

<https://doi.org/10.1038/s43856-026-01734-z>

Integrative proteogenomic analyses identify plasma proteins that impact the risk of ischemic stroke

Check for updates

Lazaros Belbasis , Adam von Ende, Parag Gajendragadkar, Elsa Valdes-Marquez, Federico Murgia , Cornelia van Duijn & Jemma C. Hopewell

Abstract

Background Understanding the proteins implicated in the pathogenesis of ischemic stroke is important for elucidating disease mechanisms and informing prevention strategies. In this study, we aim to identify plasma proteins with a potentially causal effect on risk of ischemic stroke by integrating the largest available genetic data sources for plasma proteins, ischemic stroke and its risk factors.

Methods We use genome-wide association (GWA) statistics to identify *cis* protein quantitative trait loci for over 3,500 proteins across two proteomic platforms. Subsequently, we perform two-sample Mendelian randomization (MR) to assess the potentially causal relationships between plasma proteins and a) risk of ischemic stroke and its subtypes, and b) well-established cardiovascular risk factors. Downstream analyses include Bayesian colocalization, phenome-wide associations and interrogation of biological databases.

Results We identify 21 proteins with evidence of potentially causal associations with risk of ischemic stroke or its subtypes at 5% false discovery rate, with 16 supported further by colocalization. Four proteins (CEP85, KNG1, MMUT, SPATA20) represent findings not previously implicated in ischemic stroke through MR, or through cognate genes in GWA studies. Integration of evidence from phenome-wide MR, animal models and tissue-specific gene expression highlights agonists of MMUT, CEP85 and GRK5, and inhibitors of F11 and KNG1 as the most promising for further consideration as targets for prevention of ischemic stroke.

Conclusions Our study provides the most comprehensive data integration to date supporting the identification and causal relevance of plasma proteins for ischemic stroke and implicating a number of potential therapeutic targets.

Plain language summary

Ischemic stroke occurs when the blood supply to a part of the brain is interrupted. It is a major cause of death and long-term disability worldwide, but we still do not fully understand the underlying biological processes. Proteins circulating in the blood may help explain why some people are at higher risk. They may also point to new treatments. In this study, we use large-scale genetic data from over a million people. We aim to identify specific proteins in the blood that may play a role in ischemic stroke. After examining more than 3500 proteins, we identified 11 proteins relevant to ischemic stroke. After considering additional existing knowledge, 4 of them appear to be the most promising as potential drug targets. In summary, our findings improve our understanding of stroke biology and highlight new opportunities for preventing ischemic stroke.

Stroke is one of the leading contributors to disability and mortality globally¹, with ischemic stroke comprising approximately 70% of all cerebrovascular events². Ischemic stroke is a heterogeneous clinical condition. Cardioembolic stroke, large artery stroke, and small vessel stroke are the most prevalent subtypes of ischemic stroke, each with distinct aetiologies and risk factors leading to potential implications for prevention and treatment².

Proteins are crucial for understanding disease pathogenesis, and their potential causal role in disease may also have major implications for the identification of drug targets³. Technological advances have made it possible to measure thousands of proteins in large population cohorts⁴, and genome-

wide association studies (GWAS) have identified numerous protein quantitative trait loci (pQTLs)⁵. These pQTLs can be used as genetic proxies for protein levels within a Mendelian randomization (MR) framework to understand the potential causal role of a protein in a disease³.

Two previous proteome-wide MR studies identified ten plasma proteins (ABO, LPA, TNFSF12, F11, MMP12, SCARA5, CD40, TFPI, TMPRSS5, and CD6) related to ischemic stroke^{6,7}. However, the emergence of large-scale cohorts measuring thousands of proteins on diverse platforms and the expansion of GWAS data for both proteins and ischemic stroke, as well as more detailed bioinformatics resources, motivate a more

comprehensive approach to identify proteins involved in the pathogenesis of ischemic stroke and to highlight proteins that may be suitable as drug targets for the prevention of ischemic stroke.

In the present study, we aim to (1) identify plasma proteins associated with risk of ischemic stroke and its subtypes using a two-sample MR and colocalization framework, (2) examine the effect of stroke-associated proteins on well-established cardiovascular risk factors, and (3) assess the potential relevance of stroke-associated proteins as drug targets using a comprehensive range of bioinformatics resources.

Our genetic study implicates seven proteins measured on the Olink platform (GRK5, F11, MMP12, FURIN, MMUT, CEP85, PROCR) and five proteins measured on the SomaScan platform (F11, MMP12, KNG1, SH3BGRL3, LRP4) that have potentially causal relevance for ischemic stroke. In subtype-specific analyses, we also show that F11 is implicated in the risk of cardioembolic stroke, MMP12 in the risk of large artery stroke, and SPATA20 in the risk of small vessel stroke, supported by MR and colocalization. After integrating evidence from phenome-wide MR, animal models, and tissue-specific gene expression, we highlight agonists of MMUT, CEP85, and GRK5, and inhibitors of F11 and KNG1 as the most promising targets for the prevention of ischemic stroke.

Methods

Data sources

GWAS of the human plasma proteome. Genome-wide association statistics were publicly available from two large proteogenomic studies in populations of European ancestry (Supplementary Data 1)^{8,9}. In summary, we used data from previously conducted GWAS for circulating protein levels, including: (1) 35,571 UK Biobank participants based on the Olink Explore 3072 platform, which measured 2941 protein analytes representing 2923 unique proteins⁸, and (2) 35,559 participants from the Icelandic Cancer Project and deCODE Health Study based on the SomaScan version 4 platform, which measured 4907 aptamers representing 4719 proteins⁹.

GWAS of ischemic stroke and its subtypes. Genome-wide association statistics were publicly available from the GIGASTROKE Consortium. We considered summary statistics from European ancestry cohorts examining total ischemic stroke and ischemic stroke subtypes (i.e., cardioembolic stroke, large artery stroke, and small vessel disease) (Supplementary Data 1)¹⁰. The sample overlap between these data and the two proteogenomic studies ranges from 0 to 3%.

Selection of instrumental variables and data harmonization. Autosomal *cis* pQTLs, which have a direct regulatory effect on the protein expression and proven of greater relevance for downstream translation⁵, were identified in the two proteogenomic studies^{8,9}. This approach aimed to limit the potential for horizontal pleiotropy that may occur with the use of *trans* pQTLs. The *cis* region was defined as a distance of 1 Mb upstream or downstream from the end or start, respectively, of the gene encoding the protein of interest in both platforms.

cis pQTLs were selected based on the statistical criteria applied in the contributing proteogenomic studies ($p < 1.70 \times 10^{-11}$ for the study using the Olink platform, and $p < 1.80 \times 10^{-9}$ for the study using the SomaScan platform)^{8,9}. The thresholds for variant selection differ to reflect those of the original studies. Specifically, associations in the Olink study were defined as statistically significant based on adjusting the genome-wide significant threshold for the number of proteins tested⁸, whilst associations in the SomaScan study were defined as statistically significant based on adjusting the nominal significance threshold for the number of genetic variants tested per protein⁹. Ambiguous palindromic single nucleotide polymorphisms with an allele frequency between 0.42 and 0.58 (in either the proteogenomic or ischemic stroke GWAS) were excluded¹¹.

Due to the complex linkage disequilibrium (LD) structure of genetic variants within the human major histocompatibility complex (*MHC*) region and their known highly pleiotropic effects, genetic variants and proteins

encoded by genes within the *MHC* region (chr6: from 26 Mb to 36 Mb) were excluded³. To reduce the risk of weak instrument bias, genetic instruments with an *F*-statistic of at least 10 were included¹². After restricting genetic variants to those available in GIGASTROKE, clumping based on LD from 1000 Genomes Project European ancestry population using a genetic window of ± 10 Mb and an LD $r^2 < 0.0001$ was undertaken to identify independent *cis* pQTLs for plasma protein abundance. To ensure that the genetic instruments reflect causal effects from the protein to the ischemic stroke phenotypes, we applied the Steiger filtering, which tests whether each genetic instrument explains more variance in the exposure than in the ischemic stroke phenotype¹³. PLINK version 1.90 and the *ieugwasr* R package were used to perform the clumping¹⁴. We followed the recommended harmonization framework for two-sample MR analyses^{11,15}, using the *TwoSampleMR* R package¹³. The selection process for genetic instruments is summarized in Supplementary Fig. S1.

If a protein was targeted by multiple assays within the same proteomic platform, only the assay that had the largest explained variability by the independent instrumental variables was considered to maximize statistical power for MR analysis. There is substantial heterogeneity between the complementary protein information provided by the Olink and SomaScan platforms, given their technological approaches, which can result in different proteoforms of the same protein that are not directly comparable¹⁶. Consequently, when a protein was measured in both the Olink and SomaScan platforms, we selected genetic instruments separately from each platform and performed MR analyses independently.

Statistical analysis

Association of plasma proteins with stroke and stroke subtypes. The Wald ratio, defined as the ratio of the gene-outcome effect divided by the gene-exposure effect, was calculated for each of the genetic variants selected¹⁷. When more than one variant was available for a protein, the Wald ratio estimates were combined using a fixed-effect inverse variance weighted model, based on the assumption that *cis* variants act through the same molecular pathway. Statistically significant associations were declared based on a 5% false discovery rate (FDR) threshold (after applying the Benjamini-Hochberg correction)¹⁸. MR analyses were performed using the *TwoSampleMR* R package¹³.

For proteins quantified on both platforms, we assessed whether evidence of causal effects was consistent across platforms. However, genetic instruments and resulting causal effect estimates from Olink and SomaScan platforms are not directly comparable, given differences in the underlying protein information captured by these platforms and the use of normalized expression (NPX units) and protein abundance (standard deviation units), respectively.

Bi-directional (reverse) Mendelian randomization. The potential effect of genetic liability to ischemic stroke on plasma proteins was considered using a bi-directional (reverse) MR approach. To obtain independent genetic variants associated with any ischemic stroke and ischemic stroke subtypes, we performed LD clumping as previously described above, with a statistical significance threshold of $p < 5 \times 10^{-8}$. We then examined whether genetically predicted risk of any of the ischemic stroke phenotypes was related to the proteins considered. The Wald ratio was estimated for each genetic variant and combined using a random-effects model¹⁹, to allow for heterogeneity that may result from different genetic variants for ischemic stroke and its subtypes acting through different molecular pathways. Statistically significant associations were determined using a 5% FDR.

Bayesian colocalization. Evidence of colocalization is used to support the validity of the genetic variants and strengthen MR findings by reducing the risk of confounding due to LD²⁰. In the context of MR, colocalization provides evidence that the genetic variant used as an instrument influences the outcome through the exposure of interest. We examined whether the prioritized proteins share the genetic variant with

the outcomes of interest by conducting a colocalization analysis, under the assumption of a single causal variant in each genetic locus. The colocalization analysis was performed on a genetic window of ± 0.5 Mb from the lead *cis* pQTL of each protein. We implemented this analysis using the *coloc* R package with its default prior probabilities²¹.

A posterior probability greater than 80% was considered as strong evidence for colocalization, and a posterior probability between 60 and 80% was considered moderate evidence. Proteins with a statistically significant association with any ischemic stroke or its subtypes at 5% FDR, which were also supported by strong evidence for colocalization, were prioritized for subsequent analyses.

Pleiotropy-robust *cis* Mendelian randomization. As a sensitivity analysis, we applied the MR-link-2 method to proteins that showed evidence of a causal effect in the primary analysis^{22,23}. MR-link-2 is a pleiotropy-robust *cis* MR method that enables causal inference in the presence of LD and unobserved pleiotropic effects, without requiring the removal of pleiotropic genetic instruments or explicit measurement of all sources of pleiotropy. Using this approach, we evaluated the robustness of our findings to potential horizontal pleiotropy (i.e., violation of the exclusion restriction assumption). MR-link-2 tests two parameters: the causal effect of a protein on the outcome conditional on pleiotropic effects, and a pleiotropy parameter that captures the heritable component of the outcome within the *cis* protein locus that is independent of the protein. For these analyses, we used summary statistics for each protein and stroke outcome across the corresponding *cis* region, together with the LD reference panel from the 1000 Genomes Project European ancestry population.

Comparison with prior evidence on proteins related to the risk of ischemic stroke. We examined whether (1) any genetic variant within the genes encoding the prioritized proteins was associated with ischemic stroke or its subtypes in the GWAS Catalog²⁴, or (2) the prioritized proteins were reported as statistically significant for any ischemic stroke or its subtypes in a proteome-wide MR study. To compile a comprehensive list of relevant MR studies, a systematic review was conducted in PubMed up to 25th October 2025 using the following search algorithm: (“stroke” OR “cerebrovascular disease” OR “small vessel disease”) AND (“Mendelian randomization” OR “Mendelian randomisation”). MR studies were deemed eligible if they assessed the association of at least 100 proteins using plasma *cis* pQTLs measured through a high-throughput platform.

We did not consider (1) studies examining any stroke as an outcome, because this phenotype conflates ischemic and hemorrhagic stroke, which are two very heterogeneous clinical entities, (2) studies examining plasma proteins with a causal effect on outcomes related to stroke progression, (3) studies that did not use colocalization to validate their findings, and (4) studies that did not use genome-wide association statistics from large genetic consortia. From each eligible study, we extracted information on the GWAS datasets used to derive *cis* pQTLs and their effect on ischemic stroke phenotypes, along with the proteins identified.

Association of plasma proteins with ischemic stroke risk factors. We focused on five key risk factors for ischemic stroke (ever smoking²⁵, systolic blood pressure²⁶, type 2 diabetes mellitus²⁷, body mass index²⁸, and atrial fibrillation²⁹), in line with a previously published study, which suggested a causal role for these factors in the development of ischemic stroke⁶. For each risk factor, we used publicly available GWAS data from European populations to identify the genetic effect of each of the *cis* pQTLs on the ischemic stroke risk factors (Supplementary Data 1), applying the same rigorous steps described above. Statistically significant associations were based on a 5% FDR, after applying the Benjamini-Hochberg correction across the associations between all the proteins with ischemic stroke risk factors.

Phenome-wide associations. To consider whether protein abundance could have wider implications than on ischemic stroke, we examined associations of the genetic instruments derived for the prioritized proteins across the phenome. Specifically, the effects of the genetic instruments derived for the prioritized proteins were estimated on 57 blood biomarkers in the UK Biobank and 2462 disease outcomes in FinnGen (release 5, accessed through OpenGWAS), using summary-level genetic data from publicly available GWAS³⁰. Statistically significant associations were based on 5% FDR threshold.

Biological data integration

Enrichment analysis. For the proteins associated with ischemic stroke or its subtypes and supported by colocalization, we performed an enrichment analysis in Gene Ontology³¹, Reactome Pathway³², and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases³³, using the *clusterProfiler* R package³⁴. Statistically significant enrichment was based on 5% FDR. All proteins tested for an association with ischemic stroke or its subtypes were included in the analysis as the background set^{35,36}.

Protein-protein interactions and Mendelian diseases. We used the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) to examine potential protein-protein interactions among proteins with evidence of colocalization with ischemic stroke³⁷. We also searched the Online Mendelian Inheritance in Man (OMIM) to assess whether the identified proteins have been linked to any genetic disorders³⁸.

Assessment of potential pleiotropic effects of *cis* pQTLs. We assessed potential pleiotropic effects of the selected *cis* pQTLs for the statistically significant proteins by examining whether these genetic variants have a genome-wide significant effect on other plasma proteins ($p < 5 \times 10^{-8}$), by testing the genetic effects of these *cis* pQTLs on all the Olink and SomaScan proteins measured in the UK Biobank and deCODE study, respectively. We then examined whether these associations between *cis* pQTLs and proteins may be explained by mediation through shared biological pathways, which is an indication of vertical pleiotropy, or unrelated mechanisms, which indicates the presence of horizontal pleiotropy. A pleiotropic effect was attributed to mediation if the *cis*- and *trans*-associated proteins were involved in the same biological pathway based on Gene Ontology³⁹. Otherwise, we consider them as a potential result of unrelated mechanisms.

Feasibility of drug targeting. We searched Open Targets^{40,41}, Therapeutic Target Database⁴², Drug Bank⁴³, and PharmGKB⁴⁴ to examine whether the prioritized proteins have been previously considered as drug targets in experimental studies in humans or animals. Data on the tractability and do-ability of targeting these proteins were also extracted from Open Targets⁴⁵. A particular protein was considered tractable if (1) it is a membrane or a secreted protein, or (2) it binds a specific ligand or a small molecule, or (3) it has predicted pockets. Targeting a particular protein was considered do-able if (1) at least one mouse gene contains a sequence with a 100% of identity to the target protein, or (2) a high-quality chemical probe is available for targeting this protein. Information about knockout animal models focusing on proteins that are associated with the risk of ischemic stroke was extracted from Open Targets, which curates information based on the Mouse Genome Informatics database⁴⁶.

Gene expression. The Human Protein Atlas was used to assess the tissue, brain, and single-cell specificity of the identified proteins^{47,48}. We specifically examined if the prioritized proteins show high specificity in tissues or single cells that are relevant to ischemic stroke and its well-established risk factors (i.e., kidney, liver, pancreas, adipose tissue, cardiovascular system, and brain).

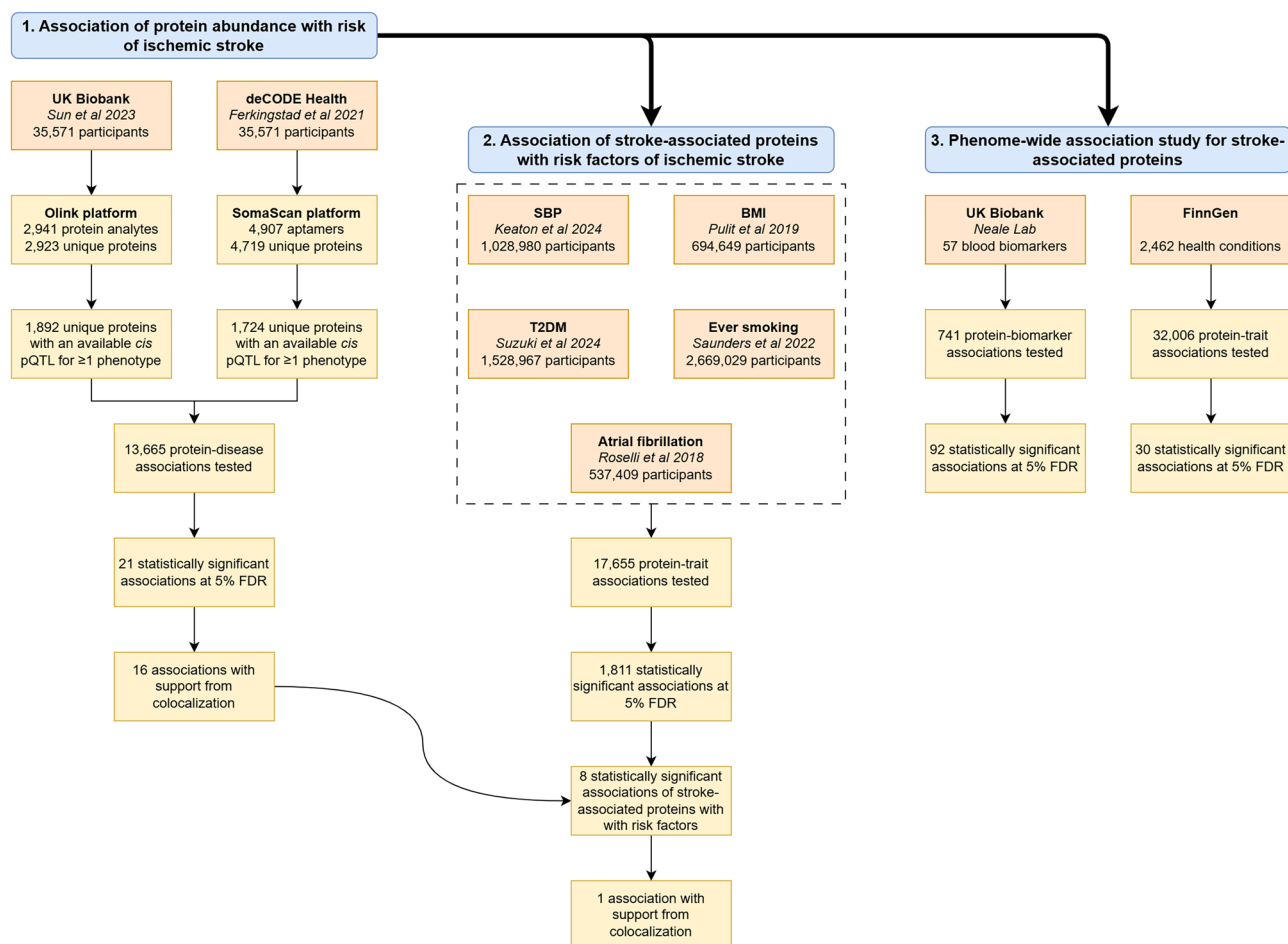


Fig. 1 | Flow chart of the study design and main results. In the first step, *cis* pQTLs from the UK Biobank (Olink platform) and the deCODE Health Study (SomaScan platform) were used to identify plasma proteins related to ischemic stroke and its subtypes. Through a two-sample Mendelian randomization approach, we identified 21 statistically significant protein-disease associations, and 16 of these protein-disease pairs also showed evidence of colocalization. In the second step, we identified

8 out of 16 statistically significant associations between plasma proteins and ischemic stroke risk factors, and one protein-risk factor pair also showed evidence of colocalization. In the final step, a phenome-wide Mendelian randomization analysis revealed 92 associations with medical conditions in FinnGen and 30 associations with blood biomarkers in the UK Biobank.

Ethics approval

In this study, we solely used publicly available summary-level GWAS data. Details on how to access these data sets are provided in the “Data availability” statement. For all included GWAS, all participants provided informed consent, and study protocols were approved by their respective ethical committees. As our analyses did not involve individual-level participant data, no additional ethical approval was required for the present study.

Results

Association of genetically predicted proteins with ischemic stroke

To systematically evaluate the evidence for a potentially causal association between plasma proteins and risk of ischemic stroke and ischemic stroke subtypes, we performed a proteome-wide two-sample MR analysis (Fig. 1). We used independent *cis* pQTLs from two complementary proteomic platforms (Olink in the UK Biobank and SomaScan in the deCODE Health Study) as genetic proxies for plasma protein abundance. In total, we tested 13,665 associations (53% using pQTLs from the Olink platform and 47% using pQTLs from the SomaScan platform). We observed 1054 (7.7%) nominally significant associations ($p < 0.05$; Supplementary Data 2), and 21 associations (0.2%) were statistically significant at a 5% false discovery rate (FDR), corresponding to $p < 6.76 \times 10^{-5}$ (Figs. 2 and 3 and Table 1). Steiger filtering did not exclude any genetic instruments and validated that the

genetic instruments for all the tested associations reflect causal effects from the protein to ischemic stroke phenotypes¹³. None of the genetic variants used in these 21 associations was protein-altering, indicating the robustness of our analysis against epitope binding effects⁴⁹.

Higher genetically-predicted levels of five proteins (F11, FURIN, KNG1, LRP4, SH2B3) were associated with higher risk of any ischemic stroke, with odds ratio ranging from 1.08 (95% CI, 1.04–1.12) per 1-SD higher LRP4 abundance to 1.77 (95% CI, 1.44–2.18) per 1-NPX higher SH2B3 abundance. Also, higher levels of six proteins (GRK5, MMP12, MMUT, CEP85, PROCN, SH3BGRL3) were associated with lower risk of any ischemic stroke with odds ratio ranging from 0.65 (95% CI, 0.54–0.79) to 0.92 (95% CI, 0.90–0.95) per 1-NPX higher CEP85 and MMP12 abundance, respectively. Furthermore, the effects of these proteins on risk of ischemic stroke were directionally consistent in the GIGASTROKE multi-ancestry population (~30% non-Europeans; Supplementary Data 3), based on the same *cis* pQTLs.

Stroke subtype-specific associations were also considered, showing higher genetically-predicted MMP12 abundance associated with lower risk of large artery stroke (OR per 1-SD higher MMP12, 0.78; 95% CI, 0.71–0.86), and higher F11 abundance with risk of cardioembolic stroke (OR per 1-SD higher, 1.30; 95% CI, 1.19–1.43). Notably, higher MSMB abundance (OR per 1-SD higher, 1.22; 95% CI, 1.11–1.34) and higher SPATA20 abundance (OR per 1-SD higher, 1.35; 95% CI, 1.19–1.54) were associated exclusively with small vessel stroke.

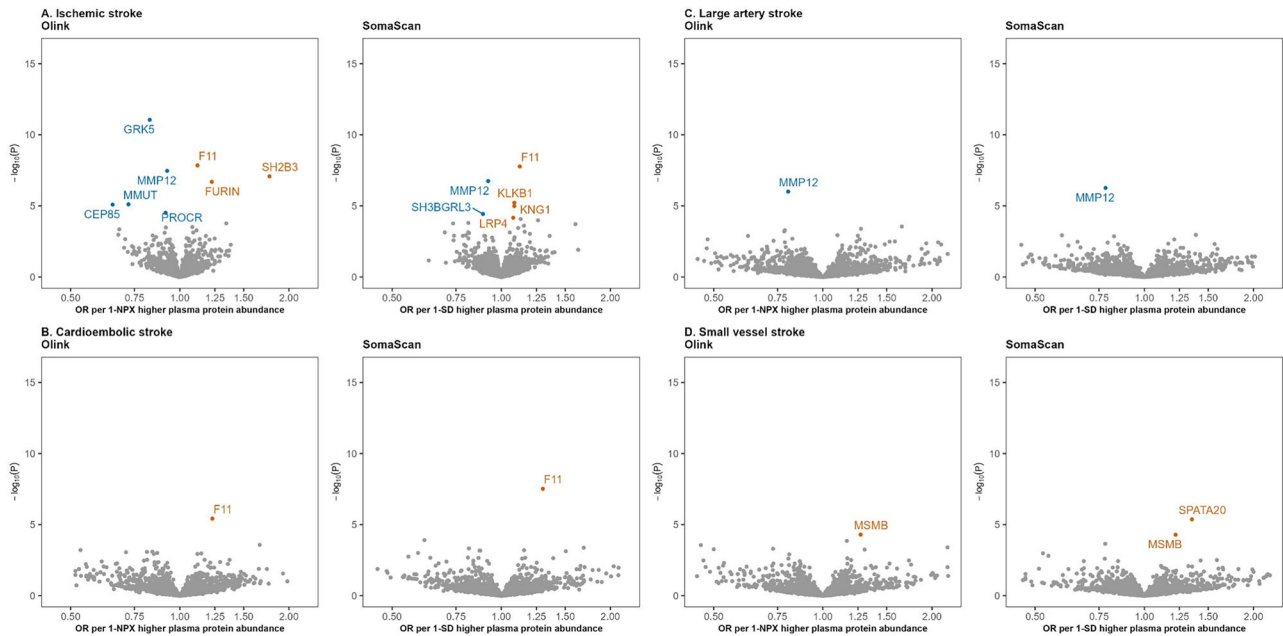


Fig. 2 | Association of plasma proteins with risk of ischemic stroke and its subtypes using the *cis* two-sample Mendelian randomization approach. Proteins that showed a statistically significant association with ischemic stroke phenotypes, based on 5% FDR, are annotated. The labeled proteins in blue indicate a negative

association with stroke risk, while those in orange indicate a positive association. The gray dots represent proteins that did not reach statistical significance. Results are shown separately for each ischemic stroke phenotype: **A** any ischemic stroke, **B** cardioembolic stroke, **C** large artery stroke, and **D** small vessel stroke.

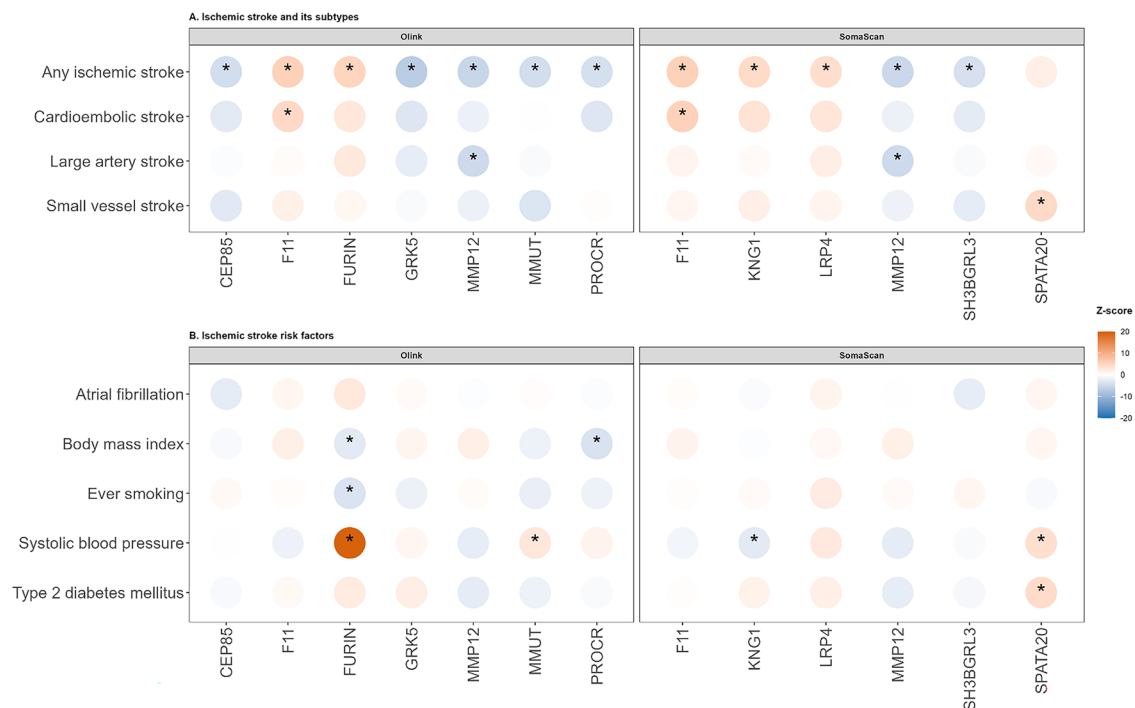


Fig. 3 | Association of ischemic stroke-associated proteins with ischemic stroke risk factors using two-sample Mendelian randomization analysis and *cis* pQTLs as instrumental variables. Only plasma proteins showing an association with ischemic stroke and/or its subtypes, with support from the two-sample *cis* MR analysis and colocalization, are shown in this figure. **A** illustrates associations with ischemic stroke and its subtypes (including cardioembolic, large artery, and small

vessel stroke). **B** shows associations with ischemic stroke risk factors (atrial fibrillation, body mass index, smoking status, systolic blood pressure, and type 2 diabetes). Positive associations (orange) and negative associations (blue) are represented by the color intensity, scaled by the Z-score. Asterisks indicate significant associations at 5% FDR.

Colocalization of plasma proteins with ischemic stroke

Assuming a single causal genetic variant at each genetic locus, we found that 16 out of 21 statistically significant associations (76%) had strong evidence for colocalization (posterior probability for $H_4 > 0.80$; Supplementary

Data 4). One additional protein-disease pair (MMP12 measured through Olink platform with large artery stroke) showed moderate evidence for colocalization with large artery stroke (posterior probability for $H_4 = 0.77$). Two associations (SH2B3 and KLKB1 with any ischemic stroke) suggested a

Table 1 | Plasma proteins with a statistically significant effect on ischemic stroke phenotypes at 5% FDR using two-sample Mendelian randomization

Platform	Assay ID	Protein	N variants ^a	OR ^b	95% CI	p ^c	Novel ^d	Colocalization ^e
Ischemic stroke								
Olink	OID20066	GRK5	1	0.83	0.78–0.87	8.82E–12	No	0.99
	OID30773	F11	2	1.12	1.08–1.16	1.41E–08	No	0.99
	OID21439	MMP12	3	0.92	0.90–0.95	3.43E–08	No	0.90
	OID21222	SH2B3	1	1.77	1.44–2.18	8.30E–08	NA	0.59
	OID21514	FURIN	2	1.23	1.13–1.32	2.02E–07	No	0.86
	OID30159	MMUT	1	0.72	0.63–0.83	7.71E–06	Yes	0.81
	OID21241	CEP85	1	0.65	0.54–0.79	8.03E–06	Yes	0.86
	OID30149	PROCR	1	0.91	0.88–0.95	3.01E–05	No	0.90
SomaScan	2190_55	F11	3	1.12	1.08–1.17	1.68E–08	No	0.99
	4496_60	MMP12	3	0.92	0.89–0.95	1.79E–07	No	0.84
	4152_58	KLKB1	3	1.09	1.05–1.13	5.92E–06	NA	0.14
	15343_337	KNG1	3	1.09	1.05–1.13	1.03E–05	Yes	0.99
	17490_4	SH3BGRL3	2	0.89	0.84–0.94	3.68E–05	No	0.84
	19558_10	LRP4	2	1.08	1.04–1.12	6.67E–05	No	0.86
Cardioembolic stroke								
Olink	OID30773	F11	2	1.23	1.13–1.34	3.76E–06	No	0.99
SomaScan	2190_55	F11	3	1.30	1.19–1.43	3.01E–08	No	0.99
Large artery stroke								
Olink	OID21439	MMP12	3	0.80	0.73–0.88	9.78E–07	No	0.77
SomaScan	4496_60	MMP12	3	0.78	0.71–0.86	5.41E–07	No	0.91
Small vessel stroke								
Olink	OID20275	MSMB	2	1.27	1.13–1.43	5.01E–05	NA	0.52
SomaScan	11117_2	SPATA20	2	1.35	1.19–1.54	4.22E–06	Yes	0.92
	10620_21	MSMB	2	1.22	1.11–1.34	5.07E–05	NA	0.46

CI confidence interval, OR odds ratio, NA not assessed.

^aFor proteins with a single genetic variant as instrumental variable, odds ratio was estimated using the Wald ratio. For proteins with at least 2 genetic variants as instrumental variables, odds ratio was estimated using the fixed-effect inverse variance-weighted model.

^bThe odds ratio represents the risk of stroke by 1 standard deviation increase in plasma protein abundance for *SomaScan* platform, and by 1 NPX unit increase in plasma protein abundance for *Olink* platform.

^cp values are two-sided and are based on Z tests. Multiple testing was addressed using the Benjamini-Hochberg false discovery rate procedure, although the table presents the nominal (unadjusted) p values.

^dNovelty was defined based on the findings of the latest genome-wide association study and previously published Mendelian randomization studies, and it was assessed only when posterior probability for colocalization exceeded 80%.

^eColocalization was performed under the assumption of a single causal variant per genetic loci. This column reports the posterior probability that each protein shares the same causal variant with an ischemic stroke phenotype at the respective genetic loci.

distinct causal variant (posterior probability for $H_3 > 0.80$). In summary, the colocalization analysis provides an additional level of support for these associations, strengthening the validity of the *cis* pQTLs as genetic instruments for these proteins³⁰.

Pleiotropy-robust *cis* Mendelian randomization

Results were broadly consistent in the sensitivity analysis using MR-link-2, a pleiotropy-robust *cis* MR method²². Of the 21 associations detected in our primary analysis, 19 (90%) remained statistically significant after adjusting for pleiotropy ($p < 0.05$; Supplementary Data 5). For SH2B3, the pleiotropy-adjusted effect on risk of ischemic stroke was not statistically significant, with the MR-link-2 pleiotropy test indicating substantial horizontal pleiotropy. This is consistent with the lack of colocalization between SH2B3 and ischemic stroke at the *cis* locus, as we previously reported. For MMUT, the pleiotropy-adjusted effect estimate was directionally consistent with the primary analysis but was not statistically significant. This finding may reflect a combination of limited statistical power due to low heritability for MMUT at the *cis* region and presence of horizontal pleiotropy.

Cross-platform comparison

We compared the plasma proteins we identified as potentially causal for ischemic stroke and its subtypes across the two different proteomic platforms (Supplementary Fig. S2 and Supplementary Data 6). Five associations (F11 with ischemic stroke and cardioembolic stroke, MMP12 with ischemic stroke and large artery stroke, and MSMB with small vessel disease) were statistically significant in both platforms with concordant direction of effect, whereas one association (KLKB1 with any ischemic stroke) was statistically significant only when we considered *cis* pQTLs from the SomaScan platform. The remaining associations could be evaluated only in one platform due to the limited overlap in the proteins measured across Olink and SomaScan platforms. A directionally consistent effect across the two platforms indicates that the inferred effect on disease risk is independent of the platform used for protein abundance measurement and pQTL identification.

Effect of genetic liability to ischemic stroke on plasma proteins

To assess the relationship of genetic liability to ischemic stroke or its subtypes with plasma proteins, we conducted a reverse MR analysis using

genetic proxies for risk of ischemic stroke or its subtypes. Associations between genetic liability to small vessel stroke and the plasma proteome could not be assessed due to absence of genome-wide significant genetic variants for European populations in the GWAS by the GIGASTROKE Consortium¹⁰.

Among 11,238 associations tested (Supplementary Data 7), 827 were statistically significant at $p < 0.05$, and 325 were statistically significant at 5% FDR. However, none of these proteins overlapped with those found to have statistically significant effects on risk of ischemic stroke and its subtypes. This observation supports that the identified plasma proteins are causally relevant to, rather than a consequence of, ischemic stroke.

Comparison of our findings with existing studies

Our systematic review identified eight proteome-wide MR studies (six in European populations, one in East Asian populations, and one in South Asian populations) examining the association of proteins with risk of ischemic stroke and/or its subtypes (Table 2)^{6,50–56}. Seven of the proteins we identified (GRK5, PROC, F11, MMP12, SH3BGL3, LRP4, and FURIN) had been previously observed in MR studies and effects were directionally consistent, thus our results support these previous findings. There was also genetic support from the GIGASTROKE Consortium, which identified genetic variants in five of the 11 genes (*GRK5*, *F11*, *PROC*, *FURIN*, and *MMP12*) encoding proteins that are associated with ischemic stroke in populations of European ancestry¹⁰. To the best of our knowledge, our study supports MR and GWAS findings for seven proteins, and identifies four proteins that have not previously been implicated in risk of ischemic stroke: *MMUT*, *CEP85*, and *KNG1* associated with ischemic stroke, and *SPATA20* with small vessel stroke. Also, by examining the data sources for ischemic stroke and pQTL selection used in the previous studies, our study offers a unique combination of two large GWAS on plasma proteins using two different proteomic platforms with the largest GWAS for ischemic stroke in populations of European ancestry.

Pleiotropic effects of cis pQTLs

Genetic variants used as instruments for a plasma protein may exhibit effects on other proteins, which may be attributed to either shared biological pathways (*vertical pleiotropy*) or unrelated mechanisms (*horizontal pleiotropy*). These effects were explored using predicted protein–protein interactions in STRING and biological pathway annotations in Gene Ontology (Supplementary Figs. S3 and S4 and Supplementary Data 8 and 9).

As might be anticipated, many pleiotropic effects were consistent with mediation through shared biological pathways. The *cis* pQTLs for MMP12 were associated with at least one of the matrix metalloproteinase family members (MMP1, MMP3, MMP7, MMP8), all involved in collagen degradation (Fig. 4A). The *cis* pQTL of GRK5 (rs10886430) was associated with CCL17, CCL22, and CCL28, all involved in the G protein-coupled receptors signaling (Fig. 4B). Similarly, the *cis* pQTL for FURIN (rs2071410) was associated with FER and FES, which are involved in the receptor tyrosine kinases signaling. Additionally, the *cis* pQTL for PROC (rs146102159) was associated with F7 and PROC, all involved in the formation of fibrin clots, indicating possible mediation (*vertical pleiotropy*, Fig. 4C). However, its association with CPNE1 could not be explained by shared biological pathways, suggesting possible confounding by unrelated mechanisms.

More complex scenarios were observed in the case of *KNG1* and *F11*, as their *cis* pQTLs were associated with multiple proteins. Shared biological mechanisms, related to blood coagulation, proteolysis, immune system, and response to stress, explain the majority of these pleiotropic effects (Fig. 4D, E). Interestingly, the *cis* pQTLs for *KNG1* were associated with *F11*, and vice versa, suggesting a biological interaction between these proteins. Moreover, the *cis* pQTLs for *SH3BGL3*, *CEP85*, *LRP4*, *MMUT* and *SPATA20* showed minimal pleiotropic effects being associated with one or two additional proteins. Although no shared biological pathways could be identified, this may reflect limitations in the current biological knowledge as captured by the existing databases rather than horizontal pleiotropy.

Table 2 | Studies examining the effect of plasma proteins on ischemic stroke and its subtypes using protein quantitative trait loci for at least 100 proteins within a Mendelian randomization and colocalization framework

First author (Year)	Ancestry	Stroke GWAS	Source of pQTLs	Proteins identified
Chen (2022) ⁶	Europeans	MEGASTROKE ⁹⁷	INTERVAL (Olink) ⁹⁸	Ischemic stroke: CD40, TFPI, MMP12, IL6RA, TMPRSS5; Large artery stroke: CD40, MMP12; Cardioembolic stroke: TMPRSS5
Wu (2022) ⁵⁰	Europeans	MEGASTROKE ⁹⁷	AGES Reykjavik study (SomaScan) ⁹⁹	Ischemic stroke: ALDH2, PTPN11, VPS36, SLC44A2 Cardioembolic stroke: L3HYPDH; Small vessel stroke: ICA1L
Wu (2024) ⁵¹	South Asians	GIGASTROKE ¹⁰	UK Biobank (Olink) ⁸	Cardioembolic stroke: GP6
Yao (2025) ⁵²	Europeans, East Asians	GIGASTROKE ¹⁰	UK Biobank (Olink) ⁸ , China Kadoorie Biobank (Olink)	Ischemic stroke: FGF5, TMPRSS5, FURIN, F11, ALDH2, ABO, GRK5, KIAA0319; PROC, MMP12
Zhang (2025) ⁵³	Europeans	MEGASTROKE ⁹⁷	INTERVAL (Olink) ⁹⁸	Ischemic stroke: MMP12
Zhao (2025) ⁵⁴	Europeans	MEGASTROKE ⁹⁷	UK Biobank (Olink) ⁸ , deCODE (SomaScan) ⁹	Ischemic stroke: FURIN; Cardioembolic stroke: F11; Small vessel stroke: VSIR, DDHD2
Zhu (2024) ⁵⁵	Europeans	GIGASTROKE ¹⁰	deCODE (SomaScan) ⁹ , ARIC (SomaScan) ¹⁰⁰	Ischemic stroke: F11, LRP4, MMP12, PTGRI1, SH3BGL3; Large artery stroke: MMP12; Cardioembolic stroke: F11, EMILIN3, NUDT9
Zou (2024) ⁵⁶	Europeans	MEGASTROKE ⁹⁷ , FinnGen, UK Biobank	KORA F4/QMDiab (SomaScan) ¹⁰¹ , INTERVAL (Olink) ⁹⁸ , Framingham Heart study (Luminex) ¹⁰² , IMPROVE (Olink) ¹⁰³ , deCODE (SomaScan) ⁹	Ischemic stroke: SWAP70, MMP12; KLKB1

The table summarizes proteome-wide Mendelian randomization studies evaluating the associations of genetically predicted plasma proteins with risk of ischemic stroke or its subtypes. For each study, we present the genetic ancestry of the study population, the genetic data sources for ischemic stroke and for identification of protein quantitative trait loci (pQTLs), and the proteins prioritized. Studies were identified through a systematic review, as described in the “Methods”.

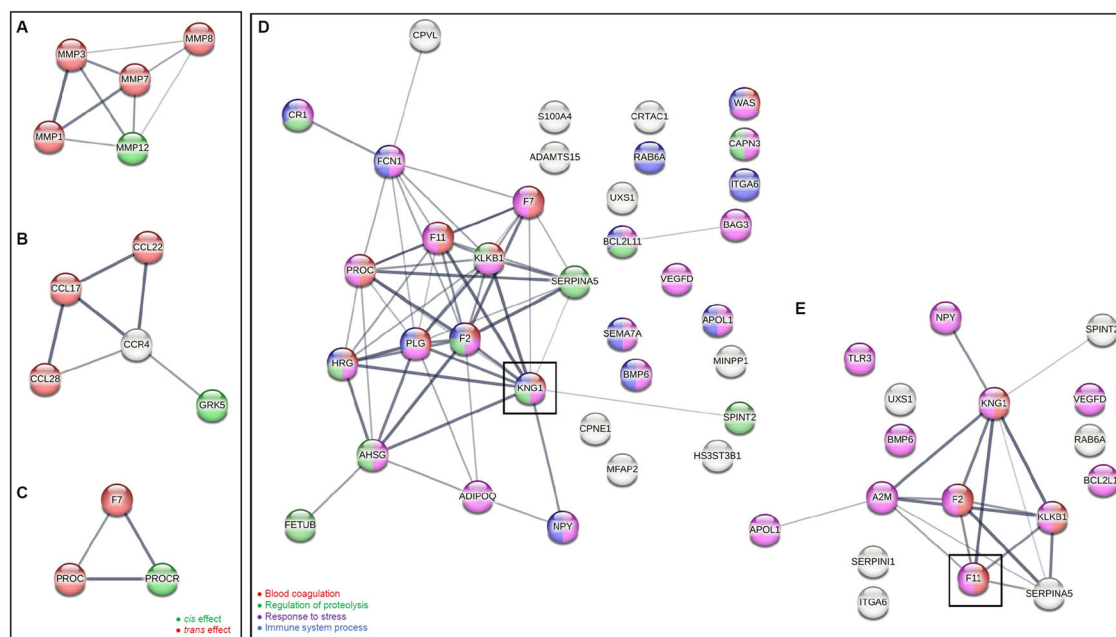


Fig. 4 | Predicted protein–protein interaction networks for pleiotropic effects of the genetic instruments. Protein–protein interaction networks were generated using the STRING database (version 12). Each node represents a protein, and edges indicate predicted interactions based on evidence from curated databases, experimental data, co-expression, and text mining. Edge thickness reflects the strength of the supporting evidence. In **A**, the *cis* pQTL for MMP12 had an effect on MMP1, MMP3, MMP7 and MMP8, which form an interconnected network in STRING and are collectively involved in collagen degradation according to Gene Ontology (GO). In **B**, the *cis* pQTL for GRK5 was associated with CCL28, CCL17 and CCL22, which participate in the G protein-coupled receptors signaling based on GO. When an additional protein (CCR4) was introduced, all these proteins form a cohesive

interaction network. In **C**, the *cis* pQTL had an effect on PROC and F7, which form an interaction network based on STRING and participate in the formation of fibrin clot based on GO. In **D**, a complex protein–protein interaction network was observed for the proteins that are affected by the *cis* pQTLs for KNG1, with most of the proteins relating to four particular biological processes: blood coagulation, regulation of proteolysis, response to stress and immune system process. In **E**, the complex protein–protein interaction network for the proteins that are affected by the *cis* pQTLs for F11 is shown. A majority of these effects can be explained by two biological processes in Gene Ontology, which are blood coagulation, and response to stress.

Effect of plasma proteins on ischemic stroke risk factors

To assess whether genetically-predicted ischemic stroke-associated proteins have an effect on established risk factors of ischemic stroke⁶, we performed a proteome-wide two-sample MR analysis, using ischemic stroke risk factors as the outcome. Among 17,655 tested associations between plasma proteins and risk factors, 1811 (10.2%) were statistically significant at 5% FDR (Supplementary Data 10). Only eight associations involved ischemic stroke-associated proteins (Fig. 3).

Genetically-predicted FURIN was associated, with systolic blood pressure (β per 1-NPX unit higher, 2.71 mmHg; 95% CI, 2.44–2.99), ever smoking (OR per 1-NPX unit higher, 0.96; 95% CI, 0.95–0.98) and BMI (β per 1-NPX unit higher, -0.02 SD; 95% CI, -0.04 to -0.01). Genetically-predicted MMUT was positively associated with systolic blood pressure (β per 1-NPX unit higher, 0.71 mmHg; 95% CI, 0.23–1.19). Genetically-predicted SPATA20 was associated with type 2 diabetes mellitus (OR per 1-SD increase, 1.05; 95% CI, 1.03–1.08) and systolic blood pressure (β per 1-SD higher, 0.30 mmHg; 95% CI, 0.15–0.45). Genetically-predicted KNG1 was associated with SBP (β per 1-SD higher, -0.18 mmHg; 95% CI, -0.31 to -0.05). Finally, genetically-predicted levels of PROC2 were associated with body mass index (β per 1-NPX unit higher, -0.02 SD; 95% CI, -0.04 to -0.01). Only the association of FURIN with systolic blood pressure showed strong evidence of colocalization (posterior probability = 0.98; Supplementary Data 11).

Assessment of phenome-wide effects

To assess whether changes in the abundance of the ischemic stroke-associated proteins have wider implications by affecting the risk of other medical conditions or the levels of blood biomarkers, we examined the phenome-wide effects of these proteins through a two-sample phenome-wide MR analysis.

We tested the potentially causal association of plasma proteins with multiple diseases in the population of FinnGen. We identified 30 statistically significant protein-disease associations at 5% FDR, involving GRK5, FURIN, MMP12, F11, and PROC2 (Supplementary Data 12 and Supplementary Fig. S5). Higher genetically-predicted abundance of F11 and PROC2 were associated with higher risks of venous thromboembolism and its subtypes (pulmonary embolism or deep venous thrombosis), despite PROC2 being associated with lower risk of ischemic stroke. In addition, higher genetically-predicted abundance of FURIN was associated with higher risks of hypertensive disorders, GRK5 with higher risk of ischemic stroke and its subtypes, and MMP12 with higher risk of allergic asthma.

We further tested the association of these proteins with multiple blood biomarkers in the population of UK Biobank. We found that 92 protein-biomarker associations were statistically significant at 5% FDR, involving all but two proteins (MMP12 and PROC2) (Supplementary Data 13 and Supplementary Fig. S6). We observed associations between ischemic stroke-associated proteins and circulating lipids (CEP85, FURIN, LRP4 and SPATA20), platelet markers (CEP85, FURIN, GRK5, F11, LRP4 and SH3BGRL3), markers of liver (FURIN, LRP4, MMUT and SPATA20) and kidney function (SH3BGRL3, FURIN and SPATA20), and glycemic traits (LRP4 and MMUT). These findings suggest potential biological pathways through which these proteins may influence the risk of ischemic stroke, whereas they may also indicate potential safety considerations if these proteins were pharmacologically targeted.

Enrichment analysis

Enrichment analysis for the 11 ischemic stroke-associated proteins revealed statistically significant enrichment for complement and coagulation cascades in KEGG pathways, and for formation of fibrin clots in Reactome Pathways, at 5% FDR. Moreover, an interaction of KNG1 with both F11 and

Table 3 | Cumulative evidence on the potential drug targets for ischemic stroke

Protein	Direction of effect ^a	Available drug ^b	Tractability ^c	Doability ^d	Phenome-wide effects ^e	Mouse models	Tissue specificity ^f	Single-cell specificity ^f	Previously reported ^g
CEP85	↓	No	○	○	Platelets	○	○	●	●
F11	↑	Inhibitors	●	○	Lymphocytes, platelets	○	●	●	○
FURIN	↑	No	●	●	Calcium, Albumin, WBC, RBC, LDL cholesterol, cystatin C, HDL cholesterol	●	●	●	○
GRK5	↓	No	●	○	Platelets, Lymphocytes	○	○	●	○
KNG1	↑	Inhibitors	●	○	Not detected	○	●	●	●
LRP4	↑	No	○	○	Calcium, Albumin, triglycerides, platelets	●	●	●	○
MMP12	↓	Inhibitors	●	●	Not detected	○	○	○	○
MMUT	↓	No	●	●	Not detected	○	●	●	●
PROCR	↓	No	●	●	Venous thromboembolism	○	●	○	○
SH3GBRL3	↓	No	●	●	Neutrophils	○	○	○	○
SPATA20	↑	No	●	●	Creatinine	●	○	○	●

^aThe direction of the effect on risk of ischemic stroke for increasing levels of protein abundance based on our two-sample Mendelian randomization.

^bBased on Open Targets, Drug Bank, Therapeutic Targets Database and PharmGKB.

^cA particular protein was considered as tractable if (a) it is a membrane or a secreted protein, or (b) it binds a specific ligand or a small molecule, or (c) has predicted pockets.

^dTargeting a particular protein was considered as do-able if (a) there is at least one gene in mice that contains a sequence with a 100% of identity with the target protein, or (b) there is a high-quality chemical probe targeting this protein.

^eThe information is based on our Mendelian randomization phenome-wide association study in the UK Biobank (for blood-based biomarkers) and FinnGen (for disease phenotypes).

^fThe information is based on Human Protein Atlas and reflects the gene expression specificity in biologically meaningful tissues and single cells for ischemic stroke.

^gThe assessment is based on GWAS Catalog and the published large-scale proteome-wide Mendelian randomization studies for ischemic stroke.

GRK5 was observed based in the STRING database. The interaction between F11 and KNG1 was further supported by our analysis, which showed that *cis* pQTLs for F11 was a *trans* pQTL for KNG1 levels, and vice versa (Supplementary Figs. S2 and S3).

Integration of existing biological evidence

We integrated data from multiple publicly available databases to evaluate the therapeutic potential of the candidate proteins. Three proteins (F11, FURIN and MMP12) have already been investigated in clinical trials with either approved or experimental inhibitors of these proteins (Table 3). Using data from the Open Targets platform, we further assessed the feasibility of targeting the proteins of interest (Table 3). Nine proteins were identified as potentially tractable, meaning that either they function as membrane or secreted proteins, bind specific ligands or small molecules, or they possess predicted binding pockets. Six proteins were highlighted as potentially promising drug targets, supported by the presence of identical gene sequences in mice or the availability of high-quality chemical probes.

We also used data from the Human Protein Atlas to examine the expression of the genes related to these proteins across multiple tissues and single cells. Six proteins showed high specificity in tissues relevant to ischemic stroke (F11, FURIN, KNG1 and MMUT in liver; LRP4 in brain; and PROCR in adipose tissue), and seven proteins showed high specificity in single cells relevant to ischemic stroke (F11, FURIN, KNG1 and MMUT in hepatocytes; MMUT and CEP85 in cardiomyocytes; LRP4 in oligodendrocytes precursor and astrocytes; and GRK5 in smooth muscle cells and microglial cells). These proteins showed direct biological relevance to ischemic stroke through their highly specific gene expression patterns.

We also reviewed known human genetic disorders from the OMIM database to better understand the biological roles of these proteins. Known genetic disorders were identified for *F11* (deficiency of Factor XI), *KNG1* (kininogen deficiency, and hereditary angioedema), *MMUT* (methylmalonic aciduria), and *LRP4* (myasthenic syndrome, Cenani-Lenz syndactyly syndrome, and sclerosteosis).

We further explored the biological relevance of these proteins to the pathogenesis of ischemic stroke and potential safety considerations using the Mouse Genome Informatics database. Knockout models have revealed that inactivation of *KNG1* and *F11* reduces the risk of thrombotic events. Deletion of *SPATA20* leads to male infertility, while loss-of-function

mutations in *FURIN* and *LRP4* are linked to various pathological phenotypes during embryonic development, indicating their potential lack of suitability as drug targets for women of reproductive age. *LRP4* knockout also led to significant abnormalities in motor neuron morphology.

Cumulative evidence on the potential for drug targeting

By synthesizing evidence from all above-mentioned sources (Table 3), agonists of MMUT, and inhibitors of KNG1 and F11 emerge as promising drug targets. Indeed, MMUT is a promising drug target due to its tractability, doability, high tissue specificity, and potential absence of associations that may represent undesirable effects, whereas loss-of-function mutations in *MMUT* have been linked to ischemic stroke in the context of methylmalonic aciduria, supporting its biological relevance. KNG1 and F11 inhibitors are also promising candidate drugs, with pharmacological inhibitors already available for both proteins, and additional supportive evidence from knockout animal models. However, caution may be needed due to their complex protein interaction network of KNG1 and F11, as we showed earlier (Fig. 4D, E). Similarly, agonists of GRK5 and CEP85 may be suitable targets for prevention of ischemic stroke with phenome-wide MR linking both of them to platelet traits.

The remaining candidate proteins display potential concerns as drug targets. Based on findings from animal models, LRP4 and FURIN inhibitors may lead to abnormalities in embryonic development, raising major safety concerns in women of reproductive age. Similarly, our phenome-wide MR showed that PROCR agonists present a trade-off: they are associated with lower risk of ischemic stroke but higher risk of venous thromboembolism. Meanwhile, the absence of tissue specificity for MMP12, SH3GBRL3 and SPATA20 would de-prioritize them as potential drug targets, given narrow gene expression (i.e., high tissue specificity) is considered desirable for a drug target due to reduced risk of side effects⁵⁷.

Discussion

After performing a comprehensive assessment of more than 3500 plasma proteins for an association with ischemic stroke and its subtypes, our study provided genetic evidence through MR and colocalization for four proteins (KNG1, MMUT, CEP85, and SPATA20), which have not been previously implicated in the risk of ischemic stroke, and seven previously-identified proteins (GRK5, F11, PROCR, FURIN, LRP4, SH3GBRL3, and MMP12)

involved in the pathogenesis of ischemic stroke. These proteins highlight particular physiological pathways for ischemic stroke, including coagulation process, regulation of blood pressure, and energy metabolism.

Our findings underscore the role of the coagulation pathway in the pathogenesis of ischemic stroke by identifying three proteins (F11, KNG1, PROCR) directly involved in coagulation. We found that higher plasma F11 leads to higher risk of ischemic stroke, which is consistent with previous MR studies⁶⁷. This finding is further supported by mouse models, where inhibition of F11 protects against acute ischemic stroke⁵⁸. Observational studies suggest that patients with severe F11 deficiency have lower incidence of ischemic stroke⁵⁹, while elevated F11 is associated with higher risk of ischemic stroke⁶⁰. Phase II randomized controlled trials (RCTs) have yielded inconclusive results in terms of efficacy and safety of F11 inhibitors for primary and secondary prevention of ischemic stroke⁶¹. Moreover, F11 interacts with KNG1, which is cleaved by kallikrein to release the pro-inflammatory mediator bradykinin⁶². Inhibiting KNG1 has been shown to reduce thrombus formation, improve cerebral blood flow, and protect from blood-brain barrier damage in mouse models⁶³. *KNG1* genetic locus has been linked to activated partial thromboplastin time further supporting its role in coagulation^{64,65}. Additionally, PROCR is involved in coagulation by acting as a receptor for protein C⁶⁶. Preclinical studies have shown that protein C has anti-inflammatory, antithrombotic and neuroprotective properties with potential benefits for protein C as a drug target^{67–69}. Protein C was associated with risk of ischemic stroke in the ARIC study⁷⁰. However, we showed that although PROCR agonists protect from risk of ischemic stroke, they are associated with higher risk of deep venous thrombosis and pulmonary embolism, which is consistent with previous literature⁷¹.

Two additional proteins (GRK5, LRP4) are indirectly involved in the coagulation process. GRK5 is a member of the G protein-coupled receptor kinases, which have been previously suggested as potential drug targets for cardiovascular diseases^{72,73}. GRK5 is involved in the regulation of thrombin-activated signaling in endothelial cells and platelets^{74,75}, which is consistent with our finding of GRK5 associated with platelet markers. Similarly, LRP4 may participate in thrombotic cerebrovascular events by acting as a receptor of von Willebrand Factor (vWF), which is a key protein in the extrinsic coagulation pathway facilitating the binding of platelets to the injured vessel wall^{76–78}. Targeting the LRP4–vWF complex has been suggested as a potential drug target for arterial wall remodeling⁷⁶. Also, *LRP4* is highly expressed in brain tissue, particularly in astrocytes and oligodendrocytes, and there is evidence that LRP4 is critical to maintain the function of the neuromuscular junction⁷⁹. Therefore, inhibiting LRP4 could lead to severe neurological adverse effects, as antibodies against LRP4 have been associated with myasthenia gravis and other myasthenia syndromes^{79,80}.

Consistent with previous MR studies, we found that higher genetically-predicted plasma levels of MMP12 is related to lower risk of ischemic stroke⁶⁷. MMP12 is involved in multiple biological functions⁸¹. In the context of ischemic stroke, a plausible mechanism supporting a protective role of MMP12 involves the regulation of coagulation through the cleavage of fibrinogen, a key glycoprotein that is converted into fibrin during clot formation, thereby reducing the substrate availability for thrombus formation⁸². In support of this hypothesis, evidence from animal models showed that knockout of *MMP12* leads to higher plasma levels of fibrinogen⁸³.

Regulation of systolic blood pressure is another important pathway involved into the pathogenesis of ischemic stroke¹⁰, and our study showed that higher plasma levels of *FURIN* were associated with both higher risk of ischemic stroke and systolic blood pressure. *FURIN* is widely expressed across multiple tissues and regulates multiple physiological functions, including vascular remodeling, progression of atherosclerotic lesions, blood clotting system and the complement system^{84–86}. This is consistent with our phenome-wide MR linking *FURIN* to circulating lipids, platelet markers, and liver and kidney function. However, mouse models indicate *FURIN* knockout is linked to developmental abnormalities, potentially limiting its suitability as a drug target for women of reproductive age⁸⁷.

Our study offers insights into the potential role of energy metabolism and tricarboxylic acid cycle in the development of ischemic stroke through MMUT (methylmalonyl-CoA mutase), which is a vitamin B₁₂-dependent mitochondrial enzyme, catalyzing the isomerisation of methylmalonyl-CoA to succinyl-CoA⁸⁸. Genetic deficiency of MMUT leads to methylmalonic aciduria, a type of inborn error of metabolism, characterized by basal ganglia stroke^{89,90}. Additionally, reduced activity of MMUT can be caused by vitamin B₁₂ deficiency, which is consistent with our phenome-wide MR linking MMUT levels with markers of anemia due to B₁₂ deficiency, including mean corpuscular volume and red blood cell distribution width. Therefore, disruption of the tricarboxylic acid cycle or brain hypo-perfusion due to anemia may explain the association of MMUT with ischemic stroke.

Our findings shed light into biological pathways through which CEP85 and SH3BGR3 may contribute to the pathogenesis of ischemic stroke. Their genes are located in the same genetic locus, previously linked to platelet markers and LDL cholesterol^{91,92}. Our phenome-wide MR further showed associations of CEP85 with platelet markers and lipids, and SH3BGR3 with platelet markers and kidney function, pointing to potential pathways contributing to cerebrovascular pathology.

Our study showed an association between genetically-predicted SPATA20 and small vessel stroke, as well as associations with type 2 diabetes and systolic blood pressure, which are established risk factors for small vessel disease⁹³. These findings offer insights into potential mechanistic pathways through which SPATA20 may contribute to the pathogenesis of small vessel stroke. Furthermore, the *SPATA20* locus has been associated with kidney function⁹⁴, and our phenome-wide MR analysis confirmed this in the UK Biobank. Indeed, there is observational evidence linking impaired kidney function with small vessel disease⁹⁵, and kidney impairment is the result of end organ damage due to both hypertension and type 2 diabetes.

In our study, we applied the MR approach, an instrumental variable technique with genetic variants serving as instruments for protein abundance to infer its causal effect on risk of ischemic stroke. MR relies on three core assumptions that should be satisfied by a genetic variant⁹⁶. First, the genetic variants should be associated with protein abundance (*relevance assumption*). This assumption was satisfied by using a strict statistical significance threshold for variant selection and requiring an *F*-statistic greater than 10. Second, the genetic variants should be independent of any confounders of the association of protein abundance with risk of ischemic stroke (*independence assumption*). Although this assumption is not directly testable, we sought to minimize its violation by selecting only *cis* pQTLs, which are less prone to pleiotropy than *trans* pQTLs. In addition, we tested the genetic effects of the selected *cis* pQTLs on all the Olink and SomaScan proteins measured in the UK Biobank and deCODE study. Third, the genetic variants should affect risk of ischemic stroke only through the protein of interest and not through alternative pathways (*exclusion restriction assumption*). While this assumption is also untestable, we assessed its potential violation by performing colocalization analyses assuming a single causal variant at the *cis* region.

Currently, eight proteome-wide MR studies have investigated the role of plasma proteome in the risk of ischemic stroke and its subtypes. The majority of these studies used smaller proteogenomic datasets alongside the meta-analysis of GWAS data for ischemic stroke by the MEGASTROKE Consortium, which includes about 34,000 ischemic stroke cases of European ancestry. In our analysis, we confirmed the associations of GRK5, F11, MMP12, LRP4, PROCR, SH3BGR3 and *FURIN* with ischemic stroke that were reported by the previous MR studies. In addition to this, we also provided genetic evidence supporting the association of MMUT, CEP85 and KNG1 with ischemic stroke, and the association of SPATA20 with small vessel stroke.

Our study leverages data on more proteomic markers, including comprehensively comparing data from two proteomic platforms and thereby providing greater discovery potential than in previously published work. In addition, we enhanced power and phenotypic depth by utilizing the largest available GWAS meta-analysis of ischemic stroke from the GIGA-STROKE Consortium, including nearly twice as many ischemic stroke cases

of European ancestry (62,100 cases) than previously available. We also provide a comprehensive framework of downstream analyses that offers additional insights into the biological evidence for the proteins identified, and we examined the effects of the stroke-associated proteins on established cardiovascular risk factors, using the largest and most up-to-date GWAS. Furthermore, we agnostically tested for potential additional effects of the stroke-associated proteins across the phenome, by performing a PheWAS in the UK Biobank and FinnGen, and conducting a comprehensive assessment of potential pleiotropic effects of the *cis* pQTLs by testing their effects on other plasma proteins in order to distinguish between occasions of mediated and confounded effects by capitalizing on biological information. This unique combination of data and downstream approaches offers meaningful biological insights and advances that have not been considered in previous proteome-wide MR studies of ischemic stroke.

Our study has several limitations. Firstly, although our study was the largest to date, we had limited power to identify plasma proteins that are specific to individual ischemic stroke subtypes. In addition, we were unable to examine plasma proteins that did not have available *cis* pQTLs. Third, MR estimates assume lifelong exposure to altered protein levels and therefore do not necessarily reflect the impact on a therapeutic intervention. Fourth, while we assessed pleiotropy by evaluating the association of *cis* pQTLs with other plasma proteins, pleiotropic effects on unmeasured proteins cannot be ruled out. Fifth, *cis* pQTLs were only available in European ancestry individuals, and further validation in other populations is needed. We were able to evaluate the associations of these *cis* pQTLs in the large-scale multi-ancestry GIGASTROKE dataset, as well as in Europeans only, and established directionally consistent causal effects. Identification of ancestry-specific causal proteins requires further study. Finally, despite integrating bioinformatics databases to distinguish between vertical and horizontal pleiotropy, our ability to do so is constrained by the current gaps in biological knowledge.

Conclusion

Leveraging large-scale genetic and proteomic data, we identified 11 plasma proteins with likely causal relevance for ischemic stroke, including seven previously-identified proteins (GRK5, F11, PROCR, FURIN, LRP4, SH3BGR13, and MMP12) and four proteins (MMUT, CEP85, KNG1, and SPATA20) that have not previously implicated in the risk of ischemic stroke. By integrating prior biological knowledge, we were able to show that these candidate proteins are involved in key physiological pathways, including coagulation, regulation of blood pressure and energy metabolism. Also, after considering potential undesirable effects and tissue specificity, CEP85, F11, GRK5, KNG1, and MMUT were the most promising candidates among those we considered for drug development aimed at ischemic stroke prevention. The continually expanding coverage of proteomic panels will facilitate the study of more plasma proteins and support development of new methods to harness *trans* pQTLs, as well as offer considerable future potential for further elucidating the role of proteins in ischemic stroke pathogenesis.

Data availability

Source data for Fig. 2 can be found in Supplementary Data 2. Source data for Fig. 3 can be found in Supplementary Data 2 and 8. Source data for Fig. 4 can be found in Supplementary Data 8 and 9. All data used in the statistical analyses of this research study are publicly available. The *cis* pQTLs that were used as instrumental variables are publicly available in the relevant publications^{8,9}. Specifically, pQTLs from the Olink platform in the UK Biobank are publicly available through Synapse (<https://www.synapse.org/Synapse:syn51364943/wiki/622119>), whereas pQTLs from the SomaScan platform in the deCODE Health Study are publicly available in the deCODE web portal (<https://www.decode.com/summarydata/>). GWAS summary statistics for ischemic stroke, ischemic stroke subtypes, and systolic blood pressure are publicly available through GWAS Catalog (<https://www.ebi.ac.uk/gwas/>), under the following accession numbers: GCST90104540 for any ischemic stroke, GCST90104542 for large artery stroke, GCST90104541 for

cardioembolic stroke, GCST90104543 for small vessel stroke, and GCST90310294 for systolic blood pressure. GWAS summary statistics for ever smoking are publicly available through GWAS and Sequencing Consortium of Alcohol and Nicotine use (<https://doi.org/10.13020/przg-dp88>). GWAS summary statistics for body mass index are publicly available through GIANT consortium (https://giant-consortium.web.broadinstitute.org/GIANT_consortium_data_files). GWAS summary statistics for type 2 diabetes are publicly available through DIAGRAM consortium (<https://www.diagram-consortium.org/downloads.html>). GWAS summary statistics for atrial fibrillation are publicly available through the Cardiovascular Knowledge Portal (<https://cvd.hugeamp.org/downloads.html>).

Code availability

We used publicly available software for the statistical analyses in our study. Mendelian randomization analyses were performed using the TwoSampleMR package (<https://mrcieu.github.io/TwoSampleMR/>). Genetic colocalization analyses were conducted using the coloc package (<https://github.com/chr1swallace/coloc>). Linkage disequilibrium clumping was performed using the ieugwasr package (<https://github.com/MRCIEU/ieugwasr>) and PLINK version 1.90 (<https://www.cog-genomics.org/plink/>). Enrichment analysis was performed using the clusterProfiler package (<https://github.com/YuLab-SMU/clusterProfiler>).

Received: 24 December 2025; Accepted: 9 June 2026;

Published online: 03 July 2026

References

1. Feigin, V. L. et al. Global, regional, and national burden of stroke and its risk factors, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet Neurol.* **20**, 795–820 (2021).
2. Hilkens, N. A., Casolla, B., Leung, T. W. & de Leeuw, F. E. Stroke. *Lancet* **403**, 2820–2836 (2024).
3. Zheng, J. et al. Phenome-wide Mendelian randomization mapping the influence of the plasma proteome on complex diseases. *Nat. Genet.* **52**, 1122–1131 (2020).
4. Suhre, K., McCarthy, M. I. & Schwenk, J. M. Genetics meets proteomics: perspectives for large population-based studies. *Nat. Rev. Genet.* **22**, 19–37 (2021).
5. Aguet, F. et al. Molecular quantitative trait loci. *Nat. Rev. Methods Prim.* **3**, 4 (2023).
6. Chen, L. et al. Systematic Mendelian randomization using the human plasma proteome to discover potential therapeutic targets for stroke. *Nat. Commun.* **13**, 6143 (2022).
7. Chong, M. et al. Novel drug targets for ischemic stroke identified through Mendelian randomization analysis of the blood proteome. *Circulation* **140**, 819–830 (2019).
8. Sun, B. B. et al. Plasma proteomic associations with genetics and health in the UK Biobank. *Nature* **622**, 329–338 (2023).
9. Ferkingstad, E. et al. Large-scale integration of the plasma proteome with genetics and disease. *Nat. Genet.* **53**, 1712–1721 (2021).
10. Mishra, A. et al. Stroke genetics informs drug discovery and risk prediction across ancestries. *Nature* **611**, 115–123 (2022).
11. Hartwig, F. P., Davies, N. M., Hemani, G. & Davey Smith, G. Two-sample Mendelian randomization: avoiding the downsides of a powerful, widely applicable but potentially fallible technique. *Int. J. Epidemiol.* **45**, 1717–1726 (2016).
12. Burgess, S., Thompson, S. G. & CRP CHD Genetics Collaboration. Avoiding bias from weak instruments in Mendelian randomization studies. *Int. J. Epidemiol.* **40**, 755–764 (2011).
13. Hemani, G., Tilling, K. & Davey Smith, G. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. *PLoS Genet.* **13**, e1007081 (2017).
14. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).

15. Burgess, S., Scott, R. A., Timpson, N. J., Davey Smith, G., Thompson, S. G. & EPIC- InterAct Consortium. Using published data in Mendelian randomization: a blueprint for efficient identification of causal risk factors. *Eur. J. Epidemiol.* **30**, 543–552 (2015).
16. Pietzner, M. et al. Synergistic insights into human health from aptamer- and antibody-based proteomic profiling. *Nat. Commun.* **12**, 6822 (2021).
17. Burgess, S., Small, D. S. & Thompson, S. G. A review of instrumental variable estimators for Mendelian randomization. *Stat. Methods Med. Res.* **26**, 2333–2355 (2017).
18. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* **57**, 289–300 (1995).
19. Bowden, J. & Holmes, M. V. Meta-analysis and Mendelian randomization: a review. *Res. Synth. Methods* **10**, 486–496 (2019).
20. Zuber, V. et al. Combining evidence from Mendelian randomization and colocalization: review and comparison of approaches. *Am. J. Hum. Genet.* **109**, 767–782 (2022).
21. Giambartolomei, C. et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* **10**, e1004383 (2014).
22. van der Graaf, A. et al. MR-link-2: pleiotropy robust cis Mendelian randomization validated in three independent reference datasets of causality. *Nat. Commun.* **16**, 6112 (2025).
23. van der Graaf, A. et al. Mendelian randomization while jointly modeling cis genetics identifies causal relationships between gene expression and lipids. *Nat. Commun.* **11**, 4930 (2020).
24. Welter, D. et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.* **42**, D1001–D1006 (2014).
25. Saunders, G. R. B. et al. Genetic diversity fuels gene discovery for tobacco and alcohol use. *Nature* **612**, 720–724 (2022).
26. Keaton, J. M. et al. Genome-wide analysis in over 1 million individuals of European ancestry yields improved polygenic risk scores for blood pressure traits. *Nat. Genet.* **56**, 778–791 (2024).
27. Suzuki, K. et al. Genetic drivers of heterogeneity in type 2 diabetes pathophysiology. *Nature* **627**, 347–357 (2024).
28. Pulit, S. L. et al. Meta-analysis of genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry. *Hum. Mol. Genet.* **28**, 166–174 (2019).
29. Roselli, C. et al. Multi-ethnic genome-wide association study for atrial fibrillation. *Nat. Genet.* **50**, 1225–1233 (2018).
30. Kurki, M. I. et al. FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature* **613**, 508–518 (2023).
31. Ashburner, M. et al. Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).
32. Jassal, B. et al. The reactome pathway knowledgebase. *Nucleic Acids Res.* **48**, D498–D503 (2020).
33. Kanehisa, M. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
34. Wu, T. et al. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innovation* **2**, 100141 (2021).
35. Tipney, H. & Hunter, L. An introduction to effective use of enrichment analysis software. *Hum. Genomics* **4**, 202 (2010).
36. Wijesooriya, K., Jadaan, S. A., Perera, K. L., Kaur, T. & Ziemann, M. Urgent need for consistent standards in functional enrichment analysis. *PLoS Comput. Biol.* **18**, e1009935 (2022).
37. Szklarczyk, D. et al. The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res.* **51**, D638–D646 (2023).
38. Amberger, J. S., Bocchini, C. A., Schiettecatte, F., Scott, A. F. & Hamosh, A. OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. *Nucleic Acids Res.* **43**, D789–D798 (2015).
39. Aleksander, S. A. et al. The Gene Ontology knowledgebase in 2026. *Nucleic Acids Res.* **54**, D1779–D1792 (2026).
40. Ochoa, D. et al. Open Targets Platform: supporting systematic drug–target identification and prioritisation. *Nucleic Acids Res.* **49**, D1302–D1310 (2021).
41. Ochoa, D. et al. The next-generation Open Targets Platform: reimagined, redesigned, rebuilt. *Nucleic Acids Res.* **51**, D1353–D1359 (2023).
42. Zhou, Y. et al. Therapeutic target database update 2022: facilitating drug discovery with enriched comparative data of targeted agents. *Nucleic Acids Res.* **50**, D1398–D1407 (2022).
43. Wishart, D. S. et al. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* **46**, D1074–D1082 (2018).
44. Whirl-Carrillo, M. et al. An evidence-based framework for evaluating pharmacogenomics knowledge for personalized medicine. *Clin. Pharm. Ther.* **110**, 563–572 (2021).
45. Brown, K. K. et al. Approaches to target tractability assessment—a practical perspective. *Medchemcomm* **9**, 606–613 (2018).
46. Baldarelli, R. M. et al. Mouse Genome Informatics: an integrated knowledgebase system for the laboratory mouse. *Genetics* **227**, iyae031 (2024).
47. Sjöstedt, E. et al. An atlas of the protein-coding genes in the human, pig, and mouse brain. *Science* **367**, eaay5947 (2020).
48. Digre, A. & Lindskog, C. The human protein atlas—integrated omics for single cell mapping of the human proteome. *Protein Sci.* **32**, e4562 (2023).
49. Lu, T., Forgetta, V., Greenwood, C. M. T., Zhou, S. & Richards, J. B. Circulating proteins influencing psychiatric disease: a Mendelian randomization study. *Biol. Psychiatry* **93**, 82–91 (2023).
50. Wu, B. S. et al. Identifying causal genes for stroke via integrating the proteome and transcriptome from brain and blood. *J. Transl. Med.* **20**, 181 (2022).
51. Wu, S. et al. Mendelian randomization and Bayesian colocalization analysis implicate glycoprotein VI as a potential drug target for cardioembolic stroke in South Asian populations. *J. Am. Heart Assoc.* **13**, e035008 (2024).
52. Yao, P. et al. Proteome-wide genetic study in East Asians and Europeans identified multiple therapeutic targets for ischemic stroke. *Stroke* **56**, 2147–2158 (2025).
53. Zhang, P., He, Y., Zhen, Q. & Zhang, Y. Systematic exploration of potential druggable genes for ischemic stroke employing genome-wide Mendelian randomization analysis. *Brain Behav.* **15**, e70857 (2025).
54. Zhao, X. et al. Proteome-wide Mendelian randomization and colocalization analysis identify therapeutic targets for stroke. *BMC Neurol.* **25**, 255 (2025).
55. Zhu, Y., Wang, Y., Cui, Z., Liu, F. & Hu, J. Identification of pleiotropic and specific therapeutic targets for cardio-cerebral diseases: a large-scale proteome-wide Mendelian randomization and colocalization study. *PLoS ONE* **19**, e0300500 (2024).
56. Zou, X. et al. Promising therapeutic targets for ischemic stroke identified from plasma and cerebrospinal fluid proteomes: a multicenter Mendelian randomization study. *Int. J. Surg.* **110**, 766–776 (2024).
57. Ryaboshapkina, M. & Hammar, M. Tissue-specific genes as an underutilized resource in drug discovery. *Sci. Rep.* **9**, 7233 (2019).
58. Leung, P. Y. et al. Inhibition of factor XII-mediated activation of factor XI provides protection against experimental acute ischemic stroke in mice. *Transl. Stroke Res.* **3**, 381–389 (2012).
59. Salomon, O., Steinberg, D. M., Koren-Morag, N., Tanne, D. & Seligsohn, U. Reduced incidence of ischemic stroke in patients with severe factor XI deficiency. *Blood* **111**, 4113–4117 (2008).
60. Suri, M. F. K., Yamagishi, K., Aleksic, N., Hannan, P. J. & Folsom, A. R. Novel hemostatic factor levels and risk of ischemic stroke: the

- Atherosclerosis Risk in Communities (ARIC) Study. *Cerebrovasc. Dis.* **29**, 497–502 (2010).
61. Gigante, B. & ten Cate, H. Factor XI inhibitors in patients with cardiovascular disease and a high risk of bleeding: a cautionary tale. *Nat. Rev. Cardiol.* **20**, 511–512 (2023).
 62. Björkqvist, J., Jämsä, A. & Renné, T. Plasma kallikrein: the bradykinin-producing enzyme. *Thromb. Haemost.* **110**, 399–407 (2013).
 63. Langhauser, F. et al. Kininogen deficiency protects from ischemic neurodegeneration in mice by reducing thrombosis, blood-brain barrier damage, and inflammation. *Blood* **120**, 4082–4092 (2012).
 64. Sabater-Lleal, M. et al. A genome-wide association study identifies *KNG1* as a genetic determinant of plasma factor XI level and activated partial thromboplastin time. *Arterioscler Thromb. Vasc. Biol.* **32**, 2008–2016 (2012).
 65. Houlihan, L. M. et al. Common variants of large effect in F12, *KNG1*, and HRG are associated with activated partial thromboplastin time. *Am. J. Hum. Genet.* **86**, 626–631 (2010).
 66. Mosnier, L. O., Zlokovic, B. V. & Griffin, J. H. The cytoprotective protein C pathway. *Blood* **109**, 3161–3172 (2007).
 67. Mosnier, L., Zlokovic, B. & Griffin, J. Cytoprotective-selective activated protein C therapy for ischaemic stroke. *Thromb. Haemost.* **112**, 883–892 (2014).
 68. Shibata, M. et al. Anti-inflammatory, antithrombotic, and neuroprotective effects of activated protein C in a murine model of focal ischemic stroke. *Circulation* **103**, 1799–1805 (2001).
 69. Griffin, J. H., Zlokovic, B. V. & Mosnier, L. O. Activated protein C: biased for translation. *Blood* **125**, 2898–2907 (2015).
 70. Folsom, A. R., Ohira, T., Yamagishi, K. & Cushman, M. Low protein C and incidence of ischemic stroke and coronary heart disease: the Atherosclerosis Risk in Communities (ARIC) Study. *J. Thromb. Haemost.* **7**, 1774–1778 (2009).
 71. Stacey, D. et al. Elucidating mechanisms of genetic cross-disease associations at the PROCRA vascular disease locus. *Nat. Commun.* **13**, 1222 (2022).
 72. Premont, R. T. & Gainetdinov, R. R. Physiological roles of G protein-coupled receptor kinases and arrestins. *Annu. Rev. Physiol.* **69**, 511–534 (2007).
 73. Pflieger, J., Gresham, K. & Koch, W. J. G protein-coupled receptor kinases as therapeutic targets in the heart. *Nat. Rev. Cardiol.* **16**, 612–622 (2019).
 74. Tirupathi, C. et al. G protein-coupled receptor kinase-5 regulates thrombin-activated signaling in endothelial cells. *Proc. Natl. Acad. Sci. USA.* **97**, 7440–7445 (2000).
 75. Downes, K. et al. G protein-coupled receptor kinase 5 regulates thrombin signaling in platelets via PAR-1. *Blood Adv.* **6**, 2319–2330 (2022).
 76. Lagrange, J. et al. The VWF/LRP4/ α V β 3-axis represents a novel pathway regulating proliferation of human vascular smooth muscle cells. *Cardiovasc. Res.* **118**, 622–637 (2022).
 77. Khrenov, A. V., Ananyeva, N. M., Griffin, J. H. & Saenko, E. L. Coagulation pathways in atherothrombosis. *Trends Cardiovasc. Med.* **12**, 317–324 (2002).
 78. Mackman, N., Tilley, R. E. & Key, N. S. Role of the extrinsic pathway of blood coagulation in hemostasis and thrombosis. *Arterioscler Thromb. Vasc. Biol.* **27**, 1687–1693 (2007).
 79. Barik, A. et al. LRP4 is critical for neuromuscular junction maintenance. *J. Neurosci.* **34**, 13892–13905 (2014).
 80. Selcen, D. et al. Impaired synaptic development, maintenance, and neuromuscular transmission in LRP4-related myasthenia. *JAMA Neurol.* **72**, 889–896 (2015).
 81. Hiller, O., Lichte, A., Oberpichler, A., Kocourek, A. & Tschesche, H. Matrix metalloproteinases collagenase-2, macrophage elastase, collagenase-3, and membrane type 1-matrix metalloproteinase impair clotting by degradation of fibrinogen and factor XII. *J. Biol. Chem.* **275**, 33008–33013 (2000).
 82. Sternlicht, M. D. & Werb, Z. How matrix metalloproteinases regulate cell behavior. *Annu. Rev. Cell Dev. Biol.* **17**, 463–516 (2001).
 83. Motterle, A. et al. Influence of matrix metalloproteinase-12 on fibrinogen level. *Atherosclerosis* **220**, 351–354 (2012).
 84. Yakala, G. K. et al. FURIN inhibition reduces vascular remodeling and atherosclerotic lesion progression in mice. *Arterioscler Thromb. Vasc. Biol.* **39**, 387–401 (2019).
 85. Chen, H. et al. FURIN suppresses the progression of atherosclerosis by promoting macrophage autophagy. *FASEB J.* **37**, e22933 (2023).
 86. Langnau, C. et al. Platelet activation and plasma levels of furin are associated with prognosis of patients with coronary artery disease and COVID-19. *Arterioscler Thromb. Vasc. Biol.* **41**, 2080–2096 (2021).
 87. Dupays, L. et al. Furin, a transcriptional target of NKX2-5, has an essential role in heart development and function. *PLoS ONE* **14**, e0212992 (2019).
 88. Nielsen, M. J., Rasmussen, M. R., Andersen, C. B. F., Nexø, E. & Moestrup, S. K. Vitamin B12 transport from food to the body's cells — a sophisticated, multistep pathway. *Nat. Rev. Gastroenterol. Hepatol.* **9**, 345–354 (2012).
 89. Mastrangelo, M. et al. Stroke and stroke-like episodes in inborn errors of metabolism: pathophysiological and clinical implications. *Mol. Genet. Metab.* **135**, 3–14 (2022).
 90. Wajner, M. Neurological manifestations of organic acidurias. *Nat. Rev. Neurol.* **15**, 253–271 (2019).
 91. Vuckovic, D. et al. The polygenic and monogenic basis of blood traits and diseases. *Cell* **182**, 1214–1231.e11 (2020).
 92. Barton, A. R., Sherman, M. A., Mukamel, R. E. & Loh, P. R. Whole-exome imputation within UK Biobank powers rare coding variant association and fine-mapping analyses. *Nat. Genet.* **53**, 1260–1269 (2021).
 93. Bordes, C., Sargurupremraj, M., Mishra, A. & Debette, S. Genetics of common cerebral small vessel disease. *Nat. Rev. Neurol.* **18**, 84–101 (2022).
 94. Sinnott-Armstrong, N. et al. Genetics of 35 blood and urine biomarkers in the UK Biobank. *Nat. Genet.* **53**, 185–194 (2021).
 95. Ikram, M. A. et al. Kidney function is related to cerebral small vessel disease. *Stroke* **39**, 55–61 (2008).
 96. Sanderson, E. et al. Mendelian randomization. *Nat. Rev. Methods Prim.* **2**, 6 (2022).
 97. Malik, R. et al. Multi-ancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nat. Genet.* **50**, 524–537 (2018).
 98. Sun, B. B. et al. Genomic atlas of the human plasma proteome. *Nature* **558**, 73–79 (2018).
 99. Emilsson, V. et al. Co-regulatory networks of human serum proteins link genetics to disease. *Science* **361**, 769–773 (2018).
 100. Zhang, J. et al. Plasma proteome analyses in individuals of European and African ancestry identify cis-pQTLs and models for proteome-wide association studies. *Nat. Genet.* **54**, 593–602 (2022).
 101. Suhre, K. et al. Connecting genetic risk to disease end points through the human blood plasma proteome. *Nat. Commun.* **8**, 14357 (2017).
 102. Yao, C. et al. Genome-wide mapping of plasma protein QTLs identifies putatively causal genes and pathways for cardiovascular disease. *Nat. Commun.* **9**, 3268 (2018).
 103. Folkersen, L. et al. Mapping of 79 loci for 83 plasma protein biomarkers in cardiovascular disease. *PLoS Genet.* **13**, e1006706 (2017).

Acknowledgements

L.B. is supported by a British Heart Foundation (BHF) Oxford Centre of Research Excellence Intermediate Transition Research Fellowship (Grant

Number: RE/18/3/34214). C.v.D. is supported by the US National Institute on Aging, the Novo Nordisk—Oxford Fellowship Programme, the Oxford—GlaxoSmithKline Institute of Molecular and Computational Medicine, the Centre for Artificial Intelligence for Precision Medicine at the University of Oxford and King Abdul Aziz University, Alzheimer’s UK, the National Institute for Health and Care Research Biomedical Research Centre: Oxford, the ZonMW, and the Alzheimer Nederland. J.C.H. is supported by the National Institute for Health and Care Research (NIHR) Biomedical Research Centre: Oxford, the BHF and BHF Oxford Centre of Research Excellence, by research grants from industry held in accordance with the policy detailed in the “competing interests” section below, and the Nuffield Department of Population Health. The computational aspects of this research were supported by the Wellcome Trust Core Award Grant Number 203141/Z/16/Z and the NIHR Biomedical Research Centre: Oxford. The views expressed are those of the author(s) and not necessarily those of the UK National Health Service, the NIHR or the UK Department of Health and Social Care.

Author contributions

L.B. conceived the study, developed the analytical strategy, performed all statistical analyses, and generated all figures, tables, and Supplementary Materials. L.B. drafted the initial version of the manuscript, with J.C.H. and C.v.D. providing critical feedback on the manuscript and contributing to its revision. A.v.E., P.G., E.V.M. and F.M. critically reviewed the final manuscript. J.C.H. and C.v.D. provided overall supervision and scientific guidance. All authors approved the final version of the manuscript.

Competing interests

The authors declare no relevant competing interests. All authors work at the Nuffield Department of Population Health, which receives research grants from industry that are governed by University of Oxford contracts that protect its independence, and has a staff policy of not taking personal payments from industry. Further details can be found at <https://www.ndph.ox.ac.uk/about/independence-of-research>.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s43856-026-01734-z>.

Correspondence and requests for materials should be addressed to Lazaros Belbasis or Jemma C. Hopewell.

Peer review information *Communications Medicine* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://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) 2026