptamer-based SomaScan 7K proteomic platform (SomaLogic). We considered the dataset with the adaptive normalization by maximum likelihood method for further analysis^{10,11}. Additionally, another aliquot of these samples was analyzed in CSF with the antibody-based Olink Explore 3072 Panel measuring over 2,900 proteins (Olink Proteomics)¹². Using this highly characterized sample, we conducted intra-platform (comparing SomaScan assays in CSF) and inter-platform (comparing SomaScan and Olink Explore in CSF) correlation analyses. We categorized proteomic measures in a single metric (1–9) accounting for reproducibility and reliability⁹². In another subset of 258 participants matched by draw date, protein-level correlations between CSF and plasma measured with SomaScan 7K at the same date were assessed.

Follow-up of significant protein-cSVD associations

Cross-platform follow-up (direct protein measurements, Olink, plasma). Protein-cSVD associations were followed up in two population-based cohorts with MRI phenotypes and plasma proteomics (Olink Explore 3072). In 3C-Dijon, 1,087 participants aged < 80 years (72.5 ± 4.1 years; 60.5% women) were included after quality control and exclusion of prevalent stroke/dementia; baseline plasma samples were profiled using proximity extension assay, following the manufacturer’s protocol⁹³ at McGill Genome Center (Montreal, Canada), measuring 2,923 unique proteins. WMH volume was estimated from multimodal MRI (T1, T2, DP; 1.5T Siemens Magnetom Scanner), and PVS burden in BG and WM was assessed using the SHIVA-PVS algorithm⁹⁴ and a validated visual scale in HIP⁹⁵. In the UK Biobank, 5,494 participants had Olink data (field ID: 1839) and brain MRI data (63.5 ± 7.9 years; 53.5% women); WMH volume and BG/WM PVS burden were estimated as in 3C-Dijon using T1-weighted images from the subset of participants with proteomics data (Supplementary Methods). All participants provided informed consent. Data preprocessing including plate-based normalization, and quality-control checks were conducted according to standardized Olink protocols.

We conducted linear and logistic regression of proteins with WMHs and PVSs adjusted for the delay between age at blood draw and age at the time of MRI, sex, batch effect, total intracranial volume (or mask volume for WMH in 3C-Dijon). WMHs and PVSs in BG and WM were inverse-normal transformed and PVSs in HIP values were dichotomized, comparing participants in the top quartile of PVS burden distribution to the rest, as previously described⁵. Data distribution was assumed to be normal but this was not formally tested. Individual data points are shown in Extended Data Fig. 10 to illustrate data distribution.

An inverse-variance weighted meta-analysis was performed using the metafor R package⁹⁶. The heterogeneity of associations across studies was assessed using the Cochran–Mantel–Haenszel statistical test. Associations with $P > 1.9 \times 10^{-3}$ (0.05/26, correcting for 26 proteins available for follow-up) were considered. Significant associations were defined by $P_{\text{FDR}} < 0.05$. In addition, results of sensitivity analyses at $P_{\text{FDR}} < 1.2 \times 10^{-2}$ are displayed, accounting for the four phenotypes tested.

Associations of plasma protein levels with WMHs and PVSs were examined stratifying by hypertension status in 3C-Dijon and the UK Biobank, followed by meta-analysis. Hypertension was defined as

SBP ≥ 140 mm Hg and diastolic blood pressure ≥ 90 mm Hg or use of antihypertensive medication (3C-Dijon: 235 hypertensive/852 non-hypertensive; UKB: 2,088/3,406). Direct analyses were further adjusted for SBP.

Protein–protein Spearman correlations were assessed in the UK Biobank using the `corrplot` R package, with significance at Bonferroni-corrected $P < 7.7 \times 10^{-5}$ ($0.05/(26 \times 26) - 26$).

Cross-ancestry follow-up (direct protein measurements, SomaScan, plasma). A subset of 858 participants with brain imaging and plasma proteomic data from the Nagahama study, a prospective population-based cohort study initiated in 2007 in Nagahama, Japan ($N = 10,082$ at baseline, median age: 57.3 (41.6–64.7) years, 68% women), were used⁹⁷ (Supplementary Methods). WMHs in Nagahama was generated using UBO detector⁹⁸. PVS burden was estimated using the aforementioned machine-learning-based SHIVA-PVS algorithm^{5,94}. Quality-control checks and proteomic measurement transformation (log2) were conducted according to standardized SomaScan protocols. After excluding participants for whom the estimation of the MRI marker was not possible, without proteomics measurements, with prevalent stroke, who had missing covariates or who had withdrawn their consent, a total of 785 participants were available for association analyses. We conducted linear regression for WMHs, WM-PVSs and BG-PVSs as continuous variables that were inverse-normal transformed and adjusted for age, sex, batch, total intracranial volume and the first four principal components. Associations at $P < 0.05$ were reported given the exploratory nature of these cross-ancestry analyses on a much smaller sample size.

Follow-up across the lifespan (pQTLs, SomaScan, plasma and CSF). We conducted 2SMR analyses using the aforementioned pQTLs in plasma and CSF (instruments) and GWAS for WMHs and PVSs (outcomes). WMH and PVS GWAS were conducted in the i-Share study, an ongoing prospective population-based cohort study of French-speaking students⁹⁹. We used a subsample of 1,748 participants aged 18–35 years, recruited in Bordeaux, France, for whom both brain MRI and genome-wide genotype data were available (mean age: 22.1 ± 2.3 years; 72.2% women)^{100–102}. All participants provided informed consent, and MRI-Share participants received compensation of 40 euros. MRI protocol, genetic data quality-control and imputation procedures are detailed elsewhere^{5,100–102}. For i-Share PVS GWAS summary statistics, we used previously described data⁵. For i-Share WMH GWAS summary statistics, we performed GWAS using the genome-wide linear mixed model implemented in REGENIE on WMH volume quantified using a recently developed algorithm¹⁰³ (after excluding eight participants with multiple sclerosis or radiologically isolated syndrome)¹⁰⁴. WMH volume was transformed using an indirect inverse-normal transformation (applying inverse-normal transformation to residuals from linear regression of WMHs adjusted for covariates (age at MRI, sex, total intracranial volume and the first four principal components of population stratification)). These analyses were restricted to SNPs with an imputation score > 0.5 and a minor allele frequency > 0.01 . Associations at $P < 0.05$ were reported given the exploratory nature of these cross-ancestry analyses on a much smaller sample size.

Clinical significance

MR. Summary statistics for any ($N = 73,652$), ischemic ($N = 62,100$) and small vessel ($N = 6,811$) stroke were derived from the GIGASTROKE study including 1,234,808 controls¹⁰⁵ and the largest GWAS for intracerebral hemorrhage (1,545 patients¹⁰⁶). For dementia, we used summary statistics of the largest GWAS for Alzheimer's disease comprising 71,880 Alzheimer's disease cases¹⁰⁷, including clinically diagnosed cases, and based on self-reported parental history as a proxy for diagnosis. We considered associations at $P < 0.05$ and reported significant findings at $P_{FDR} < 0.05$.

Observational survival analysis. We explored the relation of individual plasma protein levels (Olink Explore 3072) with incident stroke (any) and dementia (all-cause) in the UK Biobank and 3C-Dijon longitudinal population-based cohorts. After quality control and exclusion of prevalent stroke and dementia cases, $N = 53,021$ and $N = 1,087$ participants with plasma protein measurements were available in the UK Biobank and 3C-Dijon, of whom 1,400 and 1,471 developed stroke and dementia in the UK Biobank and 40 and 84 in 3C-Dijon, respectively. Event ascertainment in each cohort is detailed in Supplementary Methods. Cause-specific Cox models accounting for competing risk of death were used to explore associations with incident stroke and dementia, using age as a timescale, and adjusting for batch, sex and self-reported ancestry, and additionally educational attainment for associations with incident dementia. Analyses were then meta-analyzed using an inverse-variance weighted meta-analysis using the `metafor` R package⁹⁶. The heterogeneity of associations across studies was assessed using the Cochran–Mantel–Haenszel statistical test, and associations with $P > 2.1 \times 10^{-3}$ ($0.05/24$, correcting for 24 proteins available for follow-up) were considered. Significant associations were defined by $P_{FDR} < 0.05$ and suggestive by $P < 0.05$.

We performed sensitivity analyses using a Fine–Gray subdistribution hazard model. These analyses were conducted in the UK Biobank, representing the largest contribution to the meta-analysis ($N = 53,021$), using the `survival` R package (version 3.8-3).

Associations with dementia subtypes. In the UK Biobank, dementia subtypes were defined using algorithmically determined outcomes (field: 42022), while in 3C-Dijon, they were based on clinical diagnoses over 12 years of follow-up. This resulted in 102 participants with vascular or mixed dementia and 376 with Alzheimer's disease in 3C-Dijon, and 283 and 731 cases with vascular dementia and Alzheimer's disease, respectively, in the UK Biobank. As described above, cause-specific Cox models were performed, adjusting for the same covariates, and subsequently meta-analyzed.

Cross-ancestry association analysis with stroke. We conducted 2SMR analyses in BBJ (first cohort study¹⁰⁸), which recruited around 200,000 participants across 66 hospitals in Japan between 2003 and 2007. Proteomic profiling was conducted for a total of 2,886 individuals of East Asian ancestry from two previous studies^{109,110} with whole-genome sequencing datasets (Olink Explore 3072; mean age: 62.4 ± 14.5 years; 46.9% women). Data preprocessing and quality control were conducted according to standardized Olink protocols. Rank-based inverse-normal transformation was applied to protein-level measurements. pQTL summary statistics of serum protein levels were obtained for 19 available proteins (of the 49 cSVD-associated proteins from the discovery analysis) by meta-analyzing (using METAL¹¹¹, inverse-variance weighted method; fixed-effect model) summary statistics generated in individuals from each study separately using REGENIE (v3.2.9)¹⁰⁴ (adjusted for age, sex, age-squared, age \times sex, age-squared \times sex, batch and the first ten genotype principal components). Summary statistics of GWAS for ischemic ($N = 17,493$), large-artery atherosclerotic ($N = 1,322$), cardioembolic ($N = 747$) and small vessel ($N = 4,876$) stroke were obtained in the BBJ first cohort using REGENIE v3.2.9 (adjusted for age, sex and the first ten genotype principal components), excluding the samples used for proteomic profiling. Genotyping, quality control and imputation for BBJ samples used in the stroke GWAS were conducted as previously described¹¹², except that the imputation was performed using a reference panel combining the 1000 Genome Project phase 3 v5 reference panel and 3,256 Japanese samples (JEWEL3k) samples¹¹³. Individuals without any type of stroke or cerebral aneurysm were used as controls. Instrument selection and MR were conducted following the methods previously described using a P threshold for clumping of 1×10^{-6} (Supplementary Methods).

Biological interpretation

Pathway enrichment analysis. The GENE2FUNC analysis tool in FUMA (v1.5.4) was used to conduct gene-set enrichment analyses and detect significantly associated Gene Ontology biological processes¹¹⁴. GENE2FUNC uses a hypergeometric test to assess the overrepresentation of genes within predefined gene sets. The gene IDs used correspond to coding genes of identified proteins. We tested enrichment of the entire set of genes encoding cSVD-associated proteins identified in CSF and plasma, using the background set of genes encoding proteins tested for MR in each tissue, respectively. Benjamini–Hochberg multiple-testing correction was applied to these results ($P < 0.05$).

STEAP enrichment analysis. We performed a cell-type enrichment analysis using the STEAP tool (<https://github.com/erwinerdm/STEAP/>). This tool serves as an extension to CELLECT and integrates stratified LD score regression, MAGMA and H-MAGMA for enrichment analysis. pQTL summary statistics from the CSF and plasma datasets were preprocessed. Subsequently, expression specificity profiles were computed using single-cell RNA-seq data from human and mouse databases, including PsychENCODE DER-22, GSE67835, GSE101601, DroNc Human Hippocampus, Allen Brain Atlas MTG and LNG, Mousebrain, Tabula Muris, Descartes Human Cerebrum and Cerebellum. Cell-type enrichment analysis was conducted using MAGMA, H-MAGMA (which incorporates chromatin interaction profiles from human brain tissues in MAGMA) and stratified LD score regression. P values were Bonferroni-corrected for the number of independent cell types in each database.

Brain single-cell QTLs. Mapping of brain single-cell eQTLs was described elsewhere^{115–117}. Briefly, single-nucleus RNA-seq libraries were prepared from dorsolateral prefrontal cortex (dPFC) of 424 participants from the ROSMAP cohort using the 10x Genomics Single Cell 3' kit. Sequencing reads were processed and a unique molecular identifier count matrix was generated using Cell Ranger software (ver.6.0.0, 10x Genomics). Classification of cell types was performed by clustering cells by gene expression using the R package Seurat (ver. 4). A 'Pseudobulk' gene expression matrix was constructed by aggregating unique molecular identifier counts of the same cell type of the same donor and normalizing them to the log₂ counts per million reads mapped values. Genotyping was performed by whole-genome sequencing and GATK. Mapping of *cis*-eQTLs was performed using Matrix-eQTL (ver. 2.3) for SNPs within 1 Mb of transcription start sites. Due to the sparsity of vascular cells in brain tissue, a specific dataset from ROSMAP using in silico vasculature enrichment was used for eQTL and expression analysis. Single-nucleus RNA-seq libraries were prepared from brain samples of 409 ROSMAP participants using the 10x Genomics Single Cell 3' Kit (Supplementary Methods). Microglia states were defined from 152,459 microglial transcriptomes across 443 individuals (217 with Alzheimer's disease and 226 controls) identifying 12 transcriptional states. Microglial nuclei were obtained from post-mortem brain samples from the ROSMAP study across six brain regions (hippocampus, dPFC, mid-temporal cortex, angular gyrus, entorhinal cortex and thalamus). Using in silico sorting, 174,420 immune cells were collected from single-nucleus RNA-seq datasets using the STAR method forming 12 clusters of microglia. Those clusters were then defined as microglial states based on their molecular signature and function: MG0, hemostatic; MG1, neuronal surveillance; MG2, inflammatory I; MG3, ribosome biogenesis; MG4, lipid processing; MG5, phagocytic; MG6, stress signature; MG7, glycolytic; MG8, inflammatory II; MG10, inflammatory III; MG11, antiviral; MG12, cycling. Detailed methods regarding microglial state definitions are described elsewhere³⁵.

Proteomics-driven drug discovery

Using significant MR results from CSF and plasma, we restricted our analysis to drug-targeting proteins using four drug–gene databases

(ChEMBL, pharmGKB, DrugBank and TTD). Following this methodology, eight drug-targeting proteins were identified for WMHs (EPO, LTF, TFPI, APOE, ARSB, CTSS, CTSSB and EPHB4) and seven for PVSs (COL6A1, CTSSB, GPNMB, PCSK9, FcRIIA, heparin cofactor II, IL-6). Using public drug databases, we then curated drugs targeting those proteins in a direction compatible with a beneficial therapeutic effect against the corresponding cSVD phenotype based on MR estimates. The desired mode of action was defined as the opposite direction of the MR estimate. Once the drugs were identified, we searched the literature for a potential action of the drug.

Statistics and reproducibility

No statistical methods were used to predetermine sample sizes, but our sample sizes are similar or larger to those reported in previous publications^{12,13,32,77}. This study is based on summary statistics and observational cohort data. Randomization and blocking were not applicable as participants were not assigned to experimental groups. Data collection and preprocessing were not performed blind; however, the analyst was partially blinded to variable identities during analyses.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

We used publicly available data for analyses described in this paper, including data from the GWAS catalog (<https://www.ebi.ac.uk/gwas/>, study codes: GCST90244151, GCST011947, GCST007320, GCST90104539, GCST90162546), the DECODE project (<https://www.decode.com/summarydata/>), ChEMBL (<https://www.ebi.ac.uk/chembl/>), pharmGKB (<https://www.pharmgkb.org/>), DrugBank (<https://go.drugbank.com/>), TTD (<https://db.idrblab.net/ttd/>), CSF pQTL summary statistics available at NIAGADS (full summary statistics available to approved investigators through accession no. NG00130) and GWAS catalog (study code: GCST90421033–GCST90428040), NeuroGenomics and Informatics Center website (<https://neurogenomics.wustl.edu/open-science/raw-data/>) and ONTIME browser (<https://ontime.wustl.edu/>). Plasma proteomic data for the Knight ADRC participants are available at <https://knightadrc.wustl.edu/professionals-clinicians/request-center-resources/>. Requests for clinical or proteomic data from individual investigators will be reviewed to ensure compliance with patient confidentiality. For details on accessing available data and study protocols, see <https://knightadrc.wustl.edu/>.

Code availability

We used publicly available tools from TwoSampleMR (v0.5.686, <https://mrcieu.github.io/TwoSampleMR/>), FUMA (v1.5.4, <https://fuma.ctglab.nl/>), coloc R package (v5.2.3, <https://chr1swallace.github.io/coloc/>), STRING (v12.0, <https://string-db.org/>), FUSION (v2.7.13, https://github.com/gusevlab/fusion_twass/), PLINK (v2.0, <https://www.cog-genomics.org/plink/2.0/>), OlinkAnalyze R package (v4.3.1, <https://github.com/Olink-Proteomics/OlinkRPackage/>), MungeSumstats R package (v1.16.0, <https://github.com/neurogenomics/MungeSumstats/>), ggplot2 R package (v3.5.2, <https://github.com/tidyverse/ggplot2/>), forestplot R package (v3.1.7, <https://cran.r-project.org/web/packages/forestplot/>), Betsholtz database (<http://betsholtzlab.org/VascularSingleCells/database.html>), metafor R package (v4.8-0, <https://wviechtb.github.io/metafor/>), SHIVA-PVS algorithm (https://github.com/pboutinaud/SHIVA_PVS/, T1.PVS/v1), STEAP (<https://github.com/erwinerdm/STEAP/>). Figures were created using ggplot2 R package (<https://ggplot2.tidyverse.org/>) and BioRender.com.

References

1. Wardlaw, J. M. et al. ESO Guideline on covert cerebral small vessel disease. *Eur. Stroke J.* **6**, CXI–CLXII (2021).

2. Duering, M. et al. Neuroimaging standards for research into small vessel disease—advances since 2013. *Lancet Neurol.* **22**, 602–618 (2023).
3. Wardlaw, J. M. et al. Vascular risk factors, large-artery atheroma, and brain white matter hyperintensities. *Neurology* **82**, 1331–1338 (2014).
4. Rusina, P. V. et al. Genetic support for FDA-approved drugs over the past decade. *Nat. Rev. Drug Discov.* **22**, 864 (2023).
5. Duperron, M.-G. et al. Genomics of perivascular space burden unravels early mechanisms of cerebral small vessel disease. *Nat. Med.* **29**, 950–962 (2023).
6. Bordes, C., Sargurupremraj, M., Mishra, A. & Debette, S. Genetics of common cerebral small vessel disease. *Nat. Rev. Neurol.* **18**, 84–101 (2022).
7. Montaner, J. et al. Multilevel omics for the discovery of biomarkers and therapeutic targets for stroke. *Nat. Rev. Neurol.* **16**, 247–264 (2020).
8. Walker, K. A. et al. Large-scale plasma proteomic analysis identifies proteins and pathways associated with dementia risk. *Nat. Aging* **1**, 473–489 (2021).
9. Walker, K. A. et al. Proteomics analysis of plasma from middle-aged adults identifies protein markers of dementia risk in later life. *Sci. Transl. Med.* **15**, eadf5681 (2023).
10. Chen, L. et al. Systematic Mendelian randomization using the human plasma proteome to discover potential therapeutic targets for stroke. *Nat. Commun.* **13**, 6143 (2022).
11. Dammer, E. B. et al. Multi-platform proteomic analysis of Alzheimer's disease cerebrospinal fluid and plasma reveals network biomarkers associated with proteostasis and the matrisome. *Alzheimers Res. Ther.* **14**, 174 (2022).
12. Western, D. et al. Proteogenomic analysis of human cerebrospinal fluid identifies neurologically relevant regulation and implicates causal proteins for Alzheimer's disease. *Nat. Genet.* **56**, 2672–2684 (2024).
13. Yang, C. et al. Genomic atlas of the proteome from brain, CSF and plasma prioritizes proteins implicated in neurological disorders. *Nat. Neurosci.* **24**, 1302–1312 (2021).
14. Kuipers, S. et al. A cluster of blood-based protein biomarkers reflecting coagulation relates to the burden of cerebral small vessel disease. *J. Cereb. Blood Flow.* **42**, 1282–1293 (2022).
15. Wan, S. et al. Plasma inflammatory biomarkers in cerebral small vessel disease: a review. *CNS Neurosci. Ther.* **29**, 498–515 (2023).
16. Fornage, M. et al. Biomarkers of inflammation and MRI-defined small vessel disease of the brain: the cardiovascular health study. *Stroke* **39**, 1952–1959 (2008).
17. Satizabal, C. L., Zhu, Y. C., Mazoyer, B., Dufouil, C. & Tzourio, C. Circulating IL-6 and CRP are associated with MRI findings in the elderly: the 3C-Dijon Study. *Neurology* **78**, 720–727 (2012).
18. Jiménez-Balado, J. et al. New candidate blood biomarkers potentially associated with white matter hyperintensities progression. *Sci. Rep.* **11**, 14324 (2021).
19. Ferkingstad, E. et al. Large-scale integration of the plasma proteome with genetics and disease. *Nat. Genet.* **53**, 1712–1721 (2021).
20. Sargurupremraj, M. et al. Cerebral small vessel disease genomics and its implications across the lifespan. *Nat. Commun.* **11**, 6285 (2020).
21. Weller, R. O., Hawkes, C. A., Kalaria, R. N., Werring, D. J. & Carare, R. O. White matter changes in dementia: role of impaired drainage of interstitial fluid. *Brain Pathol.* **25**, 63–78 (2015).
22. Patankar, T. F. et al. Dilatation of the Virchow-Robin space is a sensitive indicator of cerebral microvascular disease: study in elderly patients with dementia. *AJNR Am. J. Neuroradiol.* **26**, 1512–1520 (2005).
23. Gertje, E. C., van Westen, D., Panizo, C., Mattsson-Carlsson, N. & Hansson, O. Association of enlarged perivascular spaces and measures of small vessel and Alzheimer disease. *Neurology* **96**, e193–e202 (2021).
24. Pan, P. et al. The enlarged perivascular spaces in the hippocampus is associated with memory function in patients with type 2 diabetes mellitus. *Sci. Rep.* **15**, 3644 (2025).
25. Ferguson, S. C. et al. Cognitive ability and brain structure in type 1 diabetes: relation to microangiopathy and preceding severe hypoglycemia. *Diabetes* **52**, 149–156 (2003).
26. Evans, T. E. et al. Determinants of perivascular spaces in the general population: a pooled cohort analysis of individual participant data. *Neurology* **100**, e107–e122 (2023).
27. Duperron, M.-G. et al. Burden of dilated perivascular spaces, an emerging marker of cerebral small vessel disease, is highly heritable. *Stroke* **49**, 282–287 (2018).
28. Cho, B. P. H. et al. Association of vascular risk factors and genetic factors with penetrance of variants causing monogenic stroke. *JAMA Neurol.* **79**, 1303–1311 (2022).
29. Mishra, A. et al. Association of variants in HTRA1 and NOTCH3 with MRI-defined extremes of cerebral small vessel disease in older subjects. *Brain J. Neurol.* **142**, 1009–1023 (2019).
30. Malik, R. et al. Genetically proxied HTRA1 protease activity and circulating levels independently predict risk of ischemic stroke and coronary artery disease. *Nat. Cardiovasc. Res.* **3**, 701–713 (2024).
31. ReproGen Consortium et al. An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
32. Eldjarn, G. H. et al. Large-scale plasma proteomics comparisons through genetics and disease associations. *Nature* **622**, 348–358 (2023).
33. Persyn, E. et al. Genome-wide association study of MRI markers of cerebral small vessel disease in 42,310 participants. *Nat. Commun.* **11**, 2175 (2020).
34. Malik, R. et al. Whole-exome sequencing reveals a role of HTRA1 and EGFL8 in brain white matter hyperintensities. *Brain J. Neurol.* **144**, 2670–2682 (2021).
35. Sun, N. et al. Human microglial state dynamics in Alzheimer's disease progression. *Cell* **186**, 4386–4403 (2023).
36. Pokhilko, A. et al. Global proteomic analysis of extracellular matrix in mouse and human brain highlights relevance to cerebrovascular disease. *J. Cereb. Blood Flow. Metab.* **41**, 2423–2438 (2021).
37. Joutel, A., Haddad, I., Ratelade, J. & Nelson, M. T. Perturbations of the cerebrovascular matrisome: a convergent mechanism in small vessel disease of the brain? *J. Cereb. Blood Flow.* **36**, 143–157 (2016).
38. Haffner, C. Proteostasis in cerebral small vessel disease. *Front. Neurosci.* **13**, 1142 (2019).
39. Gertje, E. C. et al. Associations between CSF markers of inflammation, white matter lesions, and cognitive decline in individuals without dementia. *Neurology* **100**, e1812–e1824 (2023).
40. Gaetani, L. et al. CSF and blood biomarkers in neuroinflammatory and neurodegenerative diseases: implications for treatment. *Trends Pharmacol. Sci.* **41**, 1023–1037 (2020).
41. Luebke, M., Parulekar, M. & Thomas, F. P. Fluid biomarkers for the diagnosis of neurodegenerative diseases. *Biomark. Neuropsychiatry* **8**, 100062 (2023).
42. Robey, T. T. & Panegyres, P. K. Cerebrospinal fluid biomarkers in neurodegenerative disorders. *Future Neurol.* **14**, FNL6 (2019).
43. Evangelou, E. et al. Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. *Nat. Genet.* **50**, 1412–1425 (2018).

44. Park, Y. H. et al. Association of blood-based transcriptional risk scores with biomarkers for Alzheimer disease. *Neurol. Genet.* **6**, e517 (2020).
45. Fournier, N. et al. FDF03, a novel inhibitory receptor of the immunoglobulin superfamily, is expressed by human dendritic and myeloid cells. *J. Immunol.* **165**, 1197–1209 (2000).
46. Charidimou, A. et al. The Boston criteria version 2.0 for cerebral amyloid angiopathy: a multicentre, retrospective, MRI–neuropathology diagnostic accuracy study. *Lancet Neurol.* **21**, 714–725 (2022).
47. Rathore, N. et al. Paired immunoglobulin-like type 2 receptor alpha G78R variant alters ligand binding and confers protection to Alzheimer's disease. *PLoS Genet.* **14**, e1007427 (2018).
48. Hook, G. et al. Cathepsin B gene knockout improves behavioral deficits and reduces pathology in models of neurologic disorders. *Pharmacol. Rev.* **74**, 600–629 (2022).
49. Hook, G., Kindy, M. & Hook, V. Cathepsin B deficiency improves memory deficits and reduces amyloid- β in hA β PP mouse models representing the major sporadic Alzheimer's disease condition. *J. Alzheimers Dis.* **93**, 33–46 (2023).
50. Bugiani, M. et al. Cathepsin A-related arteriopathy with strokes and leukoencephalopathy (CARASAL). *Neurology* **87**, 1777–1786 (2016).
51. Zhang, X. et al. Arylsulfatase B modulates neurite outgrowth via astrocyte chondroitin-4-sulfate: dysregulation by ethanol. *Glia* **62**, 259–271 (2014).
52. Vanlandewijck, M. et al. A molecular atlas of cell types and zonation in the brain vasculature. *Nature* **554**, 475–480 (2018).
53. He, L. et al. Single-cell RNA sequencing of mouse brain and lung vascular and vessel-associated cell types. *Sci. Data* **5**, 180160 (2018).
54. Valayannopoulos, V., Nicely, H., Harmatz, P. & Turbeville, S. Mucopolysaccharidosis VI. *Orphanet J. Rare Dis.* **5**, 5 (2010).
55. Mancuso, M. et al. Monogenic cerebral small-vessel diseases: diagnosis and therapy. Consensus recommendations of the European Academy of Neurology. *Eur. J. Neurol.* **27**, 909–927 (2020).
56. Solé-Guardia, G. et al. Association between hypertension and neurovascular inflammation in both normal-appearing white matter and white matter hyperintensities. *Acta Neuropathol. Commun.* **11**, 2 (2023).
57. Low, A., Mak, E., Rowe, J. B., Markus, H. S. & O'Brien, J. T. Inflammation and cerebral small vessel disease: a systematic review. *Ageing Res. Rev.* **53**, 100916 (2019).
58. Low, A. et al. In vivo neuroinflammation and cerebral small vessel disease in mild cognitive impairment and Alzheimer's disease. *J. Neurol. Neurosurg. Psychiatry* **92**, 45–52 (2021).
59. Evans, L. E. et al. Cardiovascular comorbidities, inflammation, and cerebral small vessel disease. *Cardiovasc. Res.* **117**, 2575–2588 (2021).
60. Fu, Y. & Yan, Y. Emerging role of immunity in cerebral small vessel disease. *Front. Immunol.* **9**, 67 (2018).
61. Yang, Y. et al. Epigenetic and integrative cross-omics analyses of cerebral white matter hyperintensities on MRI. *Brain* **146**, 492–506 (2023).
62. Gjoneska, E. et al. Conserved epigenomic signals in mice and humans reveal immune basis of Alzheimer's disease. *Nature* **518**, 365–369 (2015).
63. Zakharova, E. T. et al. Erythropoietin and Nrf2: key factors in the neuroprotection provided by apo-lactoferrin. *BioMetals* **31**, 425–443 (2018).
64. Gan, Y. et al. Mutant erythropoietin without erythropoietic activity is neuroprotective against ischemic brain injury. *Stroke* **43**, 3071–3077 (2012).
65. Ji, P. Pericytes: new EPO-producing cells in the brain. *Blood* **128**, 2483–2485 (2016).
66. Urrutia, A. A. et al. Prolyl-4-hydroxylase 2 and 3 coregulate murine erythropoietin in brain pericytes. *Blood* **128**, 2550–2560 (2016).
67. Spagnuolo, P. A. & Hoffman-Goetz, L. Dietary lactoferrin does not prevent dextran sulfate sodium induced murine intestinal lymphocyte death. *Exp. Biol. Med.* **233**, 1099–1108 (2008).
68. Van De Looij, Y. et al. Lactoferrin during lactation protects the immature hypoxic-ischemic rat brain. *Ann. Clin. Transl. Neurol.* **1**, 955–967 (2014).
69. Mercier, M. International approach to the assessment of chemical risks. *Sci. Total Environ.* **101**, 1–7 (1991).
70. Zhao, X. et al. Optimized lactoferrin as a highly promising treatment for intracerebral hemorrhage: pre-clinical experience. *J. Cereb. Blood Flow. Metab.* **41**, 53–66 (2021).
71. Kittur, F. S., Hung, C.-Y., Li, P. A., Sane, D. C. & Xie, J. Asialo-rhuEPO as a potential neuroprotectant for ischemic stroke treatment. *Pharmaceuticals* **16**, 610 (2023).
72. Giugliano, R. P. et al. Stroke prevention with the PCSK9 (proprotein convertase subtilisin-kexin type 9) inhibitor evolocumab added to statin in high-risk patients with stable atherosclerosis. *Stroke* **51**, 1546–1554 (2020).
73. Mazura, A. D. et al. PCSK9 acts as a key regulator of A β clearance across the blood-brain barrier. *Cell. Mol. Life Sci.* **79**, 212 (2022).
74. Mestre, H., Kostrikov, S., Mehta, R. I. & Nedergaard, M. Perivascular spaces, glymphatic dysfunction, and small vessel disease. *Clin. Sci.* **131**, 2257–2274 (2017).
75. Mok, V. et al. Race-ethnicity and cerebral small vessel disease—comparison between Chinese and White populations. *Int. J. Stroke* **9**, 36–42 (2014).
76. Wang, B. et al. Comparative studies of 2168 plasma proteins measured by two affinity-based platforms in 4000 Chinese adults. *Nat. Commun.* **16**, 1869 (2025).
77. Sun, B. B. et al. Genomic atlas of the human plasma proteome. *Nature* **558**, 73–79 (2018).
78. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* **4**, 7 (2015).
79. Knol, M. J. et al. Association of common genetic variants with brain microbleeds: a genome-wide association study. *Neurology* **95**, e3331–e3343 (2020).
80. Hemani, G., Tilling, K. & Davey Smith, G. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. *PLoS Genet.* **13**, e1007081 (2017).
81. Burgess, S., Butterworth, A. & Thompson, S. G. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet. Epidemiol.* **37**, 658–665 (2013).
82. Burgess, S. & Thompson, S. G. Interpreting findings from Mendelian randomization using the MR-Egger method. *Eur. J. Epidemiol.* **32**, 377–389 (2017).
83. Burgess, S. et al. Guidelines for performing Mendelian randomization investigations. *Wellcome Open Res.* **4**, 186 (2019).
84. Verbanck, M., Chen, C.-Y., Neale, B. & Do, R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat. Genet.* **50**, 693–698 (2018).
85. Bowden, J., Davey Smith, G. & Burgess, S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int. J. Epidemiol.* **44**, 512–525 (2015).
86. Sanderson, E., Davey Smith, G., Windmeijer, F. & Bowden, J. An examination of multivariable Mendelian randomization in the single-sample and two-sample summary data settings. *Int. J. Epidemiol.* **48**, 713–727 (2019).
87. Burgess, S. & Thompson, S. G. Multivariable Mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects. *Am. J. Epidemiol.* **181**, 251–260 (2015).

88. Giambartolomei, C. et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* **10**, e1004383 (2014).
89. Cronjé, H. T. et al. Genetic evidence implicating natriuretic peptide receptor-3 in cardiovascular disease risk: a Mendelian randomization study. *BMC Med.* **21**, 158 (2023).
90. Zou, Y., Carbonetto, P., Wang, G. & Stephens, M. Fine-mapping from summary data with the ‘Sum of Single Effects’ model. *PLoS Genet.* **18**, e1010299 (2022).
91. Li, J. & Ji, L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity* **95**, 221–227 (2005).
92. Puerta, R. et al. Head-to-Head Comparison of Aptamer- and Antibody-Based Proteomic Platforms in Human Cerebrospinal Fluid Samples from a Real-World Memory Clinic Cohort. *Int. J. Mol. Sci.* **26**, 286 (2024).
93. Lind, L. et al. Use of a proximity extension assay proteomics chip to discover new biomarkers for human atherosclerosis. *Atherosclerosis* **242**, 205–210 (2015).
94. Boutinaud, P. et al. 3D Segmentation of perivascular spaces on T1-weighted 3 Tesla MR images with a convolutional autoencoder and a U-shaped neural network. *Front. Neuroinform.* **15**, 641600 (2021).
95. Zhu, Y.-C. et al. Frequency and location of dilated Virchow-Robin spaces in elderly people: a population-based 3D MR imaging study. *AJNR Am. J. Neuroradiol.* **32**, 709–713 (2011).
96. Viechtbauer, W. Conducting meta-analyses in R with the metafor Package. *J. Stat. Softw.* **36**, 3 (2010).
97. Funada, S. et al. Longitudinal analysis of bidirectional relationships between nocturia and depressive symptoms: the Nagahama study. *J. Urol.* **203**, 984–990 (2020).
98. Jiang, J. et al. UBO Detector – a cluster-based, fully automated pipeline for extracting white matter hyperintensities. *NeuroImage* **174**, 539–549 (2018).
99. Montagni, I., Guichard, E. & Kurth, T. Association of screen time with self-perceived attention problems and hyperactivity levels in French students: a cross-sectional study. *BMJ Open* **6**, e009089 (2016).
100. Tsuchida, A. et al. Age-related variations in regional white matter volumetry and microstructure during the post-adolescence period: a cross-sectional study of a cohort of 1,713 university students. *Front. Syst. Neurosci.* **15**, 692152 (2021).
101. Tsuchida, A. et al. The MRI-Share database: brain imaging in a cross-sectional cohort of 1870 university students. *Brain Struct. Funct.* **226**, 2057–2085 (2021).
102. Le Grand, Q. et al. Genomic studies across the lifespan point to early mechanisms determining subcortical volumes. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* **7**, 616–628 (2022).
103. Tsuchida, A. et al. Early detection of white matter hyperintensities using SHIVA-WMH detector. *Hum. Brain Mapp.* **45**, e26548 (2024).
104. Mbatchou, J. et al. Computationally efficient whole-genome regression for quantitative and binary traits. *Nat. Genet.* **53**, 1097–1103 (2021).
105. Mishra, A. et al. Stroke genetics informs drug discovery and risk prediction across ancestries. *Nature* **611**, 115–123 (2022).
106. Woo, D. et al. Meta-analysis of genome-wide association studies identifies 1q22 as a susceptibility locus for intracerebral hemorrhage. *Am. J. Hum. Genet.* **94**, 511–521 (2014).
107. Jansen, I. E. et al. Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer’s disease risk. *Nat. Genet.* **51**, 404–413 (2019).
108. Nagai, A. et al. Overview of the BioBank Japan Project: study design and profile. *J. Epidemiol.* **27**, S2–S8 (2017).
109. Koyama, S. et al. Population-specific and trans-ancestry genome-wide analyses identify distinct and shared genetic risk loci for coronary artery disease. *Nat. Genet.* **52**, 1169–1177 (2020).
110. Liu, X. et al. Decoding triancestral origins, archaic introgression, and natural selection in the Japanese population by whole-genome sequencing. *Sci. Adv.* **10**, eadi8419 (2024).
111. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
112. He, Y. et al. East Asian-specific and cross-ancestry genome-wide meta-analyses provide mechanistic insights into peptic ulcer disease. *Nat. Genet.* **55**, 2129–2138 (2023).
113. Flanagan, J. et al. Population-specific reference panel improves imputation quality for genome-wide association studies conducted on the Japanese population. *Commun. Biol.* **7**, 1665 (2024).
114. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).
115. Fujita, M. et al. Cell subtype-specific effects of genetic variation in the Alzheimer’s disease brain. *Nat. Genet.* **56**, 605–614 (2024).
116. Garcia, F. J. et al. Single-cell dissection of the human brain vasculature. *Nature* **603**, 893–899 (2022).
117. Sun, N. et al. Single-nucleus multiregion transcriptomic analysis of brain vasculature in Alzheimer’s disease. *Nat. Neurosci.* **26**, 970–982 (2023).

Acknowledgements

We thank all the participants and their families, as well as the many involved institutions and their staff. This project is supported by a grant overseen by the French National Research Agency (ANR) as part of the Investment for the Future Programme ANR-18-RHUS-0002 and by the Precision and Global Vascular Brain Health Institute (VBHI) funded by the France 2030 IHU3 initiative. The project also received funding from the French National Research Agency (ANR) through the SHIVA project. Computations were performed on the Bordeaux Bioinformatics Center (CBIB) and the CREDIM computer resources, University of Bordeaux. Funding support for additional computer resources has been provided to S.D. by the Fondation Claude Pompidou. The i-Share study has received funding by the French National Agency (Agence Nationale de la Recherche, ANR), via the Investment for the Future program (grant nos. ANR-10-COHO-05 and ANR-18-RHUS-0002) and from the University of Bordeaux Initiative of Excellence (IdEX). The Three City (3C) Study is conducted under a partnership agreement among the Institut National de la Santé et de la Recherche Médicale (INSERM), the University of Bordeaux and Sanofi-Aventis. The Fondation pour la Recherche Médicale funded the preparation and initiation of the study. The 3C Study is also supported by the Caisse Nationale Maladie des Travailleurs Salariés, Direction Générale de la Santé, Mutuelle Générale de l’Éducation Nationale (MGEN), Institut de la Longévité, Conseils Régionaux d’Aquitaine and Bourgogne, Fondation de France, and Ministry of Research–INSERM Programme ‘Cohortes et collections de données biologiques’. This work was supported by grants from the National Institutes of Health (NIH; R01AG044546 (to C.C.), P01AG003991 (to C.C.), RF1AG053303 (to C.C.), RF1AG058501 (to C.C.), U01AG058922 (to C.C.), RF1AG074007 (to Y.S.), R00AG062723 (to L.I.), CCC grant R01AG059421 (to S.S.), P30 AG066546 (to A.R. and S.S.), RF1 AG063507 (to A.R. and S.S.), the Chan Zuckerberg Initiative (CZI), the Michael J. Fox Foundation (to L.I. and C.C.), the Department of Defense (W81XWH2010849, to L.I.), the Alzheimer’s Association Zenith Fellows Award (ZEN-22-848604, to C.C.) and the Bright Focus Foundation (A2021033S, to L.I.). The recruitment and clinical characterization of research participants at Washington University were supported by NIH P30AG066444 (to J. C. Morris), P01AG03991 (to J. C. Morris) and P01AG026276 (to J. C. Morris). This work was supported by access to equipment made possible by the Hope Center for Neurological

Disorders, the NeuroGenomics and Informatics Center (<https://neurogenomics.wustl.edu/>) and the Departments of Neurology and Psychiatry at Washington University School of Medicine. The Genome Research @ Ace Alzheimer Center Barcelona project (GR@ACE) is supported by Grifols, Fundació bancaria 'La Caixa', Ace Alzheimer Center Barcelona and CIBERNED. Ace Alzheimer Center Barcelona is one of the participating centers of the Dementia Genetics Spanish Consortium (DEGESCO). The FACEHBI study is supported by funds from Ace Alzheimer Center Barcelona, Grifols, Life Molecular Imaging, Araclon Biotech, Alkahest, Laboratorio de análisis Echevarne and IrsiCaixa. We acknowledge the support of the Spanish Ministry of Science and Innovation, Proyectos de Generación de Conocimiento grants PID2021-122473OA-I00, PID2021-123462OB-I00 and PID2019-106625RB-I00, ISCIII, Acción Estratégica en Salud, integrated in the Spanish National R+D+I Plan and financed by ISCIII Subdirección General de Evaluación and the Fondo Europeo de Desarrollo Regional (FEDER 'Una manera de hacer Europa') grants PI13/02434, PI16/01861, PI17/01474, PI19/00335, PI19/01240, PI19/01301, PI22/01403 and PI22/00258 and the ISCIII national grant PMP22/00022, funded by the European Union (NextGenerationEU). We acknowledge the support of CIBERNED (ISCIII) under the grants CB06/05/2004 and CB18/05/00010; the ADAPTED and MOPEAD projects, European Union/EFPIA Innovative Medicines Initiative Joint (grant numbers 115975 and 115985, respectively); from PREADAPT project, Joint Program for Neurodegenerative Diseases (JPND) grant no. AC19/00097; from HARPONE project, Agency for Innovation and Entrepreneurship (VLAIO) grant no. PR067/21 and Janssen. DESCARTES project is funded by the German Research Foundation (DFG). The Cardiovascular Health Study (CHS) research was supported by NHLBI contracts HHSN268201200036C, HHSN268200800007C, HHSN268201800001C, N01HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, N01HC85086 and 75N92021D00006, NHLBI grants U01HL080295, R01HL087652, R01HL105756, R01HL103612, R01HL120393 and U01HL130114 and R01HL172803, with additional contribution from the National Institute of Neurological Disorders and Stroke (NINDS). Additional support was provided through R01AG023629 from the National Institute on Aging. A full list of principal CHS investigators and institutions can be found at <https://chs-nhlbi.org/>. This research has been conducted using the UK Biobank Resource under application nos. 94113 and 18545. I.C. was supported by the Digital Public Health Graduate Program (DPH), a PhD program supported by the French Investment for the Future Program (grant no. 17-EURE-0019). M.H. was supported by R01NS129032. M.-G.D. was supported by an Inserm-Bettencourt CCA program. S.N. was supported by AMED (JP25tm0424228, JP25kk0305032, JP256f0137004, JP25tm0524003 and JP25tm0524009), Takeda Science Foundation and Japan Foundation for Applied Enzymology. Y. Okada was supported by JSPS KAKENHI (22H00476), AMED (JP21gm4010006, JP22km0405211, JP22ek0410075, JP22km0405217, JP22ek0109594, JP223fa627002, JP223fa627010, JP233fa627011, JP23zf0127008), JST Moonshot R&D (JPMJMS2021, JPMJMS2024), Takeda Science Foundation, Bioinformatics Initiative of Osaka University Graduate School of Medicine, Institute for Open and Transdisciplinary Research Initiatives and Center for Infectious Disease Education and Research (CiDER) and the Center for Advanced Modality and DDS (CAMaD), Osaka University. P.M.M. gratefully acknowledges personal support from the Edmond J. Safra Foundation and Lily Safra and receipt of an NIHR Senior Investigator Award. P.M.M.'s research is supported by the UK Dementia Research Institute, which is funded primarily by the UKRI Medical Research Council, and by the Imperial College Healthcare Trust NIHR Biomedical Research Centre. L.L. acknowledges the Intramural Research Program in the National Institutes on Aging. This research was supported in part by

the National Center for Advancing Translational Sciences, CTSI grant UL1TR001881, and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern California Diabetes Endocrinology Research Center. Infrastructure for the CHARGE Consortium is supported in part by the National Heart, Lung, and Blood Institute (NHLBI) grant R01HL105756. A.C. acknowledges the support of the Instituto de Salud Carlos III (ISCIII) under the grant Sara Borrell (CD22/00125) and the Spanish Ministry of Science and Innovation, Proyectos de Generación de Conocimiento grant PID2021-122473OA-I00. Support for title page creation and format was provided by AuthorArranger, a tool developed at the National Cancer Institute. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

S.D. and C.C. designed and conceived the study. C.C., D.W., J.T., D.A., P.J., M.A., M. Munter, M.L., F.M., K.M., Y.S., Y. Oda and A.K. generated the proteomic data. M.S., M.-G.D., A.M., C.T., J.I.R., B.M.P., J.C.B., W.T.L., S.S., J.M.W., M.J., F.M., K.M., Y.S., Y. Oda and A.K. generated the genomic and imaging data. I.C., D.W., C.C., J.T., P.G.-G., S.N., P.J., R.P., T.D., Y.H., G.R., M.F., P.L.D.J. and S.K. contributed to bioinformatics analyses. I.C. and S.D. wrote and edited the paper. All of the authors provided critical revision.

Competing interests

C.C. has received research support from GSK and Eisai and is a member of the advisory board of Circular Genomics and owns stocks in this company. C.C. is part of the scientific advisory board for ADmit. B.P. serves on the Steering Committee of the Yale Open Data Project funded by Johnson & Johnson. C.C. is a member of the scientific advisory board of Circular Genomics and owns stocks, and is on the scientific advisory board of ADmit and Alamar. C.C. consults for Sanofi, Novo Nordisk and Owkin. C.C. has received research support from GSK, Danaher and Eisai. P.M.M. has received an honorarium as Chair of the UKRI Medical Research Council Neuroscience and Mental Health Board until March 2024. P.M.M. acknowledges consultancy fees from Biogen, Sudo, Nimbus and GSK. P.M.M. has received speakers' honoraria from Sanofi and Redburn, and has received research or educational funds from Biogen, Merck, Bristol Myers Squibb and Nimbus. J.M.W. declares no commercial competing interests, is in receipt of various academic research grants and is chief investigator for LACunar Intervention Trials. The authors declare no competing interests with respect to research, authorship and/or publication of this article. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s43587-025-01006-w>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s43587-025-01006-w>.

Correspondence and requests for materials should be addressed to Carlos Cruchaga or Stéphanie Debette.

Peer review information *Nature Aging* thanks Shuai Yuan and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this

article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025

Ilana Caro ^{1,41}, **Daniel Western** ^{2,3,41}, **Shinichi Namba** ^{4,5,6,41}, **Na Sun**^{7,8}, **Shuji Kawaguchi** ⁹, **Yunye He** ¹⁰, **Masashi Fujita**¹¹, **Gennady Roshchupkin**¹², **Tim D'Aoust**¹, **Marie-Gabrielle Duperron**^{1,13}, **Muralidharan Sargurupremraj** ¹⁴, **Ami Tsuchida**^{1,15}, **Masaru Koido** ¹⁰, **Marziehsadat Ahmadi**¹⁶, **Chengran Yang** ^{2,3}, **Jigyasha Timsina** ^{2,3}, **Laura Ibanez**^{2,3}, **Koichi Matsuda**¹⁰, **Yutaka Suzuki** ¹⁰, **Yoshiya Oda**¹⁷, **Akinori Kanai**¹⁰, **Pouria Jandaghi**¹⁶, **Markus Munter**¹⁶, **Daniel Auld** ¹⁶, **Iana Astafeva**^{1,15}, **Raquel Puerta** ¹⁸, **Jerome I. Rotter** ¹⁹, **Bruce M. Psaty** ^{20,21,22}, **Joshua C. Bis** ²⁰, **WT Longstreth Jr**^{21,23}, **Thierry Couffinhal**²⁴, **Pablo García-González** ^{18,25}, **Vanesa Pytel**^{18,25}, **Marta Marquíe** ^{18,25}, **Amanda Cano**^{18,25}, **Mercè Boada** ^{18,25}, **Marc Joliot** ¹⁵, **Mark Lathrop**¹⁶, **Quentin Le Grand** ^{1,26}, **Lenore J. Launer** ²⁷, **Joanna M. Wardlaw** ^{28,29}, **Myriam Heiman**^{30,31}, **Agustin Ruiz** ^{14,18,25}, **Paul M. Matthews**^{32,33}, **Sudha Seshadri** ¹⁴, **Myriam Fornage** ^{34,35}, **Hieab Adams**^{36,37}, **Aniket Mishra**¹, **David-Alexandre Tréguët** ¹, **Yukinori Okada** ^{4,5,6,38,39}, **Manolis Kellis** ^{7,8}, **Philip L. De Jager** ¹¹, **Christophe Tzourio** ¹, **Yoichiro Kamatani** ^{10,42}, **Fumihiko Matsuda**^{9,42}, **Carlos Cruchaga** ^{2,3,42} ✉ & **Stéphanie Debette** ^{1,13,40,42} ✉

¹Bordeaux Population Health Research Center, Inserm U1219, Université de Bordeaux, Bordeaux, France. ²Department of Psychiatry, Washington University School of Medicine, St Louis, MO, USA. ³NeuroGenomics and Informatics Center, Washington University School of Medicine, St Louis, MO, USA.

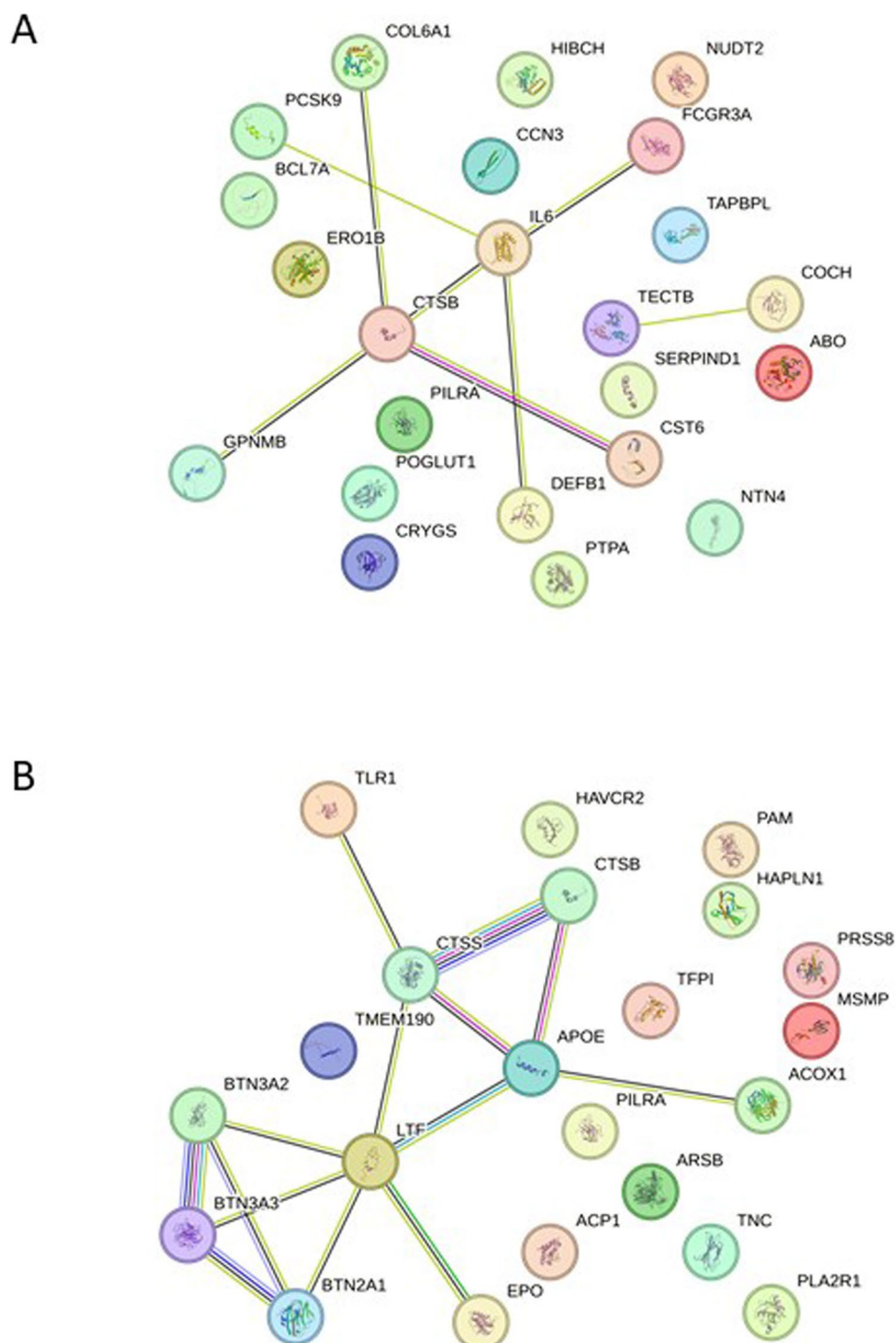
⁴Department of Genome Informatics, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan. ⁵Department of Statistical Genetics, Osaka University Graduate School of Medicine, Suita, Japan. ⁶Laboratory for Systems Genetics, RIKEN Center for Integrative Medical Sciences, Suita, Japan.

⁷MIT Computer Science and Artificial Intelligence Laboratory, Cambridge, MA, USA. ⁸Broad Institute of MIT and Harvard, Cambridge, MA, USA. ⁹Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan. ¹⁰Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan. ¹¹Department of Neurology, Center for Translational and Computational Neuroimmunology, Columbia University Irving Medical Center, New York, NY, USA. ¹²Department of Epidemiology, Erasmus MC University Medical Center Rotterdam, Rotterdam, the Netherlands. ¹³Department of Neurology, Institute for Neurodegenerative Diseases, CHU de Bordeaux, Bordeaux, France. ¹⁴Glenn Biggs Institute for Alzheimer's and Neurodegenerative Diseases, University of Texas Health Sciences Center, San Antonio, TX, USA. ¹⁵Institut des Maladies Neurodégénératives, IMN-UMR5293, Université de Bordeaux, CNRS, CEA, Bordeaux, France. ¹⁶Victor Phillip Dahdaleh Institute of Genomic Medicine, McGill University, Montreal, Quebec, Canada.

¹⁷Graduate School of Medicine, The University of Tokyo, Tokyo, Japan. ¹⁸Ace Alzheimer Center Barcelona, Universitat Internacional de Catalunya, Barcelona, Spain. ¹⁹The Institute for Translational Genomics and Population Sciences, Department of Pediatrics, The Lundquist Institute for Biomedical Innovation at Harbor-UCLA Medical Center, Torrance, CA, USA. ²⁰Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA, USA. ²¹Department of Epidemiology, University of Washington, Seattle, WA, USA. ²²Department of Health Systems and Population Health, University of Washington, Seattle, WA, USA. ²³Department of Neurology, University of Washington, Seattle, WA, USA. ²⁴Inserm U1034, Biologie Des Maladies Cardiovasculaires, Université de Bordeaux, Bordeaux, France. ²⁵CIBERNED, Network Center for Biomedical Research in Neurodegenerative Diseases, National Institute of Health Carlos III, Madrid, Spain. ²⁶Population Health Sciences, German Center for Neurodegenerative Diseases (DZNE), Bonn, Germany. ²⁷Intramural Research Program, National Institute on Aging, Baltimore, MD, USA. ²⁸Centre for Clinical Brain Sciences, University of Edinburgh, Chancellor's Building, Edinburgh, UK. ²⁹UK Dementia Research Institute Centre at the University of Edinburgh and Centre for Vascular Dementia Research, Edinburgh, UK. ³⁰Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA. ³¹The Picower Institute for Learning and Memory, Massachusetts Institute of Technology, Cambridge, MA, USA. ³²Department of Brain Sciences, Imperial College, London, UK. ³³UK Dementia Research Institute Centre at Imperial, Imperial College, London, UK. ³⁴Brown Foundation Institute of Molecular Medicine, McGovern Medical School, University of Texas Health Science Center at Houston, Houston, TX, USA. ³⁵Human Genetics Center, School of Public Health, University of Texas Health Science Center at Houston, Houston, TX, USA. ³⁶Department of Human Genetics, Radboud University Medical Center, Nijmegen, the Netherlands. ³⁷Latin American Brain Health (BrainLat), Universidad Adolfo Ibáñez, Santiago, Chile.

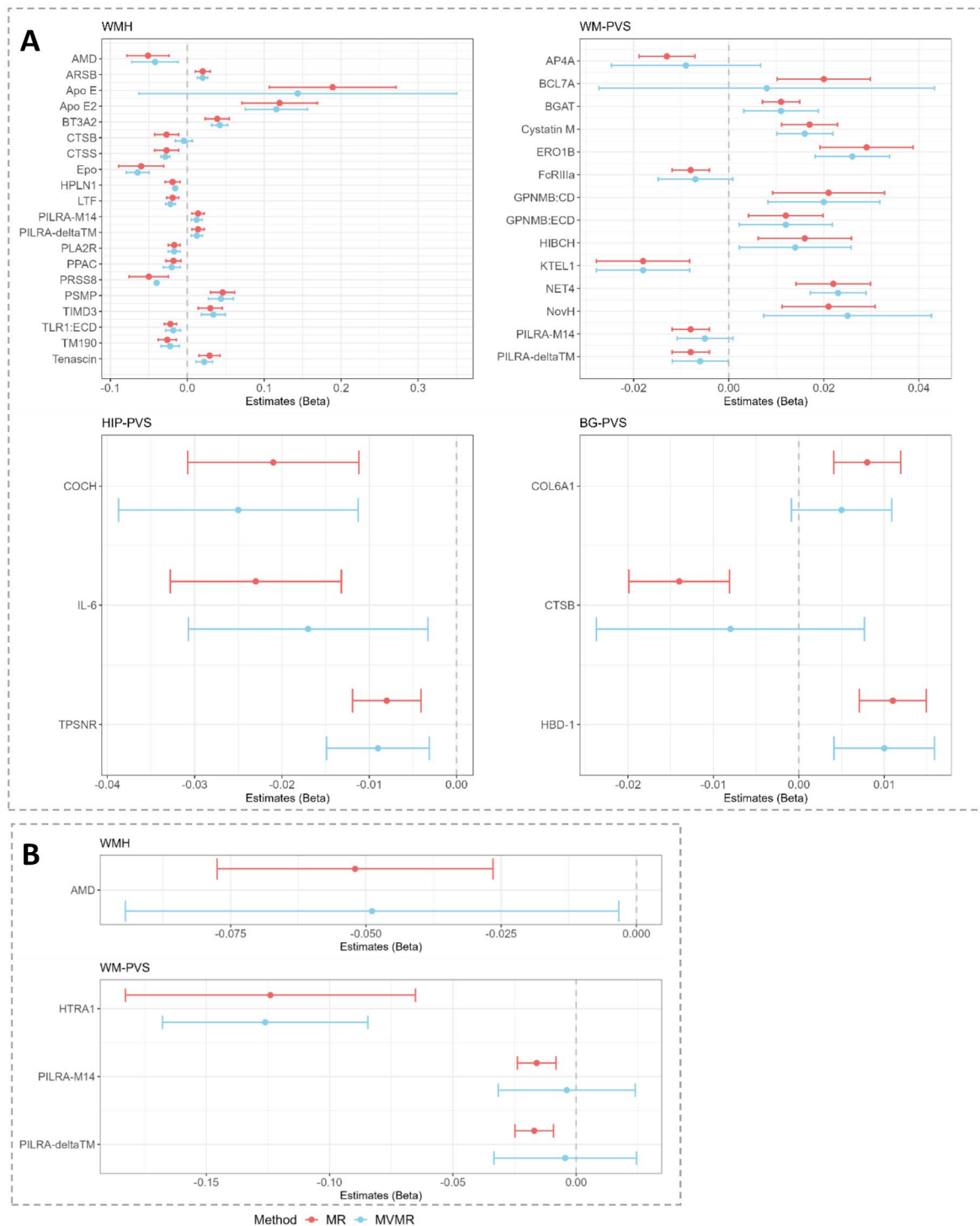
³⁸Laboratory of Statistical Immunology, Immunology Frontier Research Center (WPI-IFReC), Osaka University, Suita, Japan. ³⁹Premium Research Institute for Human Metaverse Medicine (WPI-PRIME), Osaka University, Suita, Japan. ⁴⁰Institut du Cerveau (ICM), Paris Brain Institute, INSERM U1127, UMR CNRS 7225 Paris, Sorbonne Université, Assistance Publique des Hôpitaux de Paris, Paris, France. ⁴¹These authors contributed equally: Ilana Caro, Daniel Western, Shinichi Namba. ⁴²These authors jointly supervised this work: Yoichiro Kamatani, Fumihiko Matsuda, Carlos Cruchaga, Stéphanie Debette.

✉ e-mail: cruchaga@wustl.edu; stephanie.debette@inserm.fr



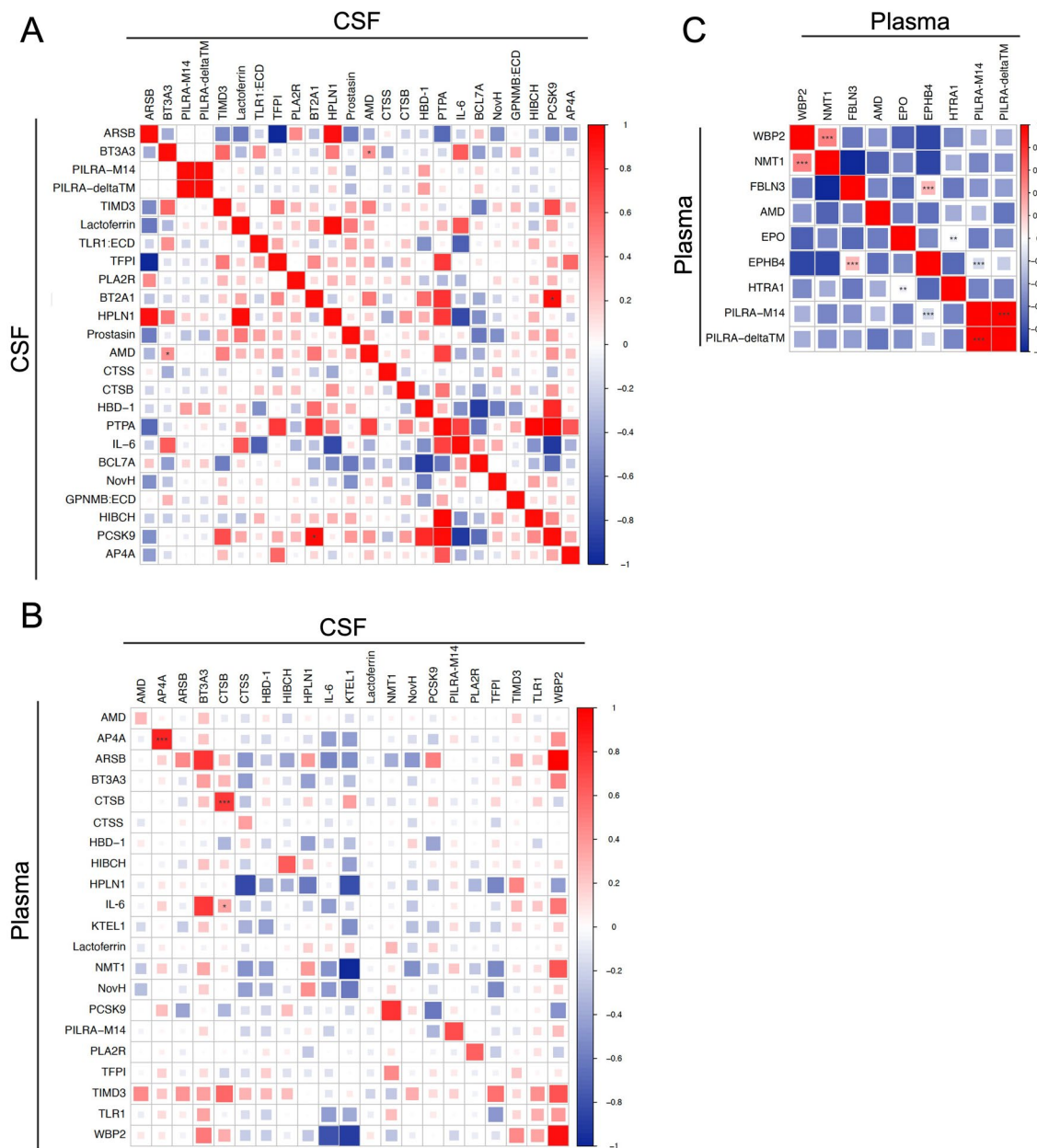
Extended Data Fig. 1 | Discovery protein-cSVD associations in CSF and plasma using cis-pQTL mendelian randomization. A. String plot of proteins associated with WMH. **B.** String plot of proteins associated with PVS (WM, BG and HIP). Network nodes represent proteins: colored nodes query proteins and first shell of interactors. Edges represent protein-protein associations.

Cyan and pink edges are known interactions, cyan: from curated databases, and pink: experimentally determined. Green and blue edges correspond to predicted interactions. Green: gene neighborhood, and blue: gene co-occurrence. Purple corresponds to protein homology, yellow to text mining and black to co-expression.



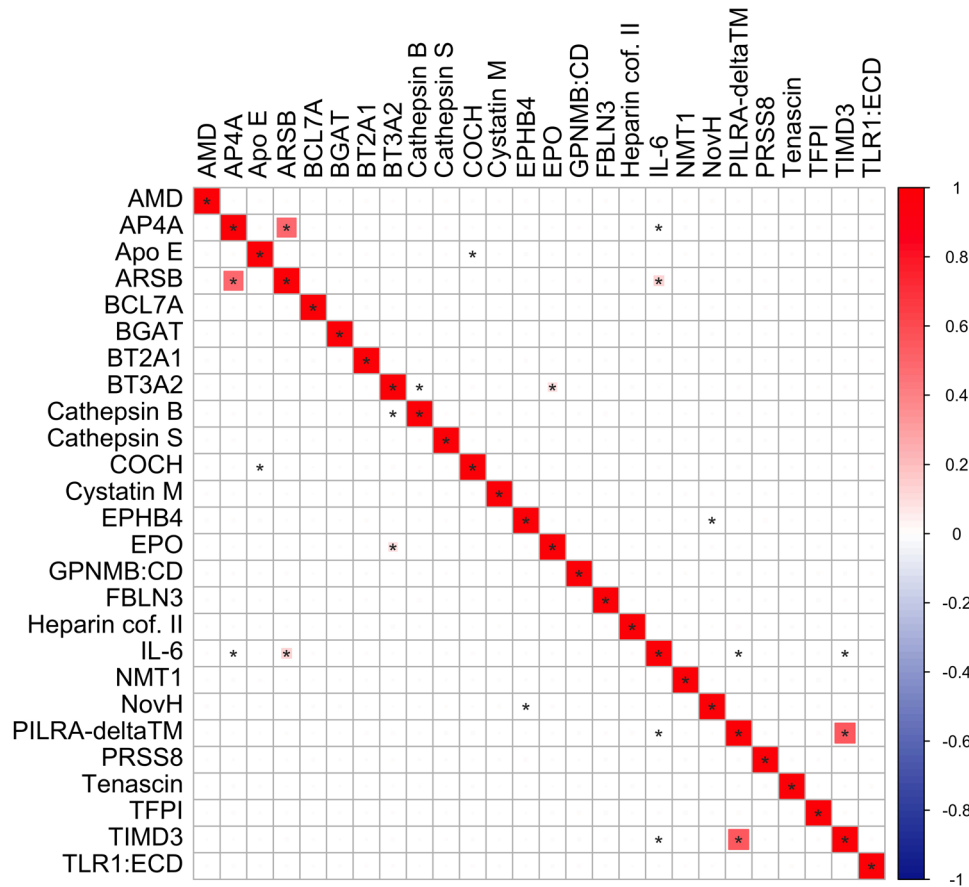
Extended Data Fig. 2 | Multivariable Mendelian randomization (MVMR) exploring the modifying effect of systolic blood pressure on the association of A. CSF proteins and B plasma proteins with MRI-cSVD. Only proteins with >1 pQTL available for analysis are represented. Dots correspond to effect estimates

(beta) and error bars to 95% confidence intervals. Forest plots correspond respectively to protein associations with WMH, WM-PVS, HIP-PVS and BG-PVS. Red lines correspond to primary MR results; blue lines correspond to multivariable MR (MVMR).

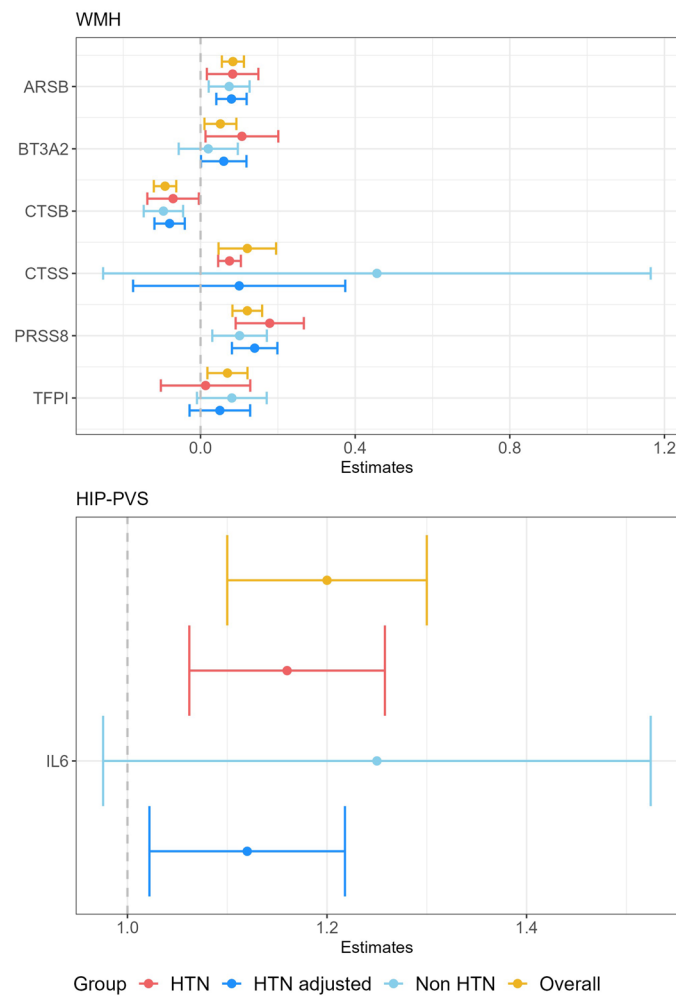


Extended Data Fig. 3 | Genetic correlation between proteins associated with MRI-cSVD using protein quantitative trait loci. A. Genetic correlation across pQTLs for 24 CSF protein levels associated with MRI-cSVD. **B.** Genetic correlation across pQTLs for 21 CSF or plasma proteins significantly associated with MRI-cSVD and available in both plasma and CSF (x axis: pQTL for CSF protein levels; y axis: pQTL for plasma protein levels). **C.** Genetic correlation across pQTLs for

9 plasma protein levels associated with MRI-cSVD. Genetic correlations were estimated using LD score regression. P-values are based on two-sided Wald tests of the null hypothesis, without adjustment for multiple comparisons. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Only proteins that converged for genetic correlation analyses are displayed.

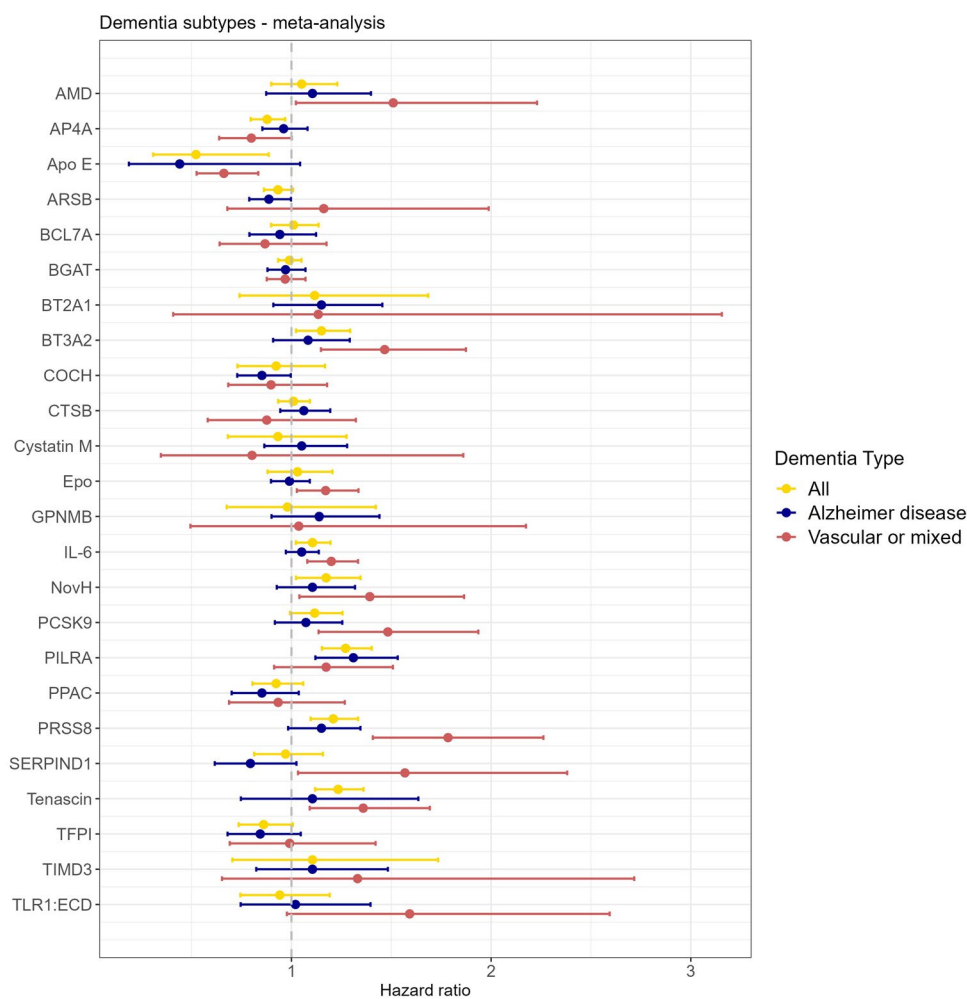


Extended Data Fig. 4 | Correlation of protein levels measured in the UKB across the 26 cSVD-associated proteins. Correlations were estimated using Spearman correlation (two-sided). Unadjusted p-values are displayed. * $p < 7.7 \times 10^{-3}$ (Bonferroni corrected threshold).

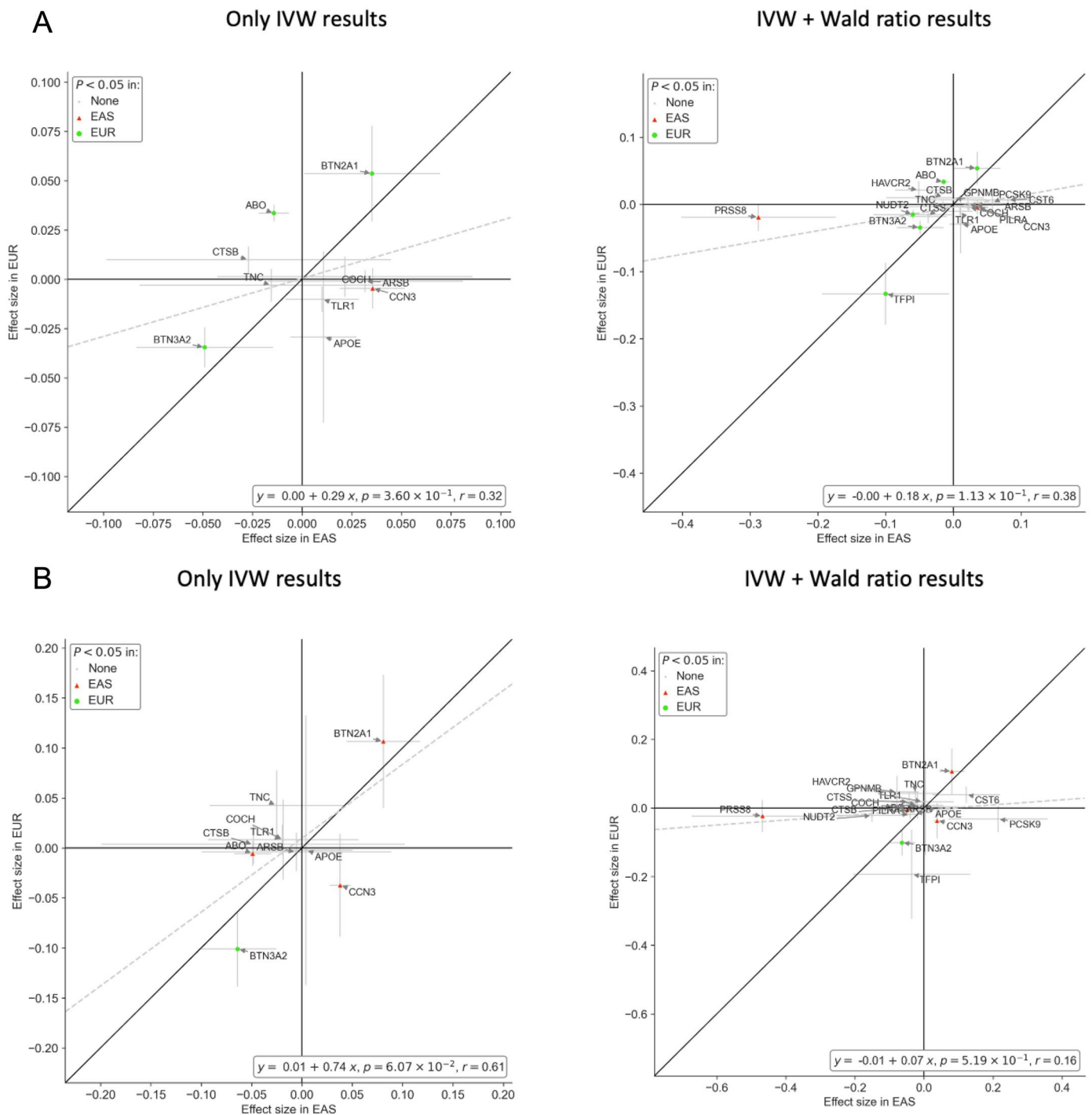


Extended Data Fig. 5 | Association of plasma protein levels with MRI-cSVD stratified on hypertension status. Forest plots correspond to protein associations with WMH and HIP-PVS. Results corresponds to meta-analyses of association results in 3C-Dijon and UK Biobank (N = 6,581; 2,088 hypertensive/3,406 non-hypertensive) for WMH; and 3C-Dijon only for HIP-PVS

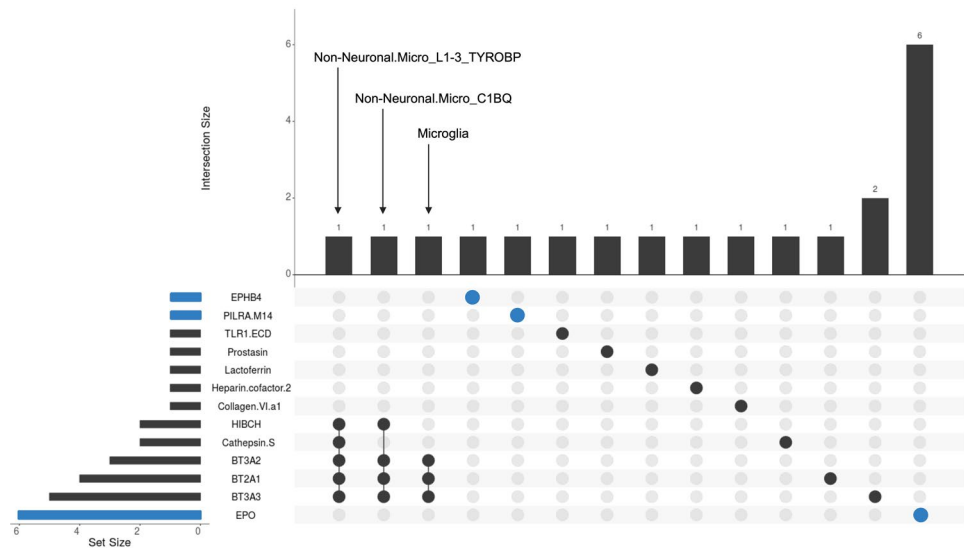
(N = 1,087; 235 hypertensive/852 non-hypertensive), stratified on hypertension status: hypertensive (HTN), non-hypertensive (Non-HTN), all combined (Overall) and adjusted on systolic blood pressure (HTN adjusted). Dots represent effect estimates (beta for WMH and odds ratio for HIP-PVS) and error bars correspond to 95% confidence intervals.



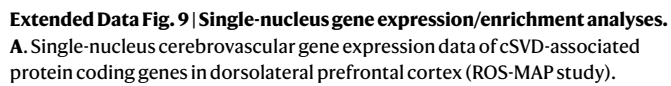
Extended Data Fig. 6 | Protein-cSVD associations with dementia subtypes (vascular and Alzheimer's disease): meta-analysis of 3C-Dijon and UK Biobank. Results of cause-specific Cox models (Methods). N-vascular dementia=385; N-Alzheimer=1,107). Dots represent hazard ratios and errors bars corresponds to 95% confidence intervals.

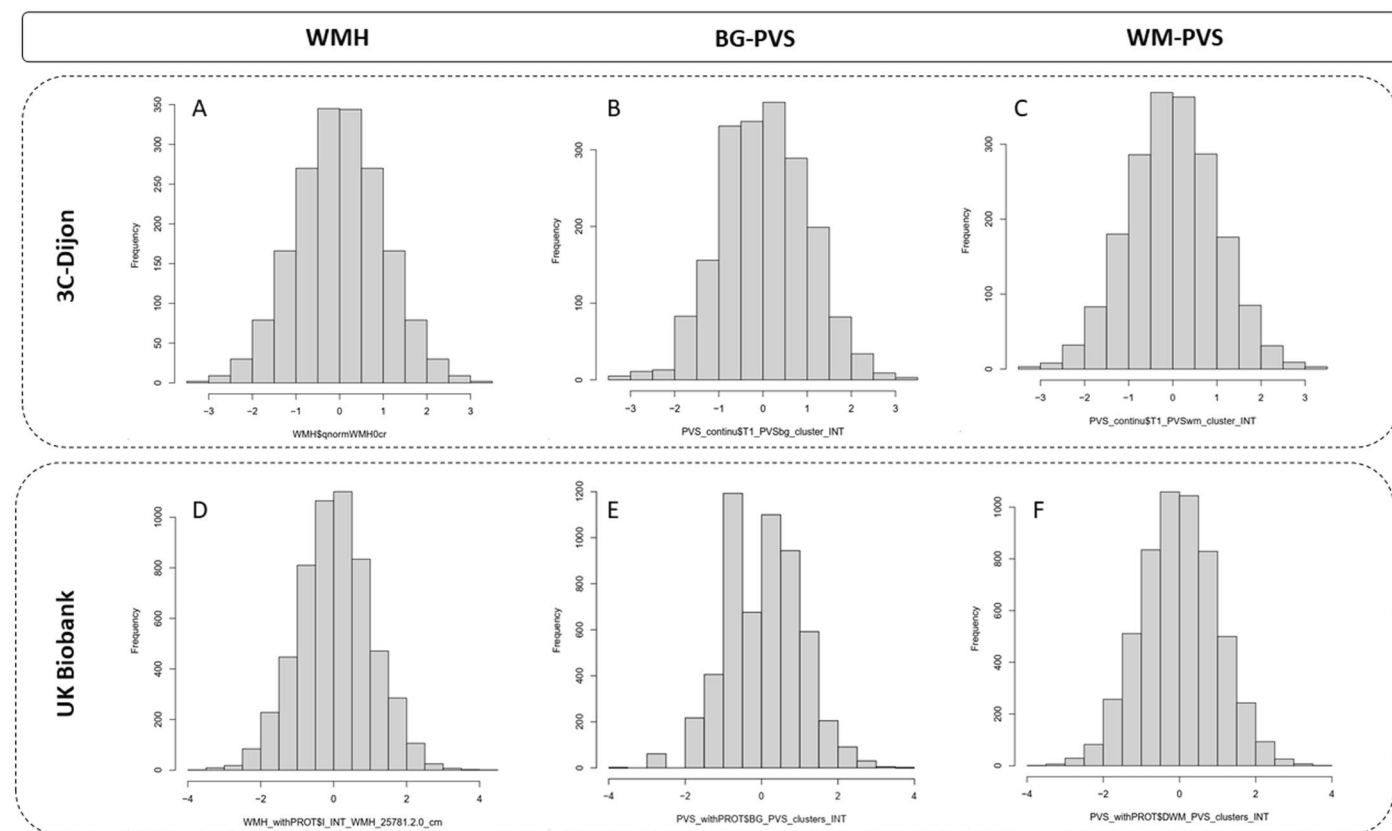


Extended Data Fig. 7 | Comparison of effect size estimates for MR associations of cSVD-associated proteins with ischemic stroke and small vessel stroke between Europeans (EUR) and East-Asians (EAS). A. Ischemic Stroke. B. Small vessel Stroke.



Extended Data Fig. 8 | Cell-type enrichment in single cell RNA-seq databases using STEAP. Upset plot displays the number of significant enrichment results by protein (pQTL) horizontally and by cell-type vertically. CSF pQTLs are in black and plasma pQTLs are in blue. Details are displayed in Supplementary Table 25. Human and mouse single-cell databases are used in this analysis (Methods).





Extended Data Fig. 10 | Distribution of MRI markers of cSVD in 3C-Dijon and UK-Biobank. Histogram of white matter hyperintensities (WMH) and perivascular spaces (PVS) distribution after normal inverse transformation in 3C-Dijon (**A-C**) and the UK Biobank (**D-F**).

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- | | |
|-------------------------------------|--|
| n/a | Confirmed |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> The exact sample size (<i>n</i>) for each experimental group/condition, given as a discrete number and unit of measurement |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided
<i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A description of all covariates tested |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> For null hypothesis testing, the test statistic (e.g. <i>F</i> , <i>t</i> , <i>r</i>) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted
<i>Give P values as exact values whenever suitable.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Estimates of effect sizes (e.g. Cohen's <i>d</i> , Pearson's <i>r</i>), indicating how they were calculated |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection	No software was used for data collection
Data analysis	We used publicly available tools from TwoSampleMR (v0.5.686, https://mrcieu.github.io/TwoSampleMR/), FUMA (v1.5.4, https://fuma.ctglab.nl/), coloc R package (v5.2.3, https://chr1swallace.github.io/coloc/), STRING (v12.0, https://string-db.org/), FUSION (v2.7.13 https://github.com/gusevlab/fusion_twas), PLINK (v2.0, https://www.cog-genomics.org/plink/2.0/), OlinkAnalyze R package (v4.3.1, https://github.com/Olink-Proteomics/OlinkRPackage), MungeSumstats R package (v1.16.0, https://github.com/neurogenomics/MungeSumstats), ggplot2 R package (v3.5.2, https://github.com/tidyverse/ggplot2), forestplot R package (v3.1.7, https://cran.r-project.org/web/packages/forestplot), Betsholtz database (http://betsholtzlab.org/VascularSingleCells/database.html), metafor R package (v4.8-0, https://wwiechth.github.io/metafor/), SHIVA-PVS algorithm (https://github.com/pboutinaud/SHIVA_PVS , T1.PVS/v1), STEAP (https://github.com/erwinerdm/STEAP). Figures were created using ggplot2 R package (https://ggplot2.tidyverse.org/) and BioRender (https://app.biorender.com).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

We used publicly available data for analyses described in this manuscript, including data from GWAS catalog (<https://www.ebi.ac.uk/gwas/>, study code: GCST90244151, GCST011947, GCST007320, GCST90104539, GCST90162546), the DECODE project (<https://www.decode.com/summarydata/>), ChEMBL (<https://www.ebi.ac.uk/chembl/>), pharmGKB (<https://www.pharmgkb.org/>), DrugBank (<https://go.drugbank.com/>), TTD (<https://db.idrblab.net/ttd/>), CSF pQTL summary statistics available at NIAGADS (full summary statistics available to approved investigators through accession #NG00130), NeuroGenomics and Informatics Center website (<https://neurogenomics.wustl.edu/open-science/raw-data/>) and ONTIME browser (<https://ontime.wustl.edu/>). Plasma proteomic data for the Knight ADRC participants are available at <https://knightadrc.wustl.edu/professionals-clinicians/request-center-resources/>. Requests for clinical or proteomic data from individual investigators will be reviewed to ensure compliance with patient confidentiality. For details on accessing available data and study protocols, see <https://knightadrc.wustl.edu>.

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender

Our primary analyses rely on summary statistics used for mendelian randomization: all of the underlying GWAS included genetically determined sex as an adjustment variable. For individual-level data association analyses between protein levels and MRI or clinical outcomes, self-reported sex was used as an adjustment variable. Due to limited power (and unavailability of sex-specific GWAS summary statistics for our discovery analysis), we did not run sex-specific analyses in the individual-level follow-up.

Reporting on race, ethnicity, or other socially relevant groupings

Analyses were conducted using GWAS summary statistics of European and East-Asian ancestry participants separately; ancestries were genetically inferred using principal components analyses. Individual-level analyses were also conducted separately in European and East-Asian ancestry groups, equally defined using genetically inferred principal components, and additionally adjusted for principal components of population stratification.

Population characteristics

For Mendelian Randomization, protein quantitative trait loci were used as exposure and obtained from up to 35,559 participants of Europeans ancestry; GWAS summary statistics of WMH (N=48,454) and PVS (N up to 40,025) were used as outcome. Further exploration in young adults (22.1±2.3 years, 72.2% women, N=1,748) were used for GWAS of WMH and PVS and used as outcome in the MR. Further explorations using individual-level datasets leveraged the 3C-Dijon cohort (72.5 ±4.1 years, 60.5% women, N=1,043), the UK Biobank (63.5±7.9 years, 53.5% women, N=5,494) and the Nagahama (68.3±5.3 years, 64.4% women, N=785). For individual-level analyses, ancestry-specific linear or logistic regression have been performed, adjusting for the delay between age at blood draw and age at the time of MRI, sex, intracranial volume (for WMH), relevant principal components of population stratification, and batch effect of proteomic measurements. Further analysis were stratified based on hypertension status. Detailed population characteristics by study and phenotype definitions can be found in the Methods and Supplementary Methods sections and in Supplementary Table 12.

Recruitment

All participating studies had population-based recruitment strategies. Details can be found in the Methods section and Supplementary Methods.

Ethics oversight

Study protocols were approved for all studies by the appropriate boards at their respective institutions, these have been reported in the Supplementary Methods. Study protocols were approved for all studies by the appropriate boards at their respective institutions:

- 3C-Dijon: Ethics committee of the University Hospital of Kremlin-Bicêtre
- iSHARE: Commission Nationale Informatique et Libertés (CNIL); CCTIRS (Comité Consultatif sur le Traitement de l'Information en matière de Recherche dans le domaine de la Santé); CPP (Comité de Protection des Personnes); l'Agence Nationale de Sécurité du Médicament et Hors Produit de Santé (ANSM)
- Nagahama : Ethics committee of Kyoto University Graduate School of Medicine and the Nagahama Municipal Review Board
- UK Biobank: National Research Ethics Service Committee North West-Haydock (reference 11/NW/0382)
- ACE: The study protocols are approved by the Clinical Research Ethics Commission of the Hospital Clinic (Barcelona, Spain)

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- ☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	We conducted a multi-pronged approach to assess the association of protein from CSF and plasma with MRI markers of cerebral small vessel disease: WMH and PVS. First, we conducted a Mendelian Randomization using the largest and unique CSF pQTL dataset (N=3,107; measuring 7,028 proteins) and the largest plasma pQTL dataset (N=35,559, measuring 4,907 proteins) as exposure; and the largest GWAS meta-analyses of WMH (N=48,454) and extensive PVS in its 3 sublocations (white matter, basal ganglia, hippocampus) (N up to 38,525) as outcome. Those datasets were of Europeans ancestry only. As stated in the Methods section “No statistical methods were used to pre-determine sample sizes but our sample sizes are similar or larger to those reported in previous publications ^{13,81,83,84.} ”
Data exclusions	Participants were excluded from analysis if they had co-morbidities modifying the measurement of the phenotype (prevalent stroke, prevalent dementia, multiple sclerosis or radiologically isolated syndrome), missing covariates or withdrawn consent. Proteins were excluded if they failed quality control filters described in the Supplementary Methods (page 13).
Replication	We conducted multiple follow-up analyses, across ancestries and across the lifespan, in the iSHARE cohort (N=1,748), the 3C-Dijon cohort (N=1,087), the UK Biobank (N=5,794) and the Nagahama cohort (N=785). These are described in detail in the Method section (pages 40-44) and Figure 1.
Randomization	We did not allocate participants into experimental groups. However we conducted Mendelian randomization analyses based on genetic associations in the study population (described in the Methods). As stated in Methods section: “This study is based on summary statistics and observational cohort data. Randomization and blocking were not applicable as participants were not assigned to experimental groups.”
Blinding	Not relevant for the present study design (strictly speaking MRI measurements of WMH and PVS burden were measured blinded to the protein levels and vice versa). MR was conducted using summary statistics of available datasets, not individual level data.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern
<input checked="" type="checkbox"/>	<input type="checkbox"/> Plants

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input type="checkbox"/>	<input checked="" type="checkbox"/> MRI-based neuroimaging

Plants

Seed stocks	NA.
Novel plant genotypes	NA.
Authentication	NA.

Magnetic resonance imaging

Experimental design

Design type	Population-based brain imaging study.
-------------	---------------------------------------

Design specifications	The recruitment strategy for each cohort studies (3C-Dijon, iSHARE, Nagahama and UK Biobank) is described in the Methods section.
Behavioral performance measures	There are no behavioral performance measures in this study.

Acquisition

Imaging type(s)	Brain MRI with T1-, T2-weighted, FLAIR and proton density sequences images. Study-specific information is described in the Method and Supplementary Methods section.
Field strength	1.5 and 3.0 Tesla.
Sequence & imaging parameters	Study-specific sequence and imaging parameters informations is described in Methods section (page 40-44).
Area of acquisition	Brain.
Diffusion MRI	<input type="checkbox"/> Used <input checked="" type="checkbox"/> Not used

Preprocessing

Preprocessing software	Study-specific information is described in the Methods section (page 17-23)
Normalization	Study-specific information is described in the Methods section (page 17-23)
Normalization template	Study-specific information is described in the Methods section (page 17-23)
Noise and artifact removal	Study-specific information is described in the Methods section (page 17-23)
Volume censoring	Study-specific information is described in the Methods section (page 17-23)

Statistical modeling & inference

Model type and settings	We conducted multivariable linear and logistic regression of individual proteins with WMH volume and PVS burden adjusted for the delay between age at blood draw and age at the time of MRI, sex, batch effect, total intracranial volume (or mask volume for WMH in 3C-Dijon). WMH volume and PVS burden in basal ganglia and white matter were inverse normal transformed and PVS in hippocampus values were dichotomized, comparing participants in the top quartile of PVS burden distribution to the rest. Cause-specific cox models and Fine-Gray models were used to assess the association of protein levels with incident stroke and dementia in the UK Biobank and 3C-Dijon using age as a time scale, and adjusting for batch, sex and self-reported ancestry, and additionally educational attainment for associations with incident dementia.
Effect(s) tested	Association of protein levels in whole blood with white matter hyperintensities and perivascular spaces burden in white matter, basal ganglia and hippocampus.
Specify type of analysis:	<input checked="" type="checkbox"/> Whole brain <input type="checkbox"/> ROI-based <input type="checkbox"/> Both
Statistic type for inference	Non relevant for this study.
(See Eklund et al. 2016)	
Correction	False discovery rate Benjamini Hochberg 5%.

Models & analysis

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Functional and/or effective connectivity
<input checked="" type="checkbox"/>	<input type="checkbox"/> Graph analysis
<input checked="" type="checkbox"/>	<input type="checkbox"/> Multivariate modeling or predictive analysis