Journal of Behavioral and Brain Science, 2026, 16(9), 237-257
https://www.scirp.org/journal/jbbs
ISSN Online: 2160-5874
ISSN Print: 2160-5866

Proteome-Wide Mendelian Randomization Reveals Divergent Plasma Protein Landscapes Underlying White Matter Hyperintensities and Lacunar Stroke

Hanying Gu[1], Xiuxia Shi[2], Jiangtao Zhang[3*]

[1]Emergency Department, Chengde Medical University, Chengde, China
[2]Department of Critical Care Medicine, Chengde Central Hospital, Chengde, China
[3]Department of Neurology, Chengde Central Hospital, Chengde, China
Email: *zhangjiangtao@vip.163.com

How to cite this paper: Gu, H.Y., Shi, X.X., and Zhang, J.T. (2026) Proteome-Wide Mendelian Randomization Reveals Divergent Plasma Protein Landscapes Underlying White Matter Hyperintensities and Lacunar Stroke. *Journal of Behavioral and Brain Science*, **16**, 237-257.
https://doi.org/10.4236/jbbs.2026.169010

Received: July 28, 2026
Accepted: September 12, 2026
Published: September 15, 2026

Copyright © 2026 by author(s) and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/

Open Access

Abstract

Background: Cerebral small vessel disease (CSVD) is a leading cause of stroke and dementia, with white matter hyperintensities (WMH) and lacunar stroke as its two principal neuroimaging subtypes. Although these subtypes frequently co-occur, whether their underlying molecular mechanisms are independent has not been systematically examined at the genetic level. This study aimed to apply proteome-wide Mendelian randomization (MR) to test whether the causal protein profiles of WMH and lacunar stroke are distinct, and to identify subtype-specific candidate targets. **Methods:** We employed a two-sample MR design. Genetic instruments for 1954 plasma proteins were derived from a large-scale proteogenomic dataset (N = 54,219). Outcome data came from the largest available GWAS for lacunar stroke (7338 cases, 225,258 controls) and WMH burden (N = 51,065). The Wald ratio method was used for primary MR analysis, with colocalization, SMR, and pleiotropy checks performed to assess robustness. **Results:** After FDR correction, seven proteins showed causal associations with lacunar stroke (TMPRSS5, WARS, KIAA0319, MMP12, CD40, FURIN, GRK5), and five with WMH burden (EFNA1, DNAJC21, MPI, BTN3A2, BTN2A1). There was no overlap between the two FDR-significant sets. **Conclusions:** These findings show no overlapping FDR-significant plasma protein associations between WMH and lacunar stroke, suggesting largely independent molecular mechanisms. Subtype stratification should be considered in future CSVD trial design. Several identified targets—including FURIN, WARS, and BTN3A2—have existing investigational agents, nominating them for further pharmacological evaluation.

Keywords

Mendelian Randomization, Cerebral Small Vessel Disease, White Matter Hyperintensities, Lacunar Stroke

1. Introduction

Cerebral small vessel disease (CSVD) refers to a group of pathological processes affecting the small arteries, arterioles, capillaries, and small veins of the brain [1]. As one of the most prevalent age-related cerebrovascular disorders, CSVD underlies approximately 25% of ischemic strokes and nearly all intracerebral hemorrhages, and is the leading cause of vascular cognitive impairment and dementia, imposing a substantial burden on global public health systems [2]. Genomic studies have demonstrated that CSVD is highly heritable, with multiple genetic loci established as associated with WMH volume and lacunar stroke risk [3]-[5]. In clinical practice, CSVD is primarily diagnosed through neuroimaging features, with WMH and lacunar stroke constituting its two most important imaging phenotypes: the former manifesting as diffuse signal abnormalities in the cerebral white matter, and the latter as small subcortical infarcts [6] [7]. Notably, CSVD is associated not only with stroke and cognitive impairment but also with sleep disturbances, mood disorders, and disruption of daily activity rhythms, underscoring the breadth of its clinical impact [8]-[10].

Although WMH and lacunar stroke frequently co-occur and are collectively classified as CSVD, accumulating evidence suggests that their pathophysiological mechanisms may diverge. WMH is primarily associated with endothelial dysfunction, blood-brain barrier disruption, and chronic ischemia [11] [12], whereas lacunar stroke is more directly linked to lipohyalinosis and microatherosclerosis of small penetrating arteries [2]. WMH progression is closely associated with cognitive decline, increased recurrent stroke risk, and worsened functional outcomes [13]-[16], while lacunar stroke is linked to executive dysfunction and synergistic motor-cognitive impairment [16] [17]. Additionally, metabolic factors such as visceral fat accumulation, elevated homocysteine, and reduced cerebral blood flow have been reported in association with WMH [18]-[20], whereas intracranial atherosclerosis plays a prominent role in lacunar stroke [21]-[23]. Nevertheless, this hypothesized mechanistic divergence has lacked systematic molecular-genetic evidence, and a direct proteome-wide comparison between the two subtypes has not previously been performed.

Mendelian randomization (MR) is an analytical method that uses genetic variants as instrumental variables to infer causal relationships between exposures and outcomes [24]. Because alleles are randomly allocated during gamete formation, MR analysis effectively mimics a randomized controlled trial, largely circumventing the confounding and reverse causation that plague conventional observational studies. With the emergence of large-scale protein quantitative trait locus (pQTL)

studies, proteomic MR has become a powerful tool for identifying disease-relevant causal proteins and candidate drug targets [25]. The use of cis-pQTLs as instruments is conceptually equivalent to a genetic perturbation experiment targeting a single protein, and has been validated as a recommended strategy for systematic drug target discovery [24] [25]. This approach has been successfully applied across a range of conditions including heart failure, obesity, cardiac structure, and brain structural phenotypes, demonstrating broad translational value [26]-[30]. Prior work has also identified multiple stroke-related protein targets through proteomic MR, providing an important methodological precedent for the present study [25].

The recently released UK Biobank Pharma Proteomics Project (UKB-PPP) dataset provides the largest plasma protein pQTL resource to date, covering 1954 proteins in 54,219 individuals. Complementary proteogenomic studies across diverse populations have further enriched our understanding of genetic regulation of the plasma proteome, creating an unprecedented opportunity for proteome-wide MR screening [31] [32].

Against this background, the present study aimed to leverage this resource together with the largest available GWAS data for lacunar stroke and WMH to systematically address a central scientific question: are the plasma protein molecular networks driving WMH and lacunar stroke mutually independent? Our primary objectives were: 1) to identify plasma proteins with causal relationships with lacunar stroke and WMH; 2) to directly compare the protein target profiles of the two subtypes and test the independence of their molecular mechanisms; and 3) to identify subtype-specific candidate drug targets with potential translational value.

2. Methods

2.1. Study Design

This study followed a two-sample Mendelian randomization design, with all analyses based on publicly available summary-level statistics. Reporting adhered to the STROBE-MR statement for MR studies.

2.2. Data Sources

Genetic instruments for plasma proteins were derived from the UKB-PPP study reported by Sun *et al.*, which measured plasma protein levels using the Olink Explore 3072 platform in up to 54,219 UK Biobank participants and conducted GWAS for 1954 proteins. Effective sample sizes varied by protein due to platform-specific batch effects and protein-level QC filtering; per-protein effective N ranged from approximately 34,000 to 54,000. F-statistics and r^2 values reported in **Table S1** were calculated using the effective N for each respective instrument, as reflected in the UKB-PPP summary statistics [31]. For each protein, we selected the cis-pQTL as the instrumental variable, defined as the SNP with the smallest p-value among those located within 1 Mb upstream or downstream of the transcription start and stop sites of the protein-coding gene and reaching genome-wide significance ($p < 5 \times 10^{-8}$). Only a single lead SNP was used as the instrument per

protein. The cis-pQTL design is conceptually equivalent to a natural genetic perturbation of a single protein, substantially reducing the risk of horizontal pleiotropy, and represents the recommended approach for systematic drug target discovery [24] [25]. Two quality control procedures were applied to ensure instrument validity: 1) calculation of the F-statistic ($F = \beta^2/SE^2$), with exclusion of weak instruments ($F < 10$); and 2) application of the Steiger filtering test to exclude instruments potentially reflecting reverse causation.

GWAS summary statistics for lacunar stroke were obtained from the large-scale meta-analysis reported by Traylor *et al.* (GWAS Catalog ID: GCST006906), comprising 7338 lacunar stroke cases and 225,258 controls of European ancestry [33]. Multiple large-scale GWAS have systematically characterized the genetic architecture of CSVD-related phenotypes, informing the selection of outcome data used in the present study [3]-[5] [34]. GWAS summary statistics for WMH were obtained from Patel *et al.*, who performed GWAS in 51,065 individuals of European ancestry for WMH burden and related cortical structural damage [35]. The WMH phenotype was standardized, and effect estimates (β) represent the change in WMH-related brain structural damage score per standard deviation increase in protein level.

A potential limitation of the WMH analysis is sample overlap between the exposure and outcome GWAS: both the UKB-PPP (Sun *et al.*) and the WMH GWAS (Patel *et al.*) include participants from the UK Biobank. For the lacunar stroke outcome (Traylor *et al.*, a multi-cohort meta-analysis not limited to UKB), this issue does not apply. For WMH, the exact degree of overlap is not directly inferable from published summary statistics. However, two-sample MR estimates are generally robust to moderate sample overlap (<50%), and the large effective sample sizes in both studies mitigate the risk of weak instrument bias. Nonetheless, we acknowledge this limitation and interpret the WMH MR estimates with appropriate caution; future studies using independent non-overlapping exposure-outcome datasets are warranted to confirm these findings.

All analyses were performed using GRCh37/hg19 coordinates. For each cis-pQTL, the lead SNP was extracted from the UKB-PPP summary statistics. Allele harmonization between exposure and outcome GWAS was performed using the TwoSampleMR package, ensuring that effect alleles were aligned to the same reference strand. Palindromic variants with intermediate allele frequencies (MAF > 0.42) were excluded to avoid strand ambiguity. For LD-based analyses (clumping and colocalization), the 1000 Genomes European reference panel was used. Colocalization analysis was performed within a ±500 kb window around each lead cis-pQTL.

2.3. Mendelian Randomization Analysis

As each protein was represented by a single cis-pQTL, the Wald ratio method was used to estimate causal effects: $\beta_{MR} = \beta_{GY}/\beta_{GX}$. For the binary outcome of lacunar stroke, results are reported as odds ratios (OR) with 95% confidence intervals (CI); for the continuous outcome of WMH, results are reported as β coefficients with

95% CI. Throughout the manuscript, all WMH effects are presented as β coefficients for the standardized continuous phenotype; no WMH results are reported as odds ratios. Multiple testing correction was applied using the Benjamini-Hochberg method, with FDR < 0.05 defined as statistically significant.

2.4. Colocalization Analysis

Bayesian colocalization analysis was performed for all significant MR findings using the coloc.abf() function from the R package coloc [36], with prior probabilities set at p1 = p2 = 1×10^{-4} and p12 = 1×10^{-5}. A posterior probability PP.H4 ≥ 0.7 was taken as evidence supporting a shared causal variant between the protein and the disease.

2.5. SMR Analysis

SMR software (v1.3.1) was used to perform summary-data-based Mendelian randomization with reference to the large-scale blood eQTL data published by Westra *et al.* [37]. The SMR test integrates eQTL data to examine whether gene expression levels mediate the effect of genetic variants on disease, providing independent transcriptomic-level corroboration of protein-level MR findings. As the corresponding linkage disequilibrium (LD) reference panel was unavailable, the heidioff parameter was specified to bypass the HEIDI test. Instead, we used Bayesian colocalization (PP.H4) to assess whether the MR signals were driven by shared causal variants versus LD. Colocalization and HEIDI address the same conceptual question but are not statistically equivalent; therefore, SMR findings without HEIDI testing and without multiple-testing correction should be interpreted as exploratory transcriptomic corroboration, not as confirmatory evidence [36]. A p_SMR < 0.05 was considered supportive of an SMR association.

2.6. Pleiotropy Assessment

The GWAS Catalog REST API was used to query all 12 finalized instrumental variable SNPs for associations with other known phenotypes, assessing the risk of horizontal pleiotropy. Prior studies have shown that diverse metabolic and immune factors, including homocysteine and inflammatory cytokines, can influence CSVD through pleiotropic pathways; our pleiotropy checks aimed to rule out such confounding [19] [38].

2.7. Software

All statistical analyses were conducted in R 4.5.3, using the TwoSampleMR and coloc packages.

3. Results

3.1. Overall Screening Results

Of the 1954 plasma proteins, 1890 had at least one qualifying cis-pQTL and were included in the lacunar stroke MR analysis (**Figure 1**). For WMH, 1648 proteins

were evaluable after harmonization with the outcome GWAS summary statistics (some instruments were unavailable in the WMH dataset); FDR correction was applied separately for each outcome using the respective number of evaluated proteins (m = 1890 for lacunar stroke; m = 1648 for WMH). F-statistics for all instrumental variables substantially exceeded the threshold of 10 (range: 202.2 - 10247.5), and all passed Steiger filtering, confirming adequate instrument strength and appropriate causal direction (Table S1). After FDR correction, seven proteins were significantly associated with lacunar stroke risk and five proteins with WMH burden (Figure 2(A) and Figure 2(C)). There was no overlap between the two FDR-significant target sets. To formally assess whether this divergence extended beyond the significant hits, we performed a post-hoc concordance analysis across all proteins evaluable in both outcomes: among proteins with nominal significance

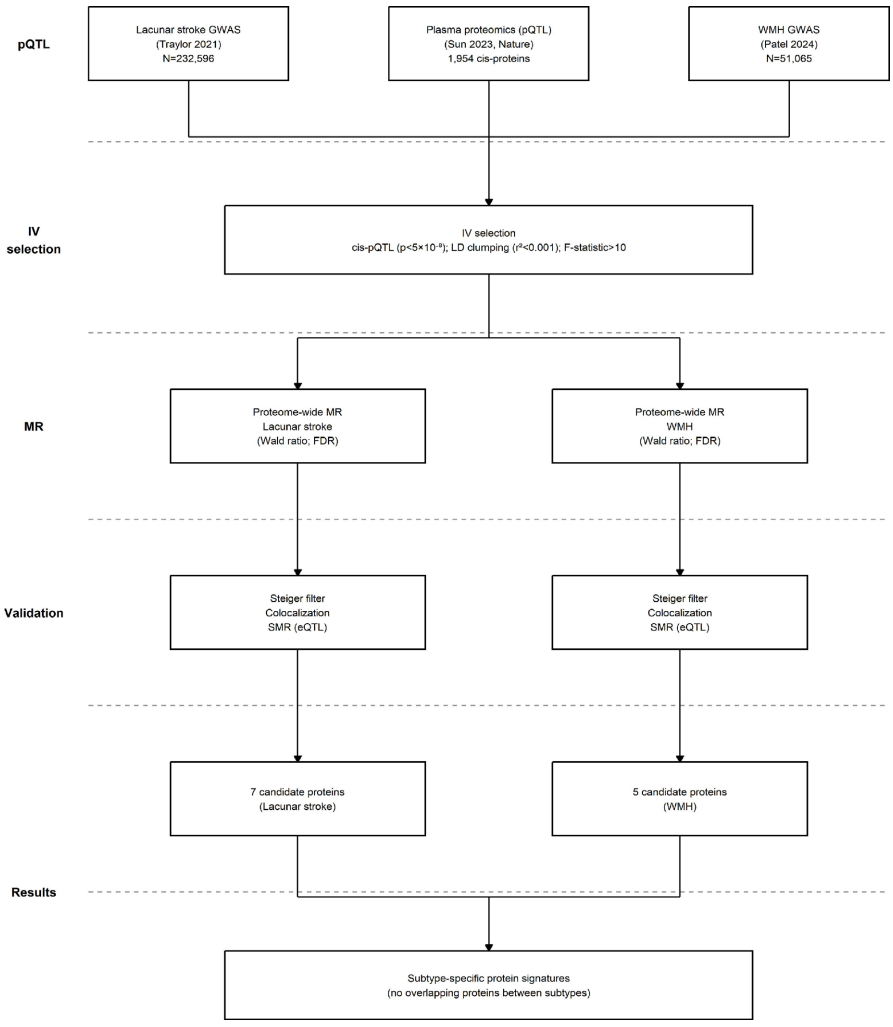

Figure 1. Study design. Two-sample MR framework: 1954 plasma proteins (UKB-PPP cis-pQTLs) tested against lacunar stroke and WMH; proteins passing FDR < 0.05 were validated by colocalization and SMR. Abbreviations: MR, Mendelian randomization; UKB-PPP, UK Biobank Pharma Proteomics Project; cis-pQTL, cis-acting protein quantitative trait locus; WMH, white matter hyperintensities; FDR, false discovery rate; SMR, summary-data-based Mendelian randomization.

Table 1. Causal protein associations with lacunar stroke and WMH.

Protein	Outcome	Lead SNP	F-statistic	Effect (95% CI or SE)	p (MR)	FDR	Steiger	PP.H4	p (SMR)	Evidence
MMP12	Lacunar stroke	rs17368814	5027.4	0.921 (0.887 - 0.956)	1.83e−05	0.014	PASS	0.348	NS	Preliminary
KIAA0319	Lacunar stroke	rs3181238	1544.5	0.876 (0.824 - 0.932)	2.35e−05	0.014	PASS	0.714	6.73e−04	Strong
FURIN	Lacunar stroke	rs2071410	612.3	1.239 (1.120 - 1.371)	3.05e−05	0.014	PASS	0.007	0.018	Preliminary
TMPRSS5	Lacunar stroke	rs7114195	10247.5	1.047 (1.023 - 1.072)	1.45e−04	0.043	PASS	0.804	NS	Strong
CD40	Lacunar stroke	rs4810485	2367.5	0.911 (0.868 - 0.956)	1.56e−04	0.043	PASS	0.019	0.028	Preliminary
GRK5	Lacunar stroke	rs10886430	1196.6	0.877 (0.818 - 0.940)	2.09e−04	0.043	PASS	0.003	0.034	Preliminary
WARS	Lacunar stroke	rs2273804	907.9	1.166 (1.075 - 1.265)	2.15e−04	0.043	PASS	0.801	4.55e−04	Strong
BTN3A2	WMH	rs9393710	7579.5	0.062 (SE = 0.011)	2.10e−08	3.46e−05	PASS	0.559	NS	Moderate
BTN2A1	WMH	rs72843784	1632.1	−0.132 (SE = 0.025)	1.18e−07	9.72e−05	PASS	0.146	NS	Preliminary
MPI	WMH	rs4886632	689.7	0.179 (SE = 0.036)	6.62e−07	3.64e−04	PASS	0.910	3.11e−05	Strong
EFNA1	WMH	rs4390169	1777.6	−0.102 (SE = 0.022)	5.38e−06	0.002	PASS	0.973	NS	Strong
DNAJC21	WMH	rs35176895	202.2	0.266 (SE = 0.067)	7.12e−05	0.024	PASS	0.940	NS	Strong

Twelve proteins passing FDR < 0.05 among 1890 tested. Effect estimates: OR (95% CI) for lacunar stroke; β (95% CI) for WMH. Abbreviations: SNP, single-nucleotide polymorphism; OR, odds ratio; CI, confidence interval; FDR, false discovery rate; PP.H4, posterior probability of a shared causal variant; SMR, summary-data-based Mendelian randomization; WMH, white matter hyperintensities; NS, not significant.

($p < 0.05$) in both outcomes, the proportion showing concordant effect directions did not exceed random expectation (binomial test $p = 0.68$), further supporting divergent molecular profiles. However, this analysis cannot exclude the possibility of shared mechanisms operating below the FDR threshold or involving proteins not captured by the current pQTL platform. The identification of FURIN and MMP12 is consistent with findings from two recent independent studies, corroborating the reliability of our analytical pipeline [39] [40].

3.2. Proteins Causally Associated with Lacunar Stroke Risk

Causal associations of the seven proteins with lacunar stroke risk are detailed in Table 1 and Figure 2(B). Elevated levels of TMPRSS5, WARS, and FURIN were associated with increased risk, with FURIN showing the largest effect size (OR = 1.239, 95% CI: 1.120 - 1.371, $p = 3.05 \times 10^{-5}$). Elevated levels of KIAA0319, MMP12,

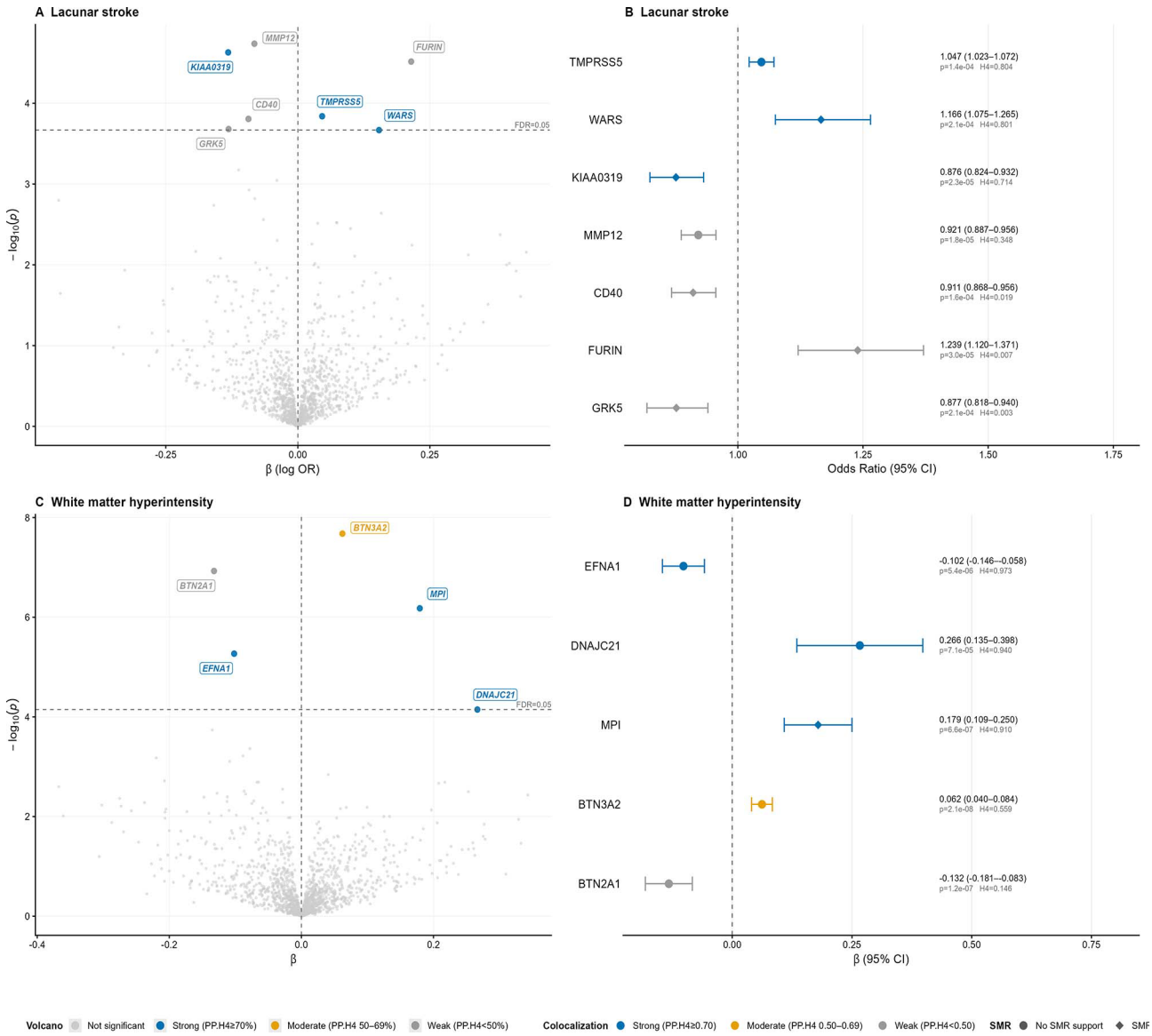

Figure 2. Proteome-Wide MR screening results. Volcano plots (A), (C) and forest plots (B), (D) for lacunar stroke (upper) and WMH (lower). Red dashed line, FDR = 0.05. Abbreviations: MR, Mendelian randomization; WMH, white matter hyperintensities; OR, odds ratio; CI, confidence interval; FDR, false discovery rate.

CD40, and GRK5 were associated with reduced risk, with MMP12 reaching the highest statistical significance ($p = 1.83 \times 10^{-5}$). Colocalization analysis provided strong evidence for shared causal variants between lacunar stroke and TMPRSS5 (PP.H4 = 0.804), WARS (PP.H4 = 0.801), and KIAA0319 (PP.H4 = 0.714) (**Figures 3(G)-(J)** for TMPRSS5 and WARS; KIAA0319 regional plot not shown); the remaining four proteins showed weaker colocalization support (PP.H4 < 0.35). SMR analysis provided independent transcriptomic-level support for WARS, KIAA0319, FURIN, and GRK5 (p_SMR < 0.05). CD40 showed opposing effect directions across two probes and, combined with very low PP.H4 (0.019), was classified as preliminary evidence.

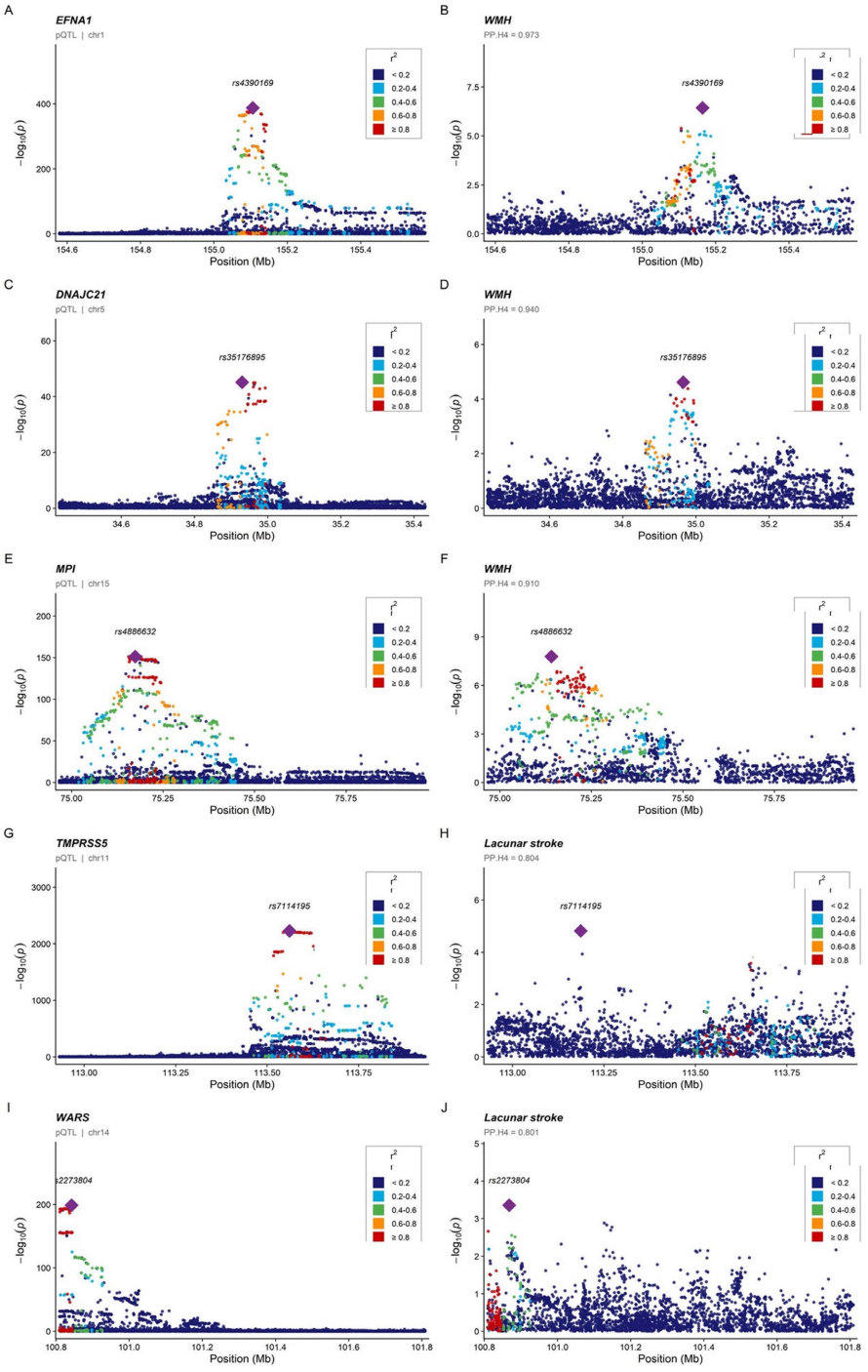

Figure 3. Representative colocalization regional plots. Aligned pQTL (upper) and GWAS (lower) signals for proteins with PP.H4 ≥ 0.70. (A)-(F) WMH proteins: EFNA1, DNAJC21, MPI; (G)-(J) lacunar stroke proteins: TMPRSS5, WARS. Color scale, LD r^2 with lead SNP. Abbreviations: pQTL, protein quantitative trait locus; GWAS, genome-wide association study; WMH, white matter hyperintensities; PP.H4, posterior probability of a shared causal variant; LD, linkage disequilibrium; SNP, single-nucleotide polymorphism.

3.3. Proteins Causally Associated with WMH Burden

Causal associations of the five proteins with WMH burden are detailed in **Table**

1 and Figure 2(D). Elevated levels of DNAJC21 (β = +0.266, p = 7.12×10^{-5}), MPI (β = +0.179, p = 6.62×10^{-7}), and BTN3A2 (β = +0.062, p = 2.10×10^{-8}) were associated with greater WMH burden, with BTN3A2 reaching the highest statistical significance in this study. Elevated levels of EFNA1 (β = −0.102, p = 5.38×10^{-6}) and BTN2A1 (β = −0.132, p = 1.18×10^{-7}) were associated with lower WMH burden, suggesting protective effects. In colocalization analysis, EFNA1 (PP.H4 = 0.973), DNAJC21 (PP.H4 = 0.940), and MPI (PP.H4 = 0.910) demonstrated very strong colocalization evidence (Figures 3(A)-(F)), representing the most well-supported causal targets in this study. BTN3A2 showed intermediate colocalization support (PP.H4 = 0.559) and BTN2A1 weaker support (PP.H4 = 0.146). In SMR analysis, MPI obtained the strongest transcriptomic corroboration (p_SMR = 3.1×10^{-5}); the remaining four proteins did not reach SMR significance, possibly reflecting tissue-specific regulation in brain or vascular endothelium.

3.4. Sensitivity Analyses

F-statistics for all 12 instrumental variables (202.2 - 10247.5) substantially exceeded the threshold for weak instruments, and Steiger tests consistently supported the causal direction from protein to disease (Figure S1, Table S1). GWAS Catalog pleiotropy checks showed that rs7114195 (TMPRSS5) and rs2273804 (WARS) were associated only with their respective protein levels, with no known pleiotropic pathways; no significant pleiotropic associations that could confound causal inference were identified for the remaining SNPs (Figure S2).

4. Discussion

The present study systematically addresses a longstanding clinical question: are WMH and lacunar stroke driven by the same plasma protein molecular network? Our findings show no overlapping FDR-significant associations between the two subtypes, and a formal cross-phenotype concordance analysis across all tested proteins revealed consistently divergent effect profiles. This suggests that the two subtypes are driven by largely—though not necessarily completely—independent molecular mechanisms. However, we emphasize that absence of overlapping FDR-significant signals does not constitute proof of complete mechanistic independence; shared mechanisms operating below the statistical threshold, or involving proteins not measured by the current pQTL platform, cannot be excluded. Nonetheless, the observed divergence at the proteome-wide level provides a molecular rationale for considering WMH and lacunar stroke as distinct entities in future clinical trial stratification. The clinical implications are substantial. Current CSVD clinical trials routinely enroll both imaging phenotypes together without subtype stratification, which may partly explain why some interventions have yielded inconsistent efficacy [41]. Our findings strongly support the evaluation of WMH and lacunar stroke as independent stratified endpoints in future clinical trial designs.

4.1. Interpretation of the Divergent Protein Profiles

Clinically, WMH and lacunar stroke have long been grouped together under the broad umbrella of CSVD, and this "one-size-fits-all" perspective has to some extent constrained the development of precision therapies [6]. The absence of overlapping FDR-significant protein associations, together with the divergent effect directions observed across nominally significant signals, suggests that these two subtypes may operate through substantially different molecular pathways. However, we caution against overinterpreting this divergence as proof of complete mechanistic independence; shared mechanisms—particularly those involving proteins not captured by the current platform or operating below the FDR threshold—remain possible. Research into monogenic CSVD (such as CADASIL) has suggested that small vessel disease arising from different genetic backgrounds operates through distinct pathological mechanisms; the present study provides analogous molecular-genetic support at the level of sporadic CSVD [42] [43]. Two recent independent proteomic MR studies each focused on specific stroke subtypes but neither performed a cross-subtype comparison of WMH versus lacunar stroke [39] [40]. By simultaneously incorporating both subtypes and directly comparing their protein target profiles, the present study fills this methodological gap and reveals a structural finding that is only discernible through systematic comparison. This strongly suggests that treatment strategies effective for one CSVD subtype may be entirely ineffective for the other, and that subtype-specific precision intervention is the necessary direction forward.

4.2. Lacunar Stroke Targets: Novel Findings and Independent Validation

Among the seven lacunar stroke-associated proteins, the identification of FURIN and MMP12 is consistent with independent findings from Zhao *et al.* and Liu *et al.* respectively, and this cross-study concordance enhances confidence in these two targets [39] [40]. Among the novel findings, elevated transmembrane serine protease 5 (TMPRSS5) increased risk; the role of this protease in vascular biology remains poorly characterized, and it may affect vascular wall integrity or inflammatory responses through cleavage of specific substrates—no specific inhibitors are currently available and the mechanism requires further elucidation. Elevated tryptophanyl-tRNA synthetase (WARS) also increased risk; beyond its canonical role in protein synthesis, this enzyme can function as a cytokine involved in inflammatory and immune responses, suggesting that immune dysregulation may contribute to lacunar stroke pathogenesis. This is conceptually consistent with prior reports linking inflammatory cytokines to CSVD. Although WARS inhibitors have been the subject of recent patent applications, it is important to emphasize that substantial preclinical and clinical work remains to establish target specificity, direction of modulation, CNS exposure, and safety before any therapeutic application can be considered [38] [44]. KIAA0319, a protein associated with neurodevelopment, showed elevated levels conferring a protective effect, suggesting

that pathways maintaining neuronal or glial homeostasis are important for preventing lacunar stroke; the protective role of cognitive reserve and neural plasticity observed in WMH-related research may operate through similar mechanisms in lacunar stroke [13]. FURIN participates in the activation of multiple growth factors and cytokines, and its inhibitors (e.g., BOS-981, MI-1851) have entered preclinical or phase I clinical investigation, indicating translational drug development potential. Elevated MMP12 showed a protective association, possibly through extracellular matrix remodeling that maintains healthy vascular wall architecture; it is noteworthy that hemorrhagic stroke and ischemic small vessel disease share common pathological substrates in white matter injury, and the protective role of MMP12 in extracellular matrix remodeling may operate through this shared pathway [45].

4.3. WMH Targets: Novelty and Multi-System Pathways

None of the five WMH-associated proteins has been previously reported in the literature. Elevated mannose-6-phosphate isomerase (MPI) increased WMH burden; as a key node in glucose metabolism, MPI suggests that disturbances in carbohydrate metabolism may directly participate in WMH pathology. Associations between visceral fat, metabolic abnormalities, and CSVD have been reported, and MPI provides a new molecular node for the metabolic-vascular pathway [18] [46]. MPI also obtained the strongest SMR support ($p\_SMR = 3.1 \times 10^{-5}$), further strengthening the causal evidence. Elevated ephrin-A1 (EFNA1) conferred protection; as a molecule involved in vascular endothelial signaling, it may attenuate WMH progression by promoting angiogenesis or maintaining endothelial barrier function [47]. The genetic association between aortic distensibility and WMH implicates vascular compliance in WMH development, and EFNA1-mediated endothelial signaling may represent the molecular basis of this pathway [34]. EFNA1 showed the strongest colocalization evidence in this study (PP.H4 = 0.973), indicating a highly credible shared causal variant between its genetic regulation and WMH. DNAJC21, a member of the HSP40 co-chaperone family, showed elevated levels associated with increased WMH risk, possibly linked to cellular stress responses and protein misfolding; the synergistic relationship between neurodegeneration and WMH suggests that proteostasis failure may be an important mechanism driving WMH progression [12]. BTN3A2 (risk factor) and BTN2A1 (protective factor) both belong to the Butyrophilin family of immune regulatory proteins, and their opposing effect directions suggest that fine-tuned regulation of adaptive or innate immunity plays an important role in WMH pathogenesis. Prior research has reported associations between leukocyte telomere length and cardiovascular disease risk, and immune aging may serve as a potential bridge linking immune function to WMH [27]. The antibody drug ICT01, targeting BTN3A2, has entered phase I/II clinical trials in oncology (EVICTION trial), providing a direct lead for drug repurposing in WMH. Furthermore, lymphatic dysfunction has been demonstrated to correlate with CSVD severity, and the immune regula-

tory functions of Butyrophilin family proteins may contribute to WMH development through effects on intracerebral waste clearance [48].

4.4. Drug Tractability and Subtype-Specific Therapeutic Strategies

The findings of the present study nominate several subtype-specific candidate targets for further investigation. For risk-increasing proteins (lacunar stroke: FURIN, WARS; WMH: BTN3A2, MPI, DNAJC21), inhibitory or downregulatory strategies may be worth exploring; for protective proteins (lacunar stroke: MMP12, CD40, GRK5, KIAA0319; WMH: EFNA1, BTN2A1), strategies to activate their function or downstream pathways could be considered. However, we emphasize that these genetic findings represent hypothesis-generating evidence only; therapeutic translation will require rigorous validation of target specificity, the appropriate direction of pharmacological modulation, CNS exposure, and safety profiles before any clinical application can be considered. For protective proteins (lacunar stroke: MMP12, CD40, GRK5, KIAA0319; WMH: EFNA1, BTN2A1), strategies to activate their function or downstream pathways should be explored. This subtype-specific approach grounded in genetic evidence has the potential to break through the current impasse of limited effective interventions in CSVD [41] [49]. The hemorrhagic complications associated with cerebral amyloid angiopathy and the differential effects of antiplatelet therapy across imaging subtypes both underscore that imaging- and molecular-subtype-guided precision intervention is key to improving CSVD treatment efficacy.

4.5. Limitations

This study has several limitations. First, single-SNP instrumental variables were used, precluding sensitivity analyses such as MR-Egger that require multiple instruments; however, the cis-pQTL design itself substantially reduces horizontal pleiotropy risk, and all F-statistics substantially exceeded the conventional weak instrument threshold of 10 (range: 202 - 10,248), ruling out weak instrument bias. Second, the HEIDI test could not be performed; instead, we used Bayesian colocalization to assess whether MR signals were driven by shared causal variants versus LD. Colocalization and HEIDI address the same conceptual question but are not statistically equivalent; therefore, SMR findings without HEIDI testing should be interpreted as exploratory. Third, pQTL and eQTL data were primarily derived from blood, whereas CSVD is a brain disease, and tissue specificity may influence some results; blood protein pQTLs have been validated for drug target discovery across multiple diseases, supporting the external validity of our approach. Fourth, all GWAS data were derived from European-ancestry populations, and generalizability to other ethnic groups requires further investigation; however, the consistency of effect directions supports the existence of true associations. Finally, for targets with weaker colocalization evidence (e.g., FURIN, GRK5, BTN2A1), causal relationships should be considered preliminary and await functional experimental confirmation.

5. Conclusion

In summary, through proteome-wide Mendelian randomization analysis, the present study systematically identified subtype-specific plasma protein targets for the two principal subtypes of CSVD—lacunar stroke and WMH. The absence of overlapping FDR-significant protein profiles, together with the divergent effect directions observed across nominally significant associations, suggests that the two subtypes are driven by largely independent molecular mechanisms. However, this interpretation is based on the current set of measurable plasma proteins and statistical thresholds; shared mechanisms cannot be definitively excluded. This study independently validated recently identified targets including FURIN and MMP12, while identifying a series of novel candidate targets including TMPRSS5, WARS, and MPI. Among these, FURIN, WARS, and BTN3A2 have existing investigational agents, nominating them as tractable targets for further pharmacological evaluation. However, we caution that these findings are hypothesis-generating and do not imply immediate clinical translatability; extensive preclinical and clinical validation is required. These findings deepen our understanding of CSVD heterogeneity and establish a molecular foundation for ultimately achieving individualized treatment tailored to different patient subtypes.

Acknowledgments

We thank all participants and investigators of the original genome-wide association studies for making their summary-level data publicly available.

Author Contributions

HG: Conceptualization; data curation; formal analysis; methodology; visualization; writing - original draft. XS: Data interpretation; writing - review and editing. JZ: Conceptualization; funding acquisition; supervision; writing - review and editing.

Funding

This work was supported by the Hebei Provincial Department of Science and Technology (grant number: 20377764D).

Data Availability Statement

The plasma proteomic pQTL summary statistics were obtained from Sun *et al.* (2023) (Nature 622:329-338) and are available through the supplementary data of that publication and the UK Biobank data access mechanism (https://www.ukbiobank.ac.uk/). GWAS summary statistics for lacunar stroke (GCST006906) are available via the GWAS Catalog (https://www.ebi.ac.uk/gwas/). GWAS summary statistics for WMH are available from the corresponding author of Patel *et al.* (2024) upon reasonable request.

Ethics Approval and Consent to Participate

Ethical review and approval were not required for this study, which used only publicly accessible summary-level GWAS data. No individual-level data were accessed, and no participants were directly involved. Clinical trial number: not applicable.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Pantoni, L. (2010) Cerebral Small Vessel Disease: From Pathogenesis and Clinical Characteristics to Therapeutic Challenges. *The Lancet Neurology*, **9**, 689-701. https://doi.org/10.1016/s1474-4422(10)70104-6

[2] Wardlaw, J.M., Smith, C. and Dichgans, M. (2019) Small Vessel Disease: Mechanisms and Clinical Implications. *The Lancet Neurology*, **18**, 684-696. https://doi.org/10.1016/s1474-4422(19)30079-1

[3] Chauhan, G., Arnold, C.R., Chu, A.Y., Fornage, M., Reyahi, A., Bis, J.C., *et al.* (2016) Identification of Additional Risk Loci for Stroke and Small Vessel Disease: A Meta-Analysis of Genome-Wide Association Studies. *The Lancet Neurology*, **15**, 695-707. https://doi.org/10.1016/s1474-4422(16)00102-2

[4] Persyn, E., Hanscombe, K.B., Howson, J.M.M., Lewis, C.M., Traylor, M. and Markus, H.S. (2020) Genome-Wide Association Study of MRI Markers of Cerebral Small Vessel Disease in 42,310 Participants. *Nature Communications*, **11**, Article No. 2175. https://doi.org/10.1038/s41467-020-15932-3

[5] Sargurupremraj, M., Suzuki, H., Jian, X., Sarnowski, C., Evans, T.E., Bis, J.C., *et al.* (2020) Cerebral Small Vessel Disease Genomics and Its Implications across the Lifespan. *Nature Communications*, **11**, Article No. 6285. https://doi.org/10.1038/s41467-020-19111-2

[6] Wardlaw, J.M., Smith, E.E., Biessels, G.J., Cordonnier, C., Fazekas, F., Frayne, R., *et al.* (2013) Neuroimaging Standards for Research into Small Vessel Disease and Its Contribution to Ageing and Neurodegeneration. *The Lancet Neurology*, **12**, 822-838. https://doi.org/10.1016/s1474-4422(13)70124-8

[7] Duering, M., Biessels, G.J., Brodtmann, A., Chen, C., Cordonnier, C., de Leeuw, F., *et al.* (2023) Neuroimaging Standards for Research into Small Vessel Disease—Advances since 2013. *The Lancet Neurology*, **22**, 602-618. https://doi.org/10.1016/s1474-4422(23)00131-x

[8] Zhou, R., Zhong, W., Xia, Y., Ying, Y., Wang, Y., Yang, L., *et al.* (2025) The Relationship between Cerebral Small Vessel Disease, Sleep Quality, and Cognitive Impairment among Community-Dwelling Older Adults: Exploring the Role of Glymphatic Function. *European Journal of Neurology*, **32**, e70384. https://doi.org/10.1111/ene.70384

[9] Zhang, X., Tang, Y., Xie, Y., Ding, C., Xiao, J., Jiang, X., *et al.* (2017) Total Magnetic Resonance Imaging Burden of Cerebral Small-Vessel Disease Is Associated with Post-Stroke Depression in Patients with Acute Lacunar Stroke. *European Journal of Neurology*, **24**, 374-380. https://doi.org/10.1111/ene.13213

[10] Zuurbier, L.A., Ikram, M.A., Luik, A.I., Hofman, A., Van Someren, E.J.W., Vernooij, M.W., *et al.* (2015) Cerebral Small Vessel Disease Is Related to Disturbed 24-h Activ-

ity Rhythms: A Population-Based Study. *European Journal of Neurology*, **22**, 1482-1487. https://doi.org/10.1111/ene.12775

[11] Wardlaw, J.M., Smith, C. and Dichgans, M. (2013) Mechanisms of Sporadic Cerebral Small Vessel Disease: Insights from Neuroimaging. *The Lancet Neurology*, **12**, 483-497. https://doi.org/10.1016/s1474-4422(13)70060-7

[12] Parent, O., Alasmar, Z., Osborne, S., Bussy, A., Costantino, M., Fouquet, J.P., *et al.* (2026) Characterizing Spatiotemporal White Matter Hyperintensity Pathophysiology *in Vivo* to Disentangle Vascular and Neurodegenerative Contributions. *Nature Communications*, **17**, Article No. 4623. https://doi.org/10.1038/s41467-026-70832-2

[13] Arola, A., Laakso, H.M., Pitkänen, J., Koikkalainen, J., Lötjönen, J., Korvenoja, A., *et al.* (2021) Associations of Cognitive Reserve and Psychological Resilience with Cognitive Functioning in Subjects with Cerebral White Matter Hyperintensities. *European Journal of Neurology*, **28**, 2622-2630. https://doi.org/10.1111/ene.14910

[14] Fruhwirth, V., Enzinger, C., Fandler-Höfler, S., Kneihsl, M., Eppinger, S., Ropele, S., *et al.* (2021) Baseline White Matter Hyperintensities Affect the Course of Cognitive Function after Small Vessel Disease-Related Stroke: A Prospective Observational Study. *European Journal of Neurology*, **28**, 401-410. https://doi.org/10.1111/ene.14593

[15] Park, J.H., Heo, S.H., Lee, M.H., Kwon, H.S., Kwon, S.U. and Lee, J.S. (2019) White Matter Hyperintensities and Recurrent Stroke Risk in Patients with Stroke with Small-Vessel Disease. *European Journal of Neurology*, **26**, 911-918. https://doi.org/10.1111/ene.13908

[16] Jokinen, H., Laakso, H.M., Ahlström, M., Arola, A., Lempiäinen, J., Pitkänen, J., *et al.* (2022) Synergistic Associations of Cognitive and Motor Impairments with Functional Outcome in Covert Cerebral Small Vessel Disease. *European Journal of Neurology*, **29**, 158-167. https://doi.org/10.1111/ene.15108

[17] Jokinen, H., Laakso, H.M., Arola, A., Paajanen, T.I., Virkkala, J., Särkämö, T., *et al.* (2025) Executive Functions and Processing Speed in Covert Cerebral Small Vessel Disease. *European Journal of Neurology*, **32**, e16533. https://doi.org/10.1111/ene.16533

[18] Yamashiro, K., Tanaka, R., Tanaka, Y., Miyamoto, N., Shimada, Y., Ueno, Y., *et al.* (2014) Visceral Fat Accumulation Is Associated with Cerebral Small Vessel Disease. *European Journal of Neurology*, **21**, 667-673. https://doi.org/10.1111/ene.12374

[19] Cao, Y., Su, N., Zhang, D., Zhou, L., Yao, M., Zhang, S., *et al.* (2021) Correlation between Total Homocysteine and Cerebral Small Vessel Disease: A Mendelian Randomization Study. *European Journal of Neurology*, **28**, 1931-1938. https://doi.org/10.1111/ene.14708

[20] Patankar, T., Widjaja, E., Chant, H., McCollum, C., Baldwin, R., Jeffries, S., *et al.* (2006) Relationship of Deep White Matter Hyperintensities and Cerebral Blood Flow in Severe Carotid Artery Stenosis. *European Journal of Neurology*, **13**, 10-16. https://doi.org/10.1111/j.1468-1331.2006.01115.x

[21] Miwa, K., Koga, M., Tanaka, K., Yakushiji, Y., Sasaki, M., Kudo, K., *et al.* (2026) Impact of Intracranial Atherosclerosis and Cerebral Small Vessel Disease on Overt Vascular Events during Antithrombotic Therapy. *European Journal of Neurology*, **33**, e70635. https://doi.org/10.1111/ene.70635

[22] Wang, Y., Cai, X., Li, H., Jin, A., Jiang, L., Chen, W., *et al.* (2023) Association of Intracranial Atherosclerosis with Cerebral Small Vessel Disease in a Community-Based Population. *European Journal of Neurology*, **30**, 2700-2712. https://doi.org/10.1111/ene.15908

[23] Park, J.H., Kwon, H.M., Lee, J., Kim, D.S. and Ovbiagele, B. (2015) Association of Intracranial Atherosclerotic Stenosis with Severity of White Matter Hyperintensities. *European Journal of Neurology*, **22**, 44-e3. https://doi.org/10.1111/ene.12431

[24] Lin, Z. and Pan, W. (2024) A Robust Cis-Mendelian Randomization Method with Application to Drug Target Discovery. *Nature Communications*, **15**, Article No. 6072. https://doi.org/10.1038/s41467-024-50385-y

[25] Chen, L., Peters, J.E., Prins, B., Persyn, E., Traylor, M., Surendran, P., *et al.* (2022) Systematic Mendelian Randomization Using the Human Plasma Proteome to Discover Potential Therapeutic Targets for Stroke. *Nature Communications*, **13**, Article No. 6143. https://doi.org/10.1038/s41467-022-33675-1

[26] Shah, A.M., Myhre, P.L., Arthur, V., Dorbala, P., Rasheed, H., Buckley, L.F., *et al.* (2024) Large Scale Plasma Proteomics Identifies Novel Proteins and Protein Networks Associated with Heart Failure Development. *Nature Communications*, **15**, Article No. 528. https://doi.org/10.1038/s41467-023-44680-3

[27] Levin, M.G., Tsao, N.L., Singhal, P., Liu, C., Vy, H.M.T., Paranjpe, I., *et al.* (2022) Genome-Wide Association and Multi-Trait Analyses Characterize the Common Genetic Architecture of Heart Failure. *Nature Communications*, **13**, Article No. 6914. https://doi.org/10.1038/s41467-022-34216-6

[28] Zaghlool, S.B., Sharma, S., Molnar, M., Matías-García, P.R., Elhadad, M.A., Waldenberger, M., *et al.* (2021) Revealing the Role of the Human Blood Plasma Proteome in Obesity Using Genetic Drivers. *Nature Communications*, **12**, Article No. 1279. https://doi.org/10.1038/s41467-021-21542-4

[29] Qiao, J., Wang, Q., Zhao, Y., Chang, M., Sun, S., Zhang, P., *et al.* (2025) Contribution of Leukocyte Telomere Length to Cardiovascular Disease Onset from Genome-Wide Cross-Trait Analysis. *Nature Communications*, **16**, Article No. 8677. https://doi.org/10.1038/s41467-025-63707-5

[30] Ren, P., Hou, X., Li, Z., You, J., Li, Y., Zhang, W., *et al.* (2025) Atlas of Proteomic Signatures of Brain Structure and Its Links to Brain Disorders. *Nature Communications*, **16**, Article No. 5092. https://doi.org/10.1038/s41467-025-60185-7

[31] Sun, B.B., Chiou, J., Traylor, M., Benner, C., Hsu, Y., Richardson, T.G., *et al.* (2023) Plasma Proteomic Associations with Genetics and Health in the UK Biobank. *Nature*, **622**, 329-338. https://doi.org/10.1038/s41586-023-06592-6

[32] Xu, F., Yu, E.Y., Cai, X., Yue, L., Jing, L., Liang, X., *et al.* (2023) Genome-Wide Genotype-Serum Proteome Mapping Provides Insights into the Cross-Ancestry Differences in Cardiometabolic Disease Susceptibility. *Nature Communications*, **14**, Article No. 896. https://doi.org/10.1038/s41467-023-36491-3

[33] Traylor, M., Persyn, E., Tomppo, L., Klasson, S., Abedi, V., Bakker, M.K., *et al.* (2021) Genetic Basis of Lacunar Stroke: A Pooled Analysis of Individual Patient Data and Genome-Wide Association Studies. *The Lancet Neurology*, **20**, 351-361. https://doi.org/10.1016/s1474-4422(21)00031-4

[34] Francis, C.M., Futschik, M.E., Huang, J., Bai, W., Sargurupremraj, M., Teumer, A., *et al.* (2022) Genome-Wide Associations of Aortic Distensibility Suggest Causality for Aortic Aneurysms and Brain White Matter Hyperintensities. *Nature Communications*, **13**, Article No. 4505. https://doi.org/10.1038/s41467-022-32219-x

[35] Patel, Y., Shin, J., Sliz, E., Tang, A., Mishra, A., Xia, R., *et al.* (2024) Genetic Risk Factors Underlying White Matter Hyperintensities and Cortical Atrophy. *Nature Communications*, **15**, Article No. 9517. https://doi.org/10.1038/s41467-024-53689-1

[36] Giambartolomei, C., Vukcevic, D., Schadt, E.E., Franke, L., Hingorani, A.D., Wallace, C., *et al.* (2014) Bayesian Test for Colocalisation between Pairs of Genetic Association

Studies Using Summary Statistics. *PLOS Genetics*, **10**, e1004383. https://doi.org/10.1371/journal.pgen.1004383

[37] Westra, H., Peters, M.J., Esko, T., Yaghootkar, H., Schurmann, C., Kettunen, J., *et al.* (2013) Systematic Identification of Trans eQTLs as Putative Drivers of Known Disease Associations. *Nature Genetics*, **45**, 1238-1243. https://doi.org/10.1038/ng.2756

[38] Li, Z., Miao, L., Zhang, T., Thomas, A.M. and Li, S. (2024) Causal Relationship of Inflammatory Cytokines and Serum Metabolites in Cerebral Small Vessel Disease: A Two-Step Mendelian Randomization Study. *European Journal of Neurology*, **31**, e16443. https://doi.org/10.1111/ene.16443

[39] Zhao, X., He, M., Zhou, D., Li, Z., Liu, L., Yang, R., *et al.* (2025) Proteome-Wide Mendelian Randomization and Colocalization Analysis Identify Therapeutic Targets for Stroke. *BMC Neurology*, **25**, Article No. 255. https://doi.org/10.1186/s12883-025-04239-9

[40] Liu, H., Zhang, X., Ma, H., Nguyen, T.N., Ling, Y., Mo, S., *et al.* (2026) Identification of Potential Therapeutic Targets for Stroke and Its Subtypes by Integrating Proteomes and Genetics from Human Plasma. *Brain Communications*, **8**, fcaf457. https://doi.org/10.1093/braincomms/fcaf457

[41] Al-Shahi Salman, R., Minks, D.P., Mitra, D., Rodrigues, M.A., Bhatnagar, P., du Plessis, J.C., *et al.* (2019) Effects of Antiplatelet Therapy on Stroke Risk by Brain Imaging Features of Intracerebral Haemorrhage and Cerebral Small Vessel Diseases: Subgroup Analyses of the RESTART Randomised, Open-Label Trial. *The Lancet Neurology*, **18**, 643-652. https://doi.org/10.1016/s1474-4422(19)30184-x

[42] Shen, Y., Chen, Y., Cheng, Y., Chen, C., Jeng, J. and Tang, S. (2024) Characteristics and Temporal Evolution of Asymptomatic Diffusion-Weighted Imaging Lesions in Patients with Cerebral Autosomal Dominant Arteriopathy with Subcortical Infarcts and Leukoencephalopathy (CADASIL). *European Journal of Neurology*, **31**, e16519. https://doi.org/10.1111/ene.16519

[43] Mancuso, M., Arnold, M., Bersano, A., Burlina, A., Chabriat, H., Debette, S., *et al.* (2020) Monogenic Cerebral Small-Vessel Diseases: Diagnosis and Therapy. Consensus Recommendations of the European Academy of Neurology. *European Journal of Neurology*, **27**, 909-927. https://doi.org/10.1111/ene.14183

[44] Miwa, K., Okazaki, S., Sakaguchi, M., Mochizuki, H. and Kitagawa, K. (2016) Interleukin-6, Interleukin-6 Receptor Gene Variant, Small-Vessel Disease and Incident Dementia. *European Journal of Neurology*, **23**, 656-663. https://doi.org/10.1111/ene.12921

[45] Cheng, Z., Zhang, W., Zhan, Z., Xia, L. and Han, Z. (2022) Cerebral Small Vessel Disease and Prognosis in Intracerebral Haemorrhage: A Systematic Review and Meta-Analysis of Cohort Studies. *European Journal of Neurology*, **29**, 2511-2525. https://doi.org/10.1111/ene.15363

[46] Makedonov, I., Black, S.E. and MacIntosh, B.J. (2012) Cerebral Small Vessel Disease in Aging and Alzheimer's Disease: A Comparative Study Using MRI and SPECT. *European Journal of Neurology*, **20**, 243-250. https://doi.org/10.1111/j.1468-1331.2012.03785.x

[47] Movérare-Skrtic, S., Nethander, M., Li, L., Chu, N.T.L., Dregval, O., Tian, X., *et al.* (2026) Identification of Ephrin-A1-EphA2 Signalling as a Potential Target for Fracture Prevention. *Nature Communications*, **17**, Article No. 1988. https://doi.org/10.1038/s41467-026-69863-6

[48] Xu, J., Su, Y., Fu, J., Wang, X., Nguchu, B.A., Qiu, B., *et al.* (2022) Glymphatic Dysfunction Correlates with Severity of Small Vessel Disease and Cognitive Impairment

in Cerebral Amyloid Angiopathy. *European Journal of Neurology*, **29**, 2895-2904. https://doi.org/10.1111/ene.15450

[49] Charidimou, A., Boulouis, G., Frosch, M.P., Baron, J., Pasi, M., Albucher, J.F., *et al.* (2022) The Boston Criteria Version 2.0 for Cerebral Amyloid Angiopathy: A Multicentre, Retrospective, MRI-Neuropathology Diagnostic Accuracy Study. *The Lancet Neurology*, **21**, 714-725. https://doi.org/10.1016/s1474-4422(22)00208-3

Appendix

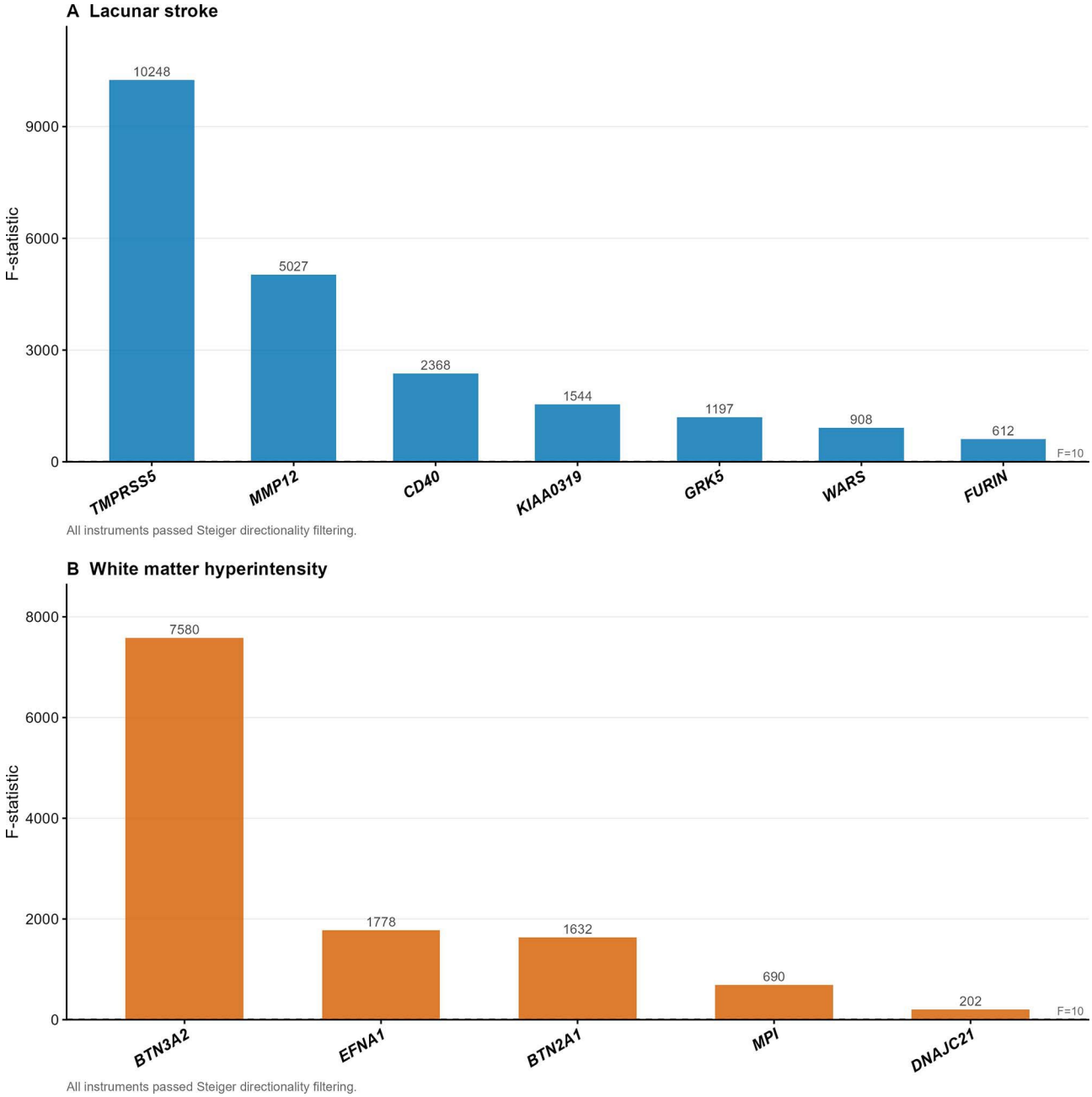

Figure S1. Instrumental variable strength. F-statistics for 12 cis-pQTL instruments. (A) Lacunar stroke; (B) WMH. Dashed line, weak instrument threshold (F = 10). All instruments passed Steiger directionality filtering. Abbreviations: cis-pQTL, cis-acting protein quantitative trait locus; WMH, white matter hyperintensities.

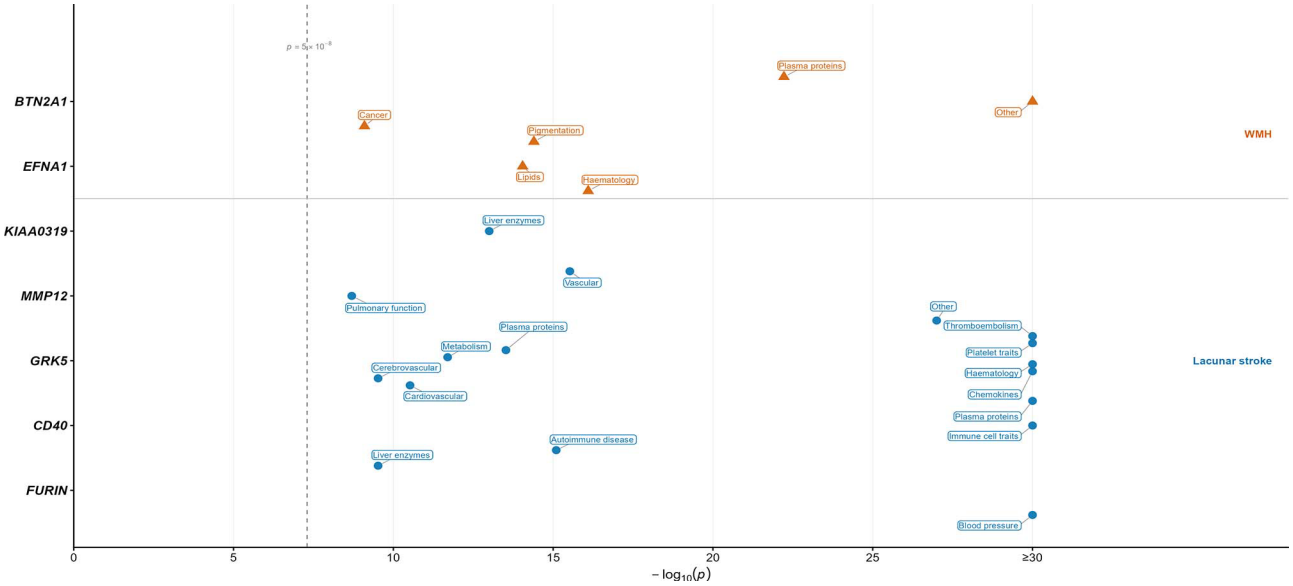

Figure S2. Pleiotropy screening. GWAS Catalog phenotypic associations for 12 lead cis-pQTL SNPs ($-\log_{10}p$). Dashed line, genome-wide significance ($p = 5 \times 10^{-8}$). Orange triangles, WMH proteins; blue circles, lacunar stroke proteins. Abbreviations: GWAS, genome-wide association study; SNP, single-nucleotide polymorphism; cis-pQTL, cis-acting protein quantitative trait locus; WMH, white matter hyperintensities.

Table S1. Instrumental variable characteristics.

Protein	Outcome	Lead SNP	F-statistic	r2 (exposure)	Steiger	OR (95% CI)	P-value	FDR	Coloc nSNPs	PP.H3	PP.H4
TMPRSS5	Lacunar stroke	rs7114195	10247.5	0.229	PASS	1.047 (1.023 - 1.072)	1.45e−04	0.043	2710	0.069	0.804
WARS	Lacunar stroke	rs2273804	907.9	0.0256	PASS	1.166 (1.075 - 1.265)	2.15e−04	0.043	3024	0.037	0.801
KIAA0319	Lacunar stroke	rs3181238	1544.5	0.0428	PASS	0.876 (0.824 - 0.932)	2.35e−05	0.014	3918	0.268	0.714
MMP12	Lacunar stroke	rs17368814	5027.4	0.1270	PASS	0.921 (0.887 - 0.956)	1.83e−05	0.014	3092	0.650	0.348
CD40	Lacunar stroke	rs4810485	2367.5	0.0641	PASS	0.911 (0.868 - 0.956)	1.56e−04	0.043	2233	0.046	0.019
FURIN	Lacunar stroke	rs2071410	612.3	0.0174	PASS	1.239 (1.120 - 1.371)	3.05e−05	0.014	3091	0.070	0.007
GRK5	Lacunar stroke	rs10886430	1196.6	0.0335	PASS	0.877 (0.818 - 0.940)	2.09e−04	0.043	3445	0.008	0.003
EFNA1	WMH	rs4390169	1777.6	0.0489	PASS	0.903 (0.864 - 0.944)	5.38e−06	0.002	1890	0.020	0.973
DNAJC21	WMH	rs35176895	202.2	0.0058	PASS	1.305 (1.144 - 1.489)	7.12e−05	0.024	4042	0.014	0.940
MPI	WMH	rs4886632	689.7	0.0196	PASS	1.196 (1.115 - 1.284)	6.62e−07	3.64e−04	1928	0.090	0.910
BTN3A2	WMH	rs9393710	7579.5	0.1800	PASS	1.064 (1.041 - 1.087)	2.10e−08	3.46e−05	3085	0.441	0.559
BTN2A1	WMH	rs72843784	1632.1	0.0451	PASS	0.876 (0.834 - 0.920)	1.18e−07	9.72e−05	3140	0.854	0.146

Abbreviations: r^2, variance in protein levels explained by IV; Steiger, PASS = IV explains more variance in protein than outcome; PP.H3/H4, posterior probability of distinct/shared causal variants. All IVs: $p < 5 \times 10^{-8}$, LD-clumped ($r^2 < 0.001$, 10,000 kb window). F-statistics and r^2 values were calculated using the per-protein effective sample size from the UKB-PPP summary statistics, which varied due to protein-level QC filtering (range: approximately 34,000 to 54,000).