Stroke

CLINICAL AND POPULATION SCIENCES

Proteome-Wide Genetic Study in East Asians and Europeans Identified Multiple Therapeutic Targets for Ischemic Stroke

Pang Yao, PhD; Mohsen Mazidi, PhD; Alfred Pozarickij, PhD; Andri Iona, DPhil; Neil Wright, PhD; Kuang Lin, PhD; Iona Y. Millwood, DPhil; Hannah Fry, MSc; Christiana Kartsonaki, DPhil; Yiping Chen, DPhil; Ling Yang, PhD; Huaidong Du, PhD; Daniel Avery, MSc; Dan Schmidt-Valle, MSc; Dianjianyi Sun, PhD; Pei Pei, BSc; Jun Lv, PhD; Canqing Yu, PhD; Michael Hill, PhD; Derrick A. Bennett, PhD; Robin G. Walters, PhD; Liming Li, MD; Robert Clarke, FRCP; Zhengming Chen, DPhil; on behalf of China Kadoorie Biobank Collaborative Group*

BACKGROUND: Analyses of genomic and proteomics data in prospective biobank studies in diverse populations may discover novel or repurposing drug targets for stroke.

METHODS: We extracted individual *cis*-protein quantitative trait locus for 2923 proteins measured using Olink Explore panel from a genome-wide association study in prospective China Kadoorie Biobank and UK Biobank, both established ≈20 years ago. These *cis*-protein quantitative trait loci were used in ancestry-specific 2-sample Mendelian randomization analyses of ischemic stroke (IS) in East Asians (n=22 664 cases) and Europeans (n=62 100 cases). We further undertook colocalization analyses to examine the shared causal variants of *cis*-protein quantitative trait locus with stroke, along with various downstream analyses (eg, phenome-wide association study, drug development lookups) to clarify mechanisms of action and druggability.

RESULTS: In Mendelian randomization analyses, the genetically predicted plasma levels of 10 proteins were significantly associated with IS in East Asians (n=2) and Europeans (n=9), with 6 proteins (FGF5 [fibroblast growth factor 5], TMPRSS5 [transmembrane protease serine 5], FURIN, F11 [coagulation factor XI], ALDH2 [aldehyde dehydrogenase 2], and ABO [histo-blood group ABO system transferase]) showing positive and 4 (GRK5 [G protein-coupled receptor kinase 5], KIAA0319 [dyslexia-associated protein KIAA0319], PROCR [endothelial protein C receptor], and MMP12 [macrophage metalloelastase 12]) showing inverse associations, all directionally consistent between East Asians and Europeans. Colocalization analyses provided strong evidence (posterior probabilities for the H4 hypothesis ≥0.7) of shared genetic variants with IS for 9 out of 10 proteins (except ABO). Moreover, 8 proteins were also causally associated, in the expected directions, with systolic blood pressure (positive/inverse: 4/2), low-density lipoprotein cholesterol (1 positive), body mass index (1 inverse), type 2 diabetes (2/1), or atrial fibrillation (3/1). Phenome-wide association study analyses and lookups in knock-out mouse models confirmed their importance for IS or stroke-related traits (eg, hematologic phenotypes). Of these 10 proteins, 1 was not druggable (ABO), 3 had known primary (F11) or potentially repurposed (ALDH2, MMP12) drug targets for stroke, and 6 (PROCR, GRK5, FGF5, FURIN, KIAA0319, and TMPRSS5) had no evidence of any drug targets.

CONCLUSIONS: Proteogenomic investigation in diverse ancestry populations identified the causal relevance of 10 proteins for IS, with several being potentially novel or repurposed targets that could be prioritized for further investigation.

GRAPHIC ABSTRACT: A graphic abstract is available for this article.

Key Words: biological specimen banks ■ genetics ■ ischemic stroke ■ Mendelian randomization analysis ■ proteomics

Correspondence to: Zhengming Chen, DPhil, Nuffield Department of Population Health, University of Oxford, Old Rd Campus, Oxford OX3 7LF, United Kingdom. Email zhengming.chen@ndph.ox.ac.uk

*Members of the China Kadoorie Biobank Collaborative Group are shown in the Supplemental Material.

Supplemental Material is available at https://www.ahajournals.org/doi/suppl/10.1161/STROKEAHA.125.050982.

For Sources of Funding and Disclosures, see page 2156.

Nonstandard Abbreviations and Acronyms

AF	atrial fibrillation
ALDH2	aldehyde dehydrogenase 2
Apo[a]	apolipoprotein(a)
CKB	China Kadoorie Biobank
CVD	cardiovascular diseases
EAS	East Asians
EUR	Europeans
F11	coagulation factor XI
FGF5	fibroblast growth factor 5
FVIII	coagulation factor VIII
GRK5	G protein-coupled receptor kinase 5
GWAS	genome-wide association study
IS	ischemic stroke
KIAA0319	dyslexia-associated protein KIAA0319
LDL-C	low-density lipoprotein cholesterol
MMP12	macrophage metalloelastase 12
MR	Mendelian randomization
OR	odds ratio
PheWAS	phenome-wide association study
PP4	posterior probabilities for the H4 hypothesis
pQTL	protein quantitative trait loci
PROCR	endothelial protein C receptor
SBP	systolic blood pressure
SCARA5	scavenger receptor class A member 5
TMPRSS5	transmembrane protease serine 5
T2D	type 2 diabetes
UKB	UK Biobank

Stroke is the second leading cause of death worldwide, causing an estimated >6.5 million annual deaths and a substantial burden of long-term disability.[1,2] Several important modifiable causes of stroke are well-established (eg, hypertension, dyslipidemia, smoking, alcohol drinking),[2] leading to improved prevention and treatment (eg, statin and blood pressure lowering treatment).[2] Despite these, the development of novel therapies for stroke based on new insights from human genetic evidence is needed to target novel causal pathways relevant to stroke or stroke types (eg, ischemic stroke [IS]) specifically. Recently, a trans-ancestry genome-wide association study (GWAS) identified association signals for stroke types at 89 (61 new) independent loci.[3] However, the mechanisms underlying many of these associations remain to be elucidated. Plasma proteins play a central role in human biology and represent a major source of therapeutic targets.[4–7] Analyses of plasma protein biomarkers, particularly when integrated with genomic data in different ancestry populations, could help to clarify disease pathogenesis and discover novel therapeutic targets for stroke.

Mendelian randomization (MR) and colocalization are 2 genetic approaches that are used to clarify the causal relevance of traits (eg, circulating proteins) and diseases (eg, stroke).[8] Previous MR analyses have identified several novel protein biomarkers for stroke, including F11 (coagulation factor XI), MMP12 (macrophage metalloelastase 12), and TMPRSS5 (transmembrane protease serine 5),[9–12] which are involved in thrombus propagation and stabilization. However, the available evidence on proteomics and stroke has been constrained by studies involving primarily European-ancestry populations, relatively small numbers of proteins with robust genetic instruments (eg, *cis*-protein quantitative trait locus [pQTLs]), or moderate case numbers in stroke GWAS consortia.[9–12] Moreover, little is known about the relevance of these emerging and novel protein biomarkers for stroke in Chinese and other ancestry populations where the stroke rates, distribution of risk factors, and genetic architecture differ greatly from Europeans.

We undertook a proteome-wide genetic investigation for IS in China Kadoorie Biobank (CKB) and UK Biobank (UKB) involving ≈3000 proteins. The study aims to (1) use *cis*-pQTLs identified in GWAS for proteins to assess their causal relevance for IS in 2-sample MR and colocalization analyses in global ancestry-specific stroke consortia; (2) reveal pathways and mechanisms of action by which proteins may influence stroke risk, including those that may be mediated through major stroke risk factors (ie, systolic blood pressure [SBP], low-density lipoprotein cholesterol [LDL-C], body mass index, type 2 diabetes [T2D], and atrial fibrillation [AF]); (3) explore the druggability, predicted side-effects and additional indications of causal proteins for stroke using phenome-wide association study (PheWAS), tissue expression, drug databases, and in vitro and in vivo experimental evidence.

METHODS

The study is reported following the STROBE guidelines (Strengthening the Reporting of Observational Studies in Epidemiology) for MR studies (Supplemental Material).

Data Availability

GWAS summary statistics for the proteins were extracted from the UK Biobank Pharma Proteomics Project discovery cohort.[13] In CKB, full proteogenomics summary statistics are available at the GWAS Catalog (accession codes pending) and on the CKB PheWeb (pheweb.ckbiobank.org).[14] All data used for this study is publicly available from the respective GWAS (Table S1).

Ethics Approval

The CKB complies with all the required ethical standards for medical research on human subjects. Ethical approvals were granted and maintained by the relevant institutional ethical research committees in the United Kingdom and China.

Study Population

The study population included CKB and UKB participants. CKB is a prospective cohort study of 512 724 adults aged from 30 to 79 years recruited from 10 geographically diverse areas across China from 2004 to 2008.[15,16] UKB is a prospective cohort of 503 317 adults aged from 40 to 70 years recruited from 2006 to 2010 across 22 assessment centers in the United Kingdom.[13] At the initial baseline visit, the participants in both CKB and UKB completed a questionnaire (interviewer-administered in CKB and touchscreen in UKB) on sociodemographic and lifestyle factors (eg, smoking, alcohol drinking, and physical activity), and medical history and medication (eg, statin), underwent physical measurements (eg, blood pressure, heart rate, height and weight, and waist and hip circumferences), and had biological samples (eg, nonfasting blood sample) taken by trained staff for long-term storage.

In both studies, prior ethical approvals were obtained, and all participants provided written informed consent.

Proteomics Assay

In CKB, the proteomics assay of 2923 unique proteins was conducted among 3977 participants, who had no prior history of cardiovascular diseases (CVD), and no use of lipid-lowering drugs (eg, statins) at the time of sample collection, but had genome-wide genotyping data available (see Table S2 for more details).[5] Stored baseline plasma samples from participants were retrieved, thawed, and subaliquoted to multiple aliquots, with 1 aliquot (100 µL) used to make 2 sets of 96-well plates (40 µL/well). These were then shipped on dry ice, the first set (batch 1) to the Olink Biosciences Laboratory, Uppsala, Sweden, and the second set (batch 2) to Olink Laboratory in Boston, USA, subsequently. Batch 1 covered 1463 unique proteins first released by Olink (Explore 1536 panel), while batch 2 covered a further 1460 unique proteins released subsequently by Olink (Explore Expansion panel).[17]

In UKB, measurements of the same 2923 proteins were generated using the Olink Explore 3072 panel in 54 306 participants selected for the UK Biobank Pharma Proteomics Project (see Table S3 for more details).[13] The retrieved plasma samples were transferred to the Olink Biosciences Laboratory, Uppsala, Sweden for measurements. The details of the UK Biobank Pharma Proteomics Project study design, sample selection, proteomics assay, and quality control have been previously described elsewhere.[13]

Discovery of pQTL for Proteins

The summary-level statistics of genetic associations with levels of 2923 circulating proteins were extracted from GWAS in CKB (East Asians [EAS]; n=3974)[14,18] and UKB (Europeans [EUR] ancestry only; n=34 557),[13] which identified pQTL, with the *cis*-pQTLs defined as variants within 500 Kbp (CKB) or 1 Mbp (UKB) of the encoded gene region that passed a conservative multiple-test-corrected threshold of $P<1.7\times10^{-11}$ (ie, $5\times10^{-8}/2923$ proteins). All other associations were categorized as *trans*-pQTLs. The *cis*-pQTLs were used in the present study as they are less prone to horizontal pleiotropy. To minimize bias from pleiotropy and ensure the robustness of selected instrumental variables, we only used the leading *cis*-pQTL (ie, lowest *P* value) for any proteins with >1 *cis*-pQTLs for subsequent analyses.

Global GWAS Consortia for Stroke and Stroke Risk Factors

To reduce confounding by population stratification, we extracted estimates for the associations of the protein instrumental variables with specific outcome measures only in EAS or EUR ancestry populations (see Table S1). The primary outcomes of the present study were IS (BioBank Japan in EAS: 22 664; GIGASTROKE [EUR only] in EUR: 62 100 cases) and any stroke (GIGASTROKE [EAS only] in EAS: 27 413; GIGASTROKE [EUR only] in EUR: 73 652 cases).[3,19] The secondary outcomes were 5 major risk factors for IS: that is SBP (EAS/EUR: 145 505/1006,863 participants),[19,20] LDL-C (146 500/1320,000 participants),[21] body mass index (173 430/700 000 participants), T2D (88 109/242 283 cases), and AF (4150/60 620 cases).

Statistical Analyses

We used 2-sample MR analyses to assess the associations between genetically predicted protein levels and study outcomes. The Wald ratio method was used to estimate odds ratios (ORs) and mean differences (95% CI) for binary (primary) and quantitative (secondary) outcomes, respectively, based on 1-SD higher levels of each protein. Colocalization analyses were performed to assess whether proteins and stroke have a shared signal at *cis*-pQTL, with fine-mapping performed using the SuSiE (v. 0.12.35) to identify 95% credible sets of likely causal variants.[22] For each *cis*-pQTLs, SuSiE was conducted using an internal CKB LD reference of 40K unrelated Chinese individuals or UKB LD reference of 40K unrelated white British individuals, with the maximum number of nonzero effects in the SuSiE set to 10, the minimum absolute correlation allowed in a credible set to 0.1, and the maximum number of iterations to 100 000. When the number of fine-mapped credible sets was found to be equal to 10, the maximum number of nonzero effects was set to 15. Colocalization using SuSiE output was performed using the coloc (v. 5.2.1) package in R. Two signals were considered as having strong evidence of colocalization if the posterior probability for shared causal variants (posterior probabilities for the H4 hypothesis [PP4]) was ≥0.7.

For proteins that were causally associated with both IS and SBP, we conducted a formal mediation analysis to assess if the total effects of proteins on IS were via SBP, both directly and indirectly, based on a 2-sample MR of proteins and IS. To decompose direct and indirect effects, we used results from 2-step MR and chose the Product method to estimate the beta of the indirect effect and the Delta method to estimate the SE.[23] For the MR of SBP and IS, we used the publicly available estimates from genetically determined SBP to IS in EAS[24] and EUR,[10] respectively. For proteins showing significant associations with stroke, a further 2-sample MR was conducted using leading *trans*-pQTL, with lookups in ancestry-specific stroke GWAS. Moreover, we screened protein expression data from the RNA database of Genotype-Tissue Expression (https://gtexportal.org/home/) to examine the tissue-specific role of the causal proteins in stroke. We further searched the Common Metabolic Diseases Knowledge Portal (https://hugeamp.org/) for associations of genes with available diseases (n=555) and phenotypes (n=642) using Human Genetic Evidence scores to evaluate the extent to which human genetic evidence supports the role of a gene in a disease or phenotype.[25] Additional

downstream analyses included single gene knock-out models (www.mousephenotype.org) and for those stroke-related proteins. To assess the potential druggability of identified proteins, we searched DrugBank and OpenTargets databases, for details of drug development.

All statistical analyses were performed using R, version 4.1.2. Benjamini-Hochberg's false discovery rate (5%) was used to correct for multiple testing.

RESULTS

Overall, GWAS in CKB and UKB identified *cis*-pQTL for 1159 and 1954 (1148 overlapped) proteins respectively (Figure 1), which were used for 2-sample MR and colocalization analyses.

In 2-sample MR of EAS (ie, CKB and BioBank Japan) involving 971 *cis*-pQTLs (F-statistic ≥30) with GWAS summary statistics of stroke, genetically determined levels of 2 proteins (ALDH2 [aldehyde dehydrogenase 2] and FGF5 [fibroblast growth factor 5]) were significantly associated with disease risk after correction for multiple testing (Figure 2A and 2B). The ORs per 1—SD higher levels of ALDH2 were 1.60 (95% CI, 1.39–1.84; *P*=8.2e—11) for IS and 1.36 (95% CI, 1.23–1.51; *P*=9.8e—9) for any stroke, while for FGF5, they were 1.14 (95% CI, 1.08–1.19; *P*=2.1e—7), and 1.15 (95% CI, 1.10–1.20; *P*=9.0e—11), respectively (Figure 3). In replication analyses of different EAS populations from GIGASTROKE consortia involving 935 *cis*-pQTLs, the same 2 proteins showed significant associations with stroke. Moreover, in colocalization analyses, there was strong evidence (PP4>0.95) of shared genetic variants between these 2 proteins and IS (Figure 3; Figure S1). The genetically determined levels of FGF5 and ALDH2 were positively associated with SBP and AF but not with LDL-C (Figure 4) and ALDH2 was also positively associated with T2D in EAS populations (Figure S2).

In 2-sample MR of UKB and GIGASTROKE-EUR involving 1455 *cis*-pQTLs (F-statistic ≥45), we

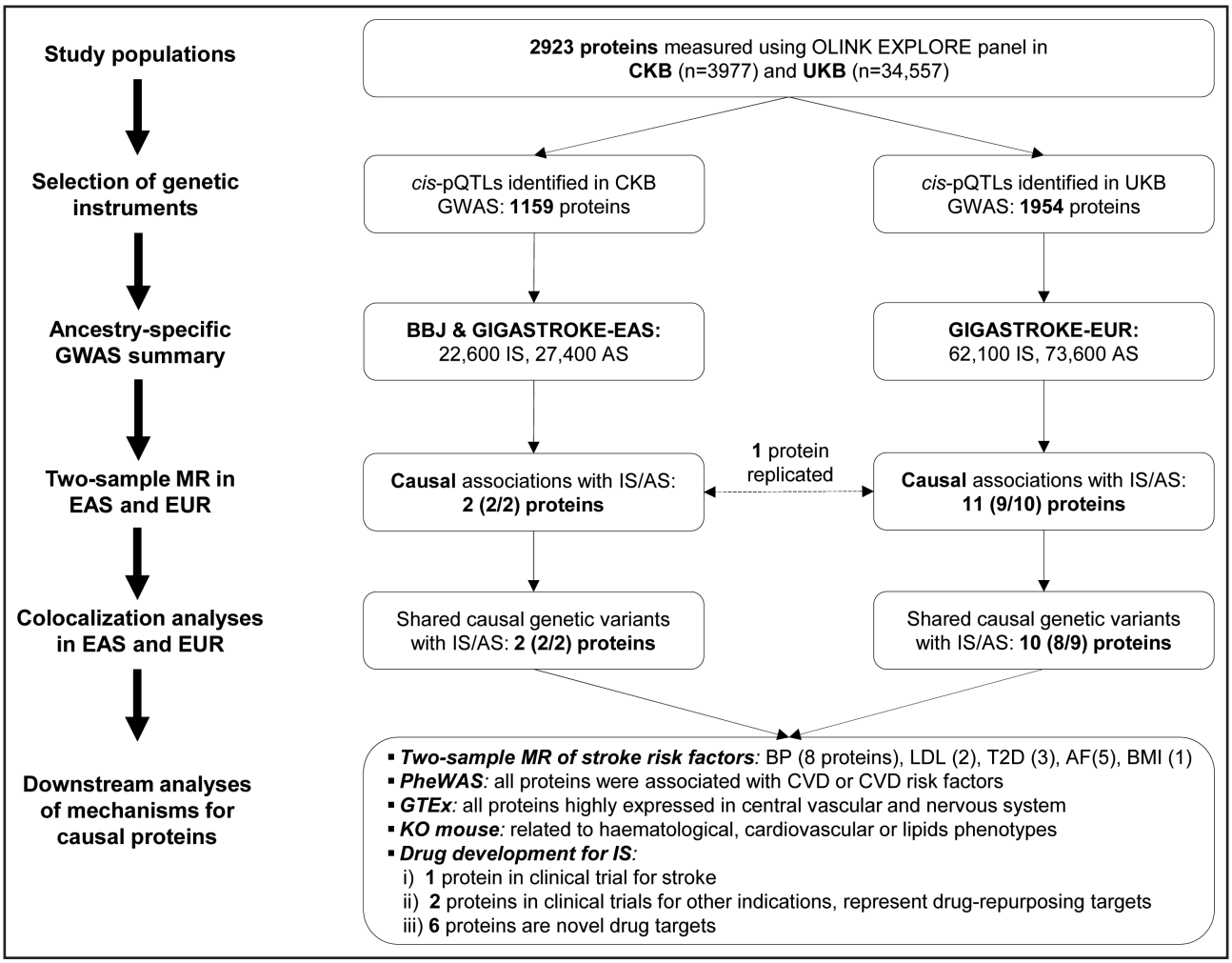

Figure 1. Overview of study design, analytic approaches, and key findings.
AF indicates atrial fibrillation; AS, any stroke; BBJ, BioBank Japan; BMI, body mass index; BP, blood pressure; CKB, China Kadoorie Biobank; EAS, East Asians; EUR, Europeans; GTEx, genotype-tissue expression; GWAS, genome-wide association study; IS, ischemic stroke; KO, knock-out; LDL, low-density lipoprotein; MR, Mendelian randomization; PheWAS, phenome-wide association study; pQTL, protein quantitative trait loci; T2D, type 2 diabetes; and UKB, UK Biobank.

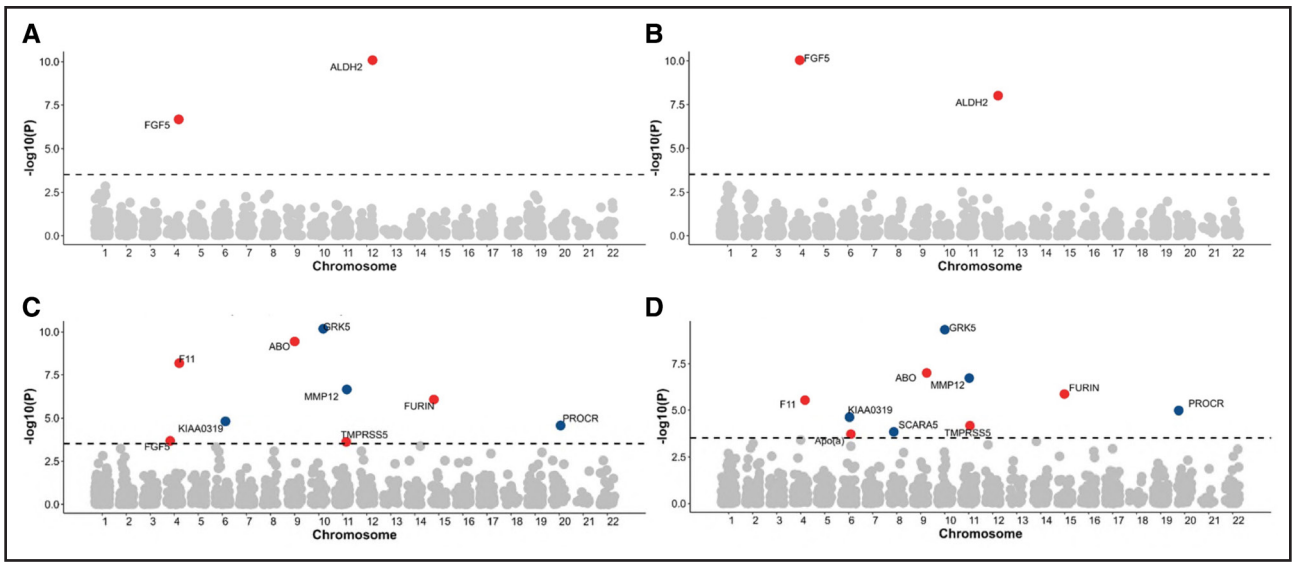

Figure 2. Associations of genetically predicted plasma levels of proteins with risk of stroke types in East Asians (EAS) and Europeans (EUR) populations.

The number of proteins included is 1159 in EAS and 1954 in EUR. Each point represents a single association test between a gene and stroke outcomes ordered by genomic position on the *x*-axis and the association strength on the *y*-axis as the $-\log_{10}(P)$. The horizontal line reflects the significant threshold of the false discovery rate $P<0.05$ and is set at the highest unadjusted *P* value that is below that threshold, with red points showing significant positive associations and blue points showing significant inverse associations, with names given for certain selected proteins.

identified a total of 11 putatively causal proteins (GRK5 [G protein-coupled receptor kinase 5], FURIN, KIAAA0319 [dyslexia-associated protein KIAA0319], PROCR [endothelial protein C receptor], TMPRSS5, F11, ABO, MMP12, FGF5, SCARA5 [scavenger receptor class A member 5], and Apo[a] [apolipoprotein(a)]) for IS (the first 9 proteins) or any stroke (10 proteins except FGF5), with the first 8 proteins overlapped between them and 1 (FGF5) overlapped with IS in EAS (Figure 2C and 2D). These resulted in a total of 12 putatively causal proteins for IS or any stroke in EAS or EUR. Of the 9 IS-associated proteins in EUR, 5 (FURIN, F11, FGF5, ABO, and TMPRSS5) showed positive associations, with ORs ranging from 1.04 (1.02–1.05) for TMPRSS5 to 1.23 (1.14–1.34) for FURIN, and 4 (GRK5, KIAA0319, PROCR5, and MMP12) showed inverse associations, with ORs ranging from 0.83 (0.78–0.87) for GRK5 to 0.93 (0.91–0.96) for MMP12 (Figure 3A). Of the 2 proteins associated only with any stroke after multiple-testing correction, 1 (Apo[a]) showed positive (1.05, 1.02–1.07) and 1 (SCARA5) showed inverse (0.92, 0.88–0.96) associations, with the remaining 8 showing similar, at least directionally, associations to those with IS (Figure 3B). In replication analyses using MEGASTROKE (EUR only), 6 out of 9 proteins (KIAA0319, MMP12, ABO, FURIN, GRK5, and TMPRSS5) were replicated for IS. All proteins (except those without *cis*-pQTL identified) shared the same direction with findings in EAS (Tables S4 and S5). Moreover, colocalization analyses provided strong support (PP4≥0.7) for shared genetic variants of 8 out of 9 proteins (except ABO with PP4 of 0.669) with IS, with 5 proteins (F11, MMP12, PROCR, KIAA0319, and GRK5) having PP4>0.95 (Figure 3; Figure S3).

Of these 11 stroke-associated proteins in EUR, 4 proteins (FGF5, FURIN, Apo[a], and TMPRSS5) were positive, and 3 (KIAA0319, PROCR, and SCARA5) were inversely associated with SBP. Moreover, 2 proteins (ABO and Apo[a]) were positively associated with LDL-C, 2 (ABO and MMP12) with T2D, 1 (PROCR) with body mass index, and 4 (Apo[a], FGF5, TMPRSS5, and MMP12) with AF (Figure 4; Figure S2f). These associations were all directionally consistent with their expected associations with stroke risk. The estimated proportion of the mediation effect of FGF5 via SBP was ≈77% in EUR but only 5% in EAS (Table S6). For FURIN, the mediation effect through SBP accounted for 41% of the total effect on IS. In contrast, the mediation effects via SBP were all small for other proteins, including KIAA0319 (8%), TMPRSS5 (6%), PROCR (1%), and ALDH2 (1%). Using leading *trans*-pQTL in MR analyses, we further replicated associations of 3 proteins (F11, MMP12, and SCARA5) with IS (Table S7).

Of the 10 IS-associated proteins in EAS or EUR, 4 (F11, ABO, MMP12, and TMPRSS5) were replicated in previous MR studies, 5 (PROCR, GRK5, FGF5, ALDH2, and FURIN) were associated with prioritized genetic signals in previous GWAS of stroke (Table). PheWAS analyses of these 10 proteins suggested that 8 proteins (ABO, F11, ALDH2, MMP12, PROCR, GRK5, FGF5, and FURIN) were associated with IS or other CVD outcomes (eg, MI), and the remaining 2 (KIAA0319 and TMPRSS5) were associated with

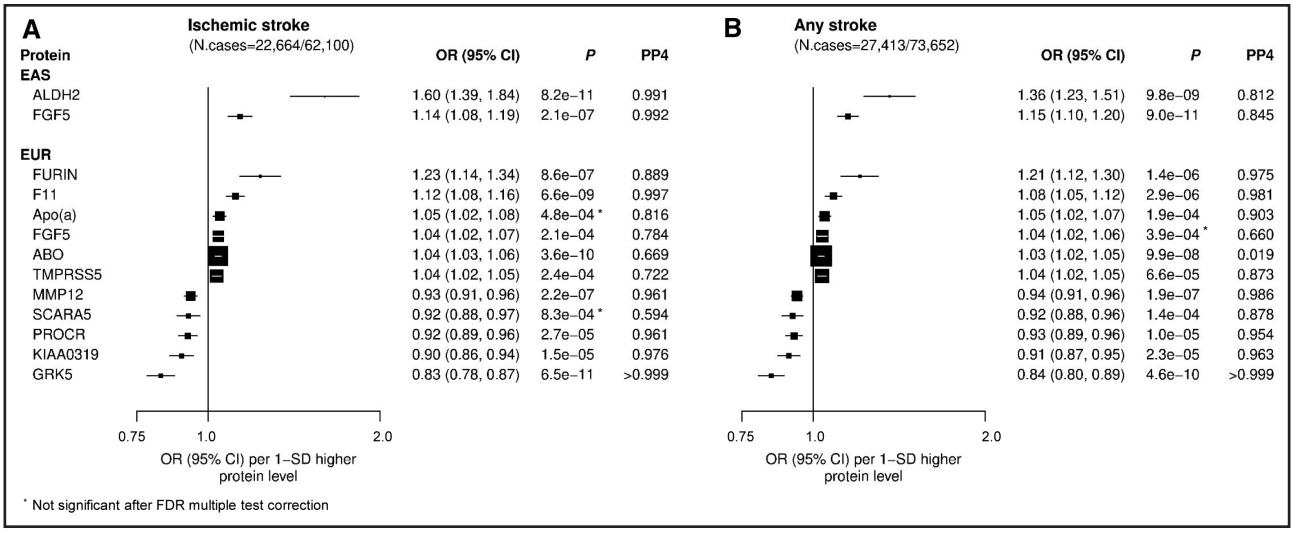

Figure 3. Adjusted odds ratios (ORs) per 1–SD higher genetically predicted levels of 12 proteins for risk of ischemic stroke and any stroke in East Asians (EAS) and Europeans (EUR) populations.
A, Ischemic stroke; (**B**) any stroke. The size of the squares is proportional to the inverse of the variance of the log ORs.

hematological (eg, platelet volume and count), hepatic (eg, alanine transaminase), renal (eg, estimated Glomerular Filtration Rate), and anthropometric (eg, height) phenotypes (Table S8). Moreover, ALDH2, ABO, and FURIN, all highly expressed in the central vascular or central nervous system, were associated with several lipid biomarkers (eg, triglyceride, total cholesterol, and high-density lipoprotein cholesterol). Furthermore, additional analyses of single-gene knock-out mouse models identified associations with several cardiovascular phenotypes (ALDH2, GRK5, and FURIN), hematopoietic phenotypes (F11, MMP12, KIAA0319, PROCR, and ALDH2), central nervous system related phenotypes (GRK5, MMP12, and KIAA0319) and integument phenotypes (ABO and FGF5). With the exception of ABO with poor druggability, of the 9 other proteins, there were no reports of associated drug targets and development for 6 proteins, while for the remaining 3 proteins (F11, ALDH2, and MMP12) with drug development, only 1 (F11) involved stroke and the remaining

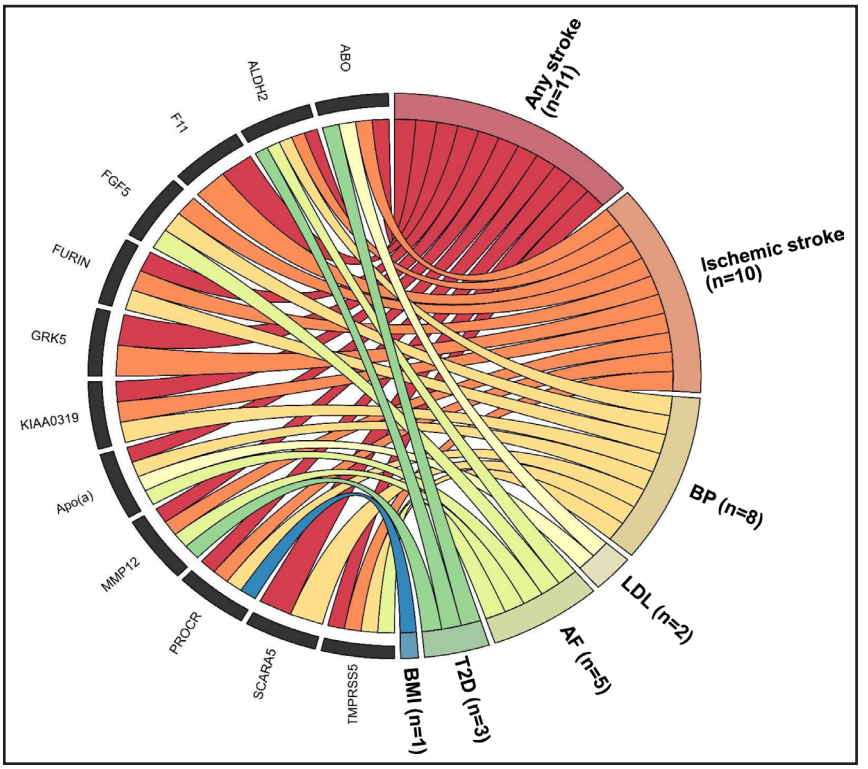

Figure 4. Chord diagram showing 12 potential causal proteins identified for risks of ischemic stroke (IS) and any stroke and their associations with 5 major stroke risk factors.
AF indicates atrial fibrillation; BMI, body mass index; BP, blood pressure; LDL, low-density lipoprotein; and T2D, type 2 diabetes.

Table. PheWAS Results of 10 Proteins Showing Genetic Associations With Ischemic Stroke in East Asians and Europeans, and Their Potential as Novel or Repurposing Drug Targets

Protein	Reported in previous		PheWAS associations (phenotypes)	Highly expressed tissues	KO mouse phenotype	Drug target development
	GWAS	MR				
Not druggable						
ABO (histo-blood group ABO system transferase)	√	√	Cardiovascular, hematological, Hepatic, glycemic, lipids and metabolic	Kidney, CVS, CNS	Integument, reproductive, and endocrine phenotype	…
In clinical trials for stroke						
F11 (coagulation factor XI)	√	√	Cardiovascular	CVS, Kidney, CNS	CNS, hematopoietic, and metabolism phenotype	Phase II–III
In clinical trials for other indications (repurpose)						
ALDH2 (aldehyde dehydrogenase)	√	…	Anthropometric, cardiovascular, hematological, hepatic, glycemic, renal, and lipids	CVS, kidney, liver	CVD, behavior/ neurological, and hematopoietic	Phases I–IV
MMP12 (macrophage metalloelastase)	√	√	Cardiovascular	Adipose, pancreas, CVS	Hematopoietic, CNS	Phase II–III
Novel drug targets						
PROCR (endothelial protein C receptor)	√	…	Cardiovascular, renal, and anthropometric	Kidney, adipose, CVS	Hematopoietic	…
GRK5 (G protein-coupled receptor kinase 5)	√	…	Hematological, anthropometric, and metabolic	CNS, CVS, adipose	CVD, CNS	…
FGF5 (fibroblast growth factor 5)	√	…	Anthropometric, cardiovascular, hepatic, renal and reproductive	CVS, CNS, Kidney	Integument	…
FURIN	√	…	Cardiovascular, hematological, hepatic, reproductive, and lipids	CVS, stem cell, endothelium	CVD	…
KIAA0319 (dyslexia-associated protein KIAA0319)	…	…	Hematological, immunologic, anthropometric, and hepatic	CNS, CVS, pancreas	Hematopoietic, CNS	…
TMPRSS5 (transmembrane protease serine 5)	…	√	Hematological, reproductive, hepatic, sleep, and circadian	CNS, CVS, kidney	…	…

ABO indicates histo-blood group ABO system transferase; CNS, central nervous system; CVD, cardiovascular diseases; CVS, cardiovascular system; GWAS, genome-wide association study; KO, knock-out; MR, Mendelian randomization; and PheWAS, phenome-wide association study.

2 involved alcohol dependence (ALDH2) and cancer (MMP12).

DISCUSSION

The present proteome-wide genetic analyses of 2923 proteins in EAS and EUR populations provided robust genetic evidence for the causal relevance of 12 proteins in the pathogenesis of IS (10 proteins) or any stroke (11 proteins). For 3 proteins (GRK5, F11, and MMP12), the associations with stroke were likely mediated through hematological traits, while for the other 9 proteins, they were likely mediated in part by major stroke risk factors, particularly SBP. PheWAS analyses and KO mouse models confirmed the importance of these proteins for stroke or stroke-related traits. Of these 10 IS-related proteins, 6 were novel, among which 5 proteins (PROCR, GRK5, FGF5, ALDH2, and FURIN) were further prioritized by GWAS signals as potential drug targets for stroke.

While previous MR analyses have identified several putatively causal proteins for stroke, these studies only involved European ancestry populations, with varying numbers of proteins assayed (n=41–4137) using different platforms and genetic instruments (n=308–653) identified from GWAS.[9–12] Nevertheless, of the 10 causal proteins identified for IS in the present study, 4 proteins (ABO, F11, MMP12, and TMPRSS5) were previously reported in MR studies, while for the remaining 6 proteins, 5 had GWAS support of their roles in IS, mainly in EUR populations (Table). Our study provided further evidence of their relevance for IS in East Asian populations and their potential as novel drug targets.

ABO is a glycosyltransferase enzyme whose activity determines ABO blood type. Consistent with our findings, previous observational studies suggested that non-O blood individuals (higher ABO enzymatic activity) had a higher risk of venous thromboembolism, IS and myocardial infarction, possibly mediated by higher levels of several clotting proteins (eg, FVIII [coagulation factor VIII] and von Willebrand Factor) in these individuals.[26] Our separate genetic analyses also demonstrated positive associations of ABO with risks of cardio-embolic or large artery subtypes of IS and levels of LDL-C and risk of T2D (Figure S2), for reasons that require further investigation. However, drugs targeting ABO may be not feasible, given the uncertain effect on blood type and potential side

effects across multiple tissues. F11 is a coagulation factor involved in thrombus propagation and stabilization,[27] whose inhibition has been hypothesized to reduce thrombosis without increasing bleeding tendency.[28] Drugs targeting F11 (F11 inhibitors such as abelacimab) are currently under investigation for stroke (URL: https://clinicaltrials.gov/; Unique identifiers: NCT04755283, NCT04304508, and NCT03766581),[29] and our results provide genetic support for the likely success of development programs. Moreover, our analyses also supported F11 as a possible target for the cardio-embolic subtype of IS (OR, 1.26 [1.16–1.37]; PP4=0.995). As for possible bleeding-adverse effects associated with lowering of F11 levels, recent studies did not find its causal association with intracerebral hemorrhage stroke risk in randomized trials of second-generation F11 lowering drugs (antisense oligonucleotide),[30] or MR study.[31]

MMP12 and ALDH2 represent potential drug-repurposing targets for the treatment of stroke. MMP12 is a member of the matrix metalloproteinase family that contributes to vascular remodeling and atherosclerosis, but to date, its relationship with stroke is unclear with conflicting evidence from observational and genetic studies. One population-based study involving 2983 participants with 450 incident IS cases found a significant positive association of MMP12 level, measured using the SomaScan platform, with IS risk (hazard ratio, 1.30 [1.16–1.45]; *P*=4.5e−6).[32] This contrasted with findings in the present study and all previous genetic studies using different assays, which showed an inverse relationship.[9–12] Moreover, our genetic analyses also found inverse associations of MMP12 with AF and T2D (Figure S2), 2 major risk factors for IS. In animal models of IS, *MMP12* gene expression in the brain increases dramatically, and its suppression reduces brain damage and promotes neurological, sensorimotor, and cognitive functional outcomes.[33] Consistent with PheWAS results (Table S8), there was also animal experimental evidence that genetic deletion of *MMP12* ameliorates cardiometabolic disease by improving insulin sensitivity, systemic inflammation, and atherosclerotic features in mice.[34] MMP12 is a drug target for cancer (breast or lung; phase III), hepatitis C infection (phase II), and chronic obstructive pulmonary disease (phase II), but is not currently being evaluated for stroke, thus highlighting potential drug-repurposing opportunities for stroke. ALDH2 is a mitochondrial enzyme involved in the detoxification of alcohol-derived acetaldehyde and endogenous aldehydes. The *ALDH2* locus had the strongest association signals with IS in EAS.[19] Although numerous studies have attributed an accumulation of aldehydes (secondary to alcohol consumption, ischemia, or elevated oxidative stress) to an increased risk of CVD, recent studies suggested BP regulation and risks of small vessel subtype of IS, AF, and T2D (Figure S2). However, we were unable to replicate the associations of ALDH2 and stroke in EUR due to a lack of *cis*-pQTL in UKB. ALDH2 inhibitors are currently being tested in >30 phases I–IV trials for various indications, including alcohol dependence, cocaine dependence, infection, and cancer, but not a stroke, thus representing its drug-repurposing opportunities for stroke.

To the best of our knowledge, 6 proteins (PROCR, GRK5, FGF5, KIAA0319, FURIN, and TMPRSS5) have not been extensively studied in stroke and represent novel drug targets. PROCR, also known as activated protein C receptor, is encoded by the *PROCR* gene and is a receptor for protein C (PROC) that enhances its activation and is an anti-coagulant serine protease activated by the blood coagulation pathway. Notably, there was no evidence of any drug development targeting PROCR, however, PROC is a drug target currently being evaluated in phase II trials (Unique identifier: NCT02222714) for treatment (PROC activator) of acute IS and is poised for an upcoming phase III trial.[35] PheWAS results confirmed its importance in CVD and kidney function (Table S8). Taken together, the available evidence provides strong support for PROCR as a potential treatment target (PROCR enhancer) for IS. In the present study, GRK5 was the protein most strongly associated with IS in UKB. In CKB, however, we did not find a significant association of GRK5 with IS, likely due to the universal presence of the rs10886430-A genotype (*cis*-pQTL in the present analyses) in EAS (>99.9%) compared with EUR (88%). GRK5, encoded by the *GRK5* gene, is a member of the GPCR kinase subfamily of the serine/threonine kinase family, which preferentially phosphorylates the activated forms of a variety of GPCRs.[36] GRK5 has also been implicated in the pathophysiology of atherosclerosis via modification of endothelial cell inflammation and desensitization of several cytokine and endothelin A and B receptors.[37] The *GRK5* locus was associated with IS and thrombosis in a cross-ancestry GWAS meta-analyses of >110 000 stroke cases,[3] but the underlying mechanisms remain to be properly elucidated. In platelet RNA-sequencing data, *GRK5* is by far the most expressed member of the GRK family, suggesting that GRK5 is likely to be the critical protein family member active in platelets.[38] Moreover, disruption of platelet *GRK5* expression by rs10886430-A (same *cis*-pQTL used in our study) is associated with reduced platelet reactivity, and MR analyses also support a causal role of rs10886430-A in decreasing the risk of stroke, pulmonary embolism, and venous thromboembolism through its effect on thrombin-induced platelet activity.[38] Hence, the available evidence provides strong support for GRK5 as a potential treatment target for IS and thrombosis, particularly in EUR.

FGF5 is a polypeptide growth factor that is highly expressed in central vascular and central nervous systems, and involved in biological processes including growth, development, neuronal functions, metabolism,

and angiogenesis.[39] In both EAS and EUR, we found a causal association of FGF5 with IS using different genetic variants. However, we found a somewhat greater effect size in CKB than in UKB (1.14 versus 1.04), which may partly reflect the differences in the leading variants used and their frequencies (UKB 28% versus CKB 42%). In animal experiments, *FGF5* upregulation was reported to increase myocardial blood flow and function, decrease myocyte apoptosis, and increase myocyte number.[40–42] Moreover, the *FGF5* locus was one of the strongest association signals with BP in both EAS and EUR.[19,20] Moreover, in addition to IS, our genetic analyses in both CKB and UKB also found a causal relevance of FGF5 for BP and AF (Figure S2), consistent with PheWAS findings. The directions of these associations were consistent with each other, indicating that the FGF5-stroke associations may be mediated partly through known stroke risk factors (eg, BP and AF). KIAA0319 is a transmembrane protein with a presumed role in neuronal migration. Several genetic association studies have implicated the *KIAA0319* gene in dyslexia susceptibility.[43] In recent animal studies, the *KIAA0319* gene is highly expressed in spinal cord neurons and spinal ganglion neurons, providing evidence that the KIAA0319 modulates axon growth and regeneration through Smad2 activation.[43] Therefore, KIAA0319 has the potential to be a treatment target to improve functional recovery following stroke via modulation of neurogenesis and neuroplasticity. Moreover, we found that genetically determined higher KIAA0319 levels were associated with lower SBP, which may partially explain the protein-stroke associations.

TMPRSS5 is a member of the type II transmembrane serine protease family,[44,45] but its function on CVD is poorly understood. Human TMPRSS5 mRNA is expressed in the brain and the protein is predominantly expressed in neurons, in their axons in the spinal cord.[45] A mouse model with mutant *TMPRSS5* had reduced proteolytic activity and suggested a role in hearing loss.[46] TMPRSS5 has been reported in a previous MR study to be associated with risks of IS and any stroke,[10] perhaps mediated by BP and AF, as supported by the present genetic analyses (Figure S2). Furthermore, PheWAS analysis revealed its association with certain hematological traits (eg, mean platelet volume, platelet count). Taken together, TMPRSS5 represents a potentially promising therapeutic target for IS, but further research is warranted to clarify the underlying mechanisms involved. FURIN (also known as PCSK3) was the most strongly IS-associated protein in EUR. It is a catalytic enzyme involved in the activation of proteins involved in inflammation and atherosclerosis.[47] We have previously reported a causal role of FURIN for IHD.[48] Moreover, we found that FURIN was also causally associated with BP, which may partially explain its role in stroke. However, we found no evidence of drug development for stroke (and IHD) related to FURIN.

SCARA5 and Apo[a] were 2 proteins that showed significant associations with any stroke but nonsignificant albeit directionally consistent association with IS after multiple testing corrections. SCARA5 is a scavenger receptor that exports ferritin-bound iron from circulation to parenchymal tissues.[49] Prior MR studies reported that lower levels of circulating iron were associated with a lower risk of stroke, which could partially explain the protective effect of SCARA5 observed in the present study. Notably, prior MR analyses also found that SCARA5 was inversely associated with both IS and subarachnoid hemorrhage,[9,12] suggesting SCARA5 represents a promising drug target for different stroke types. Apo(a), the major protein component of lipoprotein(a; Lp[a]), is a well-established causal biomarker for CVD.[50] Apo(a) is encoded by the *LPA* gene, which contains a 5.6-kb segment existing in multiple repeats (kringle IV-type 2 repeat polymorphism) that is responsible for the Apo(a) isoform variation.[50,51] Previous studies suggested that people with smaller Apo(a) isoforms are more effective at inhibiting thrombolysis, leading to a ≈2-fold higher risk of IS compared with those with larger protein isoforms.[51] Further studies are needed to determine whether smaller Apo(a) isoforms are relevant to vascular disease independent of Lp(a) concentration and other risk factors.

The present study has several strengths, including proteome-wide genetic analyses in both EAS and EUR populations, the use of the largest stroke GWAS consortia with replication involving different datasets, the use of ancestry-specific genetic instruments to assess causality, and multiple downstream analyses to assess possible mechanisms underlying these associations, including possible mediation of 2 major stroke risk factors (SBP, LDL-C). In addition to MR, we also undertook colocalization analyses to inform the causal inference. However, the present study also has limitations. First, despite being the largest to date in EAS, the number of proteins with genetic instruments in CKB was still modest compared with UKB, limiting its power to detect more significant associations, and thus only 1 protein was replicated for IS in ancestry-specific analyses though all proteins shared same direction. Second, the *cis*-pQTLs' definition differs between CKB (±500 Kbp) and UKB (±1 Mbp), but all IS-related *cis*-pQTLs identified in UKB were located within gene position (±500 Kbp), and only 7 more *cis*-pQTLs were identified in CKB using UKB criterion (±1 Mbp). Third, we only used lead *cis*-pQTLs as genetic instruments without considering other *cis*-pQTLs present for many proteins, which may limit the identification of novel associations. However, for the 10 IS-related proteins, each had only 1 *cis*-pQTL in CKB and UKB. Although the lead *cis*-pQTLs had high F-statistics (≥30), some exhibited low posterior inclusion probabilities, leading to uncertainty about their causal role at the locus. However, colocalization analyses supported the validity

of these genetic instruments. Moreover, although we used *trans*-pQTL to replicate certain results, we did not systematically explore the role of *trans*-pQTLs, which may provide further insights into stroke etiology. Fourth, ≈3.5% of participants overlapped between UKB and GIGASTROKE. Given this small proportion of overlap, the impact on our results in EUR is expected to be negligible. Additionally, in replication analyses using MEGASTROKE, which has no participant overlap with UKB, 6 out of 9 proteins were successfully replicated for IS, with the remaining 3 proteins showing nominal significance. In the genetic analysis of any stroke among EAS, there was a 24% participant overlap between CKB and GIGASTROKE. However, the findings were identical to those from the genetic analysis of IS involving CKB and BioBank Japan, which had no overlap. These replication results support the validity of our findings. Fifth, the colocalization analyses were only conducted among those significant proteins identified in 2-sample MR, which might miss some valid targets. However, our approach (ie, 2-sample MR followed by colocalization analyses) is consistent with previous studies of a similar nature.[10–12]

In summary, the present study identified 10 putatively causal proteins for IS, with 1 (ABO) not feasible as a drug target and 1 (F11) having a known drug target for IS. Of the remaining 8 proteins, 2 (MMP12, ALDH2) have drug targets that could potentially be repurposed for IS, and 6 were novel, including PROCR, GRK5, FGF5, and FURIN which could be prioritized for further in vitro and in vivo experiments and subsequent human studies to elucidate their mechanisms of action relevant for stroke. These findings demonstrated the potential of proteome-wide genetic analyses in different ancestry populations to discover novel or repurposing protein targets for stroke and many other diseases.

ARTICLE INFORMATION

Received October 16, 2025; final revision received March 18, 2025; accepted April 11, 2025.

Affiliations

Clinical Trial Service Unit & Epidemiological Studies Unit (P.Y., M.M., A.P., A.I., N.W., K.L., I.Y.M., H.F., C.K., Y.C., L.Y., H.D., D.A., D.S.-V., M.H., D.A.B., R.G.W., R.C., Z.C.) and Medical Research Council Health Research Unit (I.Y.M., H.F., C.K., Y.C., L.Y., H.D., D.A., D.S.-V., D.A.B., R.G.W., Z.C.), Nuffield Department of Population Health, University of Oxford, United Kingdom. Department of Epidemiology and Biostatistics, School of Public Health, Peking University Health Science Center, Beijing, China (D.S., J.L., C.Y., L.L.). Peking University Center for Public Health and Epidemic Preparedness & Response, Beijing, China (D.S., P.P., J.L., C.Y., L.L.). Key Laboratory of Epidemiology of Major Diseases (Peking University), Ministry of Education, Beijing, China (D.S., J.L., C.Y., L.L.).

Acknowledgments

The chief acknowledgment is to the participants, the project staff, and the Chinese Center for Disease Control and Prevention and its regional offices for assisting with the fieldwork and long-term follow-up of mortality and major morbidity. The authors thank regional health insurance bureaus for assisting with long-term follow-up of hospitalized outcomes and disease subphenotyping. Dr Yao and Z. Chen contributed to the concept and design of the study. Drs Yao and Pozarickij conducted statistical analyses and Dr Yao drafted the article. Dr Yao, Dr Mazidi, Dr Pozarickij, A. Iona, Dr Wright, Dr Lin, I.Y. Millwood, H. Fry, C. Kartsonaki, Y. Chen, Dr Yang, Dr Du, D. Avery, D. Schmidt-Valle, Dr Sun, P. Pei, Dr Lv, Dr Yu, Dr Hill, Dr Bennett, Dr Walters, Dr Li, R. Clarke, and Z. Chen were involved in the planning, acquisition, and interpretation of data. I.Y. Millwood, H. Fry, Y. Chen, D. Avery, D. Schmidt-Valle, P. Pei, and Dr Hill provided administrative, technical, or material support. All authors provided critical revisions of the article for important intellectual content. Dr Yao and Z. Chen are the guarantors of this work and take responsibility for the integrity and accuracy of the data analysis. Z. Chen supervised the work.

Data Access Statement

The CKB is a global resource for the investigation of lifestyle, environmental, blood biochemical, and genetic factors as determinants of common diseases. The CKB study group is committed to making the cohort data available to the scientific community in China, the United Kingdom, and worldwide to advance knowledge about the causes, prevention, and treatment of disease. In CKB, nongenetic data (eg, baseline, resurveys, biomarkers, and disease end points) are released periodically to bona fide researchers for open access. Details of the CKB Data Sharing Policy, data release schedules, and data request application procedures are available at https://www.ckbiobank.org/data-access. Accessing to individual participant genetic data (eg, genotyping, whole genome sequence) is currently constrained by China's Administrative Regulations on Human Genetic Resources, for which collaboration with CKB researchers is generally required, which may be subject to separate regulatory approvals in China (handled by CKB group) if it involves substantial sharing of unpublished data. Summary genetic statistics used in this study are publicly available from the relevant websites. For CKB, the summary stats for all proteins can be accessed through the CKB PheWeb browser at pheweb.ckbiobank.org.

Code Availability

Custom code was used in all statistical analyses in this report.

Sources of Funding

The China Kadoorie Biobank (CKB) baseline survey and the first resurvey were supported by the Kadoorie Charitable Foundation in Hong Kong. Long-term follow-up of the CKB study was supported by grants to Z. Chen at Oxford University from the Wellcome Trust (212946/Z/18/Z, 202922/Z/16/Z, 104085/Z/14/Z, and 088158/Z/09/Z) and grants to Dr Li from the National Key Research and Development Program of China (2016YFC0900500, 2016YFC0900501, 2016YFC0900504, and 2016YFC1303904) and the National Natural Science Foundation of China (91843302). DNA extraction and genotyping were supported by grants to Z. Chen from GlaxoSmithKline and the UK Medical Research Council (MC-PC-13049 and MC-PC-14135). The project was supported by core funding from the UK Medical Research Council (MC_UU_00017/1, MC_UU_12026/2, and MC_U137686851), Cancer Research UK (C16077/A29186 and C500/A16896) and the British Heart Foundation (CH/1996001/9454) to the Clinical Trial Service Unit and Epidemiological Studies Unit at Oxford University. Computation used the Oxford Biomedical Research Computing facility, a joint development between the Wellcome Centre for Human Genetics and the Big Data Institute supported by Health Data Research UK and the National Institute for Health and Care Research Oxford Biomedical Research Centre. Dr Yao is supported by an Intermediate Fellowship from the Nuffield Department of Population Health, University of Oxford.

Disclosures

None.

Supplemental Material

Tables S1–S8
Figures S1–S3
STROBE MR Checklist
Members of the China Kadoorie Biobank Collaborative Group

REFERENCES

1. Feigin VL, Stark BA, Johnson CO, Roth GA, Bisignano C, Abady GG, Abbasifard M, Abbasi-Kangevari M, Abd-Allah F, Abedi V, 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.* 2021;20:795–820. doi: 10.1016/S1474-4422(21)00252-0
2. Boehme AK, Esenwa C, Elkind MS. Stroke risk factors, genetics, and prevention. *Circ Res.* 2017;120:472–495. doi: 10.1161/CIRCRESAHA.116.308398
3. Mishra A, Malik R, Hachiya T, Jürgenson T, Namba S, Posner DC, Kamanu FK, Koido M, Le Grand Q, Shi M, et al; COMPASS Consortium.

Stroke genetics informs drug discovery and risk prediction across ancestries. *Nature*. 2022;611:115–123. doi: 10.1038/s41586-022-05165-3
4. Santos R, Ursu O, Gaulton A, Bento AP, Donadi RS, Bologa CG, Karlsson A, Al-Lazikani B, Hersey A, Oprea TI, et al. A comprehensive map of molecular drug targets. *Nat Rev Drug Discov*. 2017;16:19–34. doi: 10.1038/nrd.2016.230
5. Yao P, Iona A, Kartsonaki C, Said S, Wright N, Lin K, Pozarickij A, Millwood I, Fry H, Mazidi M, et al; China Kadoorie Biobank Collaborative Group. Conventional and genetic associations of adiposity with 1463 proteins in relatively lean Chinese adults. *Eur J Epidemiol*. 2023;38:1089–1103. doi: 10.1007/s10654-023-01038-9
6. Yao P, Iona A, Pozarickij A, Said S, Wright N, Lin K, Millwood I, Fry H, Kartsonaki C, Mazidi M, et al; China Kadoorie Biobank Collaborative Group. Proteomic analyses in diverse populations improved risk prediction and identified new drug targets for type 2 diabetes. *Diabetes Care*. 2024;47:1012–1019. doi: 10.2337/dc23-2145
7. Iona A, Yao P, Pozarickij A, Kartsonaki C, Said S, Wright N, Lin K, Millwood I, Fry H, Mazidi M, et al; China Kadoorie Biobank Collaborative Group. Proteogenomic analyses in relatively lean Chinese adults identify proteins and pathways that affect general and central adiposity levels. *Commun Biol*. 2024;7:1327. doi: 10.1038/s42003-024-06984-y
8. Zuber V, Grinberg NF, Gill D, Manipur I, Slob EAW, Patel A, Wallace C, Burgess S. Combining evidence from Mendelian randomization and colocalization: review and comparison of approaches. *Am J Hum Genet*. 2022;109:767–782. doi: 10.1016/j.ajhg.2022.04.001
9. Chong M, Sjaarda J, Pigeyre M, Mohammadi-Shemirani P, Lali R, Shoamanesh A, Gerstein HC, Paré G. Novel drug targets for ischemic stroke identified through Mendelian randomization analysis of the blood proteome. *Circulation*. 2019;140:819–830. doi: 10.1161/CIRCULATIONAHA.119.040180
10. Chen L, Peters JE, Prins B, Persyn E, Traylor M, Surendran P, Karthikeyan S, Yonova-Doing E, Di Angelantonio E, Roberts DJ, et al. Systematic Mendelian randomization using the human plasma proteome to discover potential therapeutic targets for stroke. *Nat Commun*. 2022;13:6143. doi: 10.1038/s41467-022-33675-1
11. Wu B-S, Chen S-F, Huang S-Y, Ou Y-N, Deng Y-T, Chen S-D, Dong Q, Yu J-T. Identifying causal genes for stroke via integrating the proteome and transcriptome from brain and blood. *J Transl Med*. 2022;20:181. doi: 10.1186/s12967-022-03377-9
12. Liu H, Zhang X, Ma H, Nguyen TN, Hua W, Mo S, Huang Q, Liu J, Zhou Y, Yang P. Identification of potential therapeutic targets for stroke and its subtypes by integrating proteomes and genetics from human plasma. *medRxiv*. 2023;2023.2008.2029.23294808. doi: 10.1101/2023.08.29.23294808
13. Sun BB, Chiou J, Traylor M, Benner C, Hsu Y-H, Richardson TG, Surendran P, Mahajan A, Robins C, Vasquez-Grinnell SG, et al; Alnylam Human Genetics. Plasma proteomic associations with genetics and health in the UK Biobank. *Nature*. 2023;622:329–338. doi: 10.1038/s41586-023-06592-6
14. Pozarickij A, Said S, Lin K, Morris S, Edris A, Iona A, Kartsonaki C, Wright N, Yao P, Fry H, et al. Ancestry diversity in the genetic determinants of the human plasma proteome enhances identification of potential drug targets. *medRxiv*. 2025;2023.2011.2013.23298365. doi: 10.1101/2023.11.13.23298365
15. Chen Z, Chen J, Collins R, Guo Y, Peto R, Wu F, Li L; China Kadoorie Biobank (CKB) collaborative group. China Kadoorie Biobank of 0.5 million people: survey methods, baseline characteristics and long-term follow-up. *Int J Epidemiol*. 2011;40:1652–1666. doi: 10.1093/ije/dyr120
16. Chen Z, Lee L, Chen J, Collins R, Wu F, Guo Y, Linksted P, Peto R. Cohort profile: the Kadoorie Study of Chronic Disease in China (KSCDC). *Int J Epidemiol*. 2005;34:1243–1249. doi: 10.1093/ije/dyi174
17. Assarsson E, Lundberg M, Holmquist G, Björkesten J, Thorsen SB, Ekman D, Eriksson A, Rennel Dickens E, Ohlsson S, Edfeldt G, et al. Homogenous 96-plex PEA immunoassay exhibiting high sensitivity, specificity, and excellent scalability. *PLoS One*. 2014;9:e95192. doi: 10.1371/journal.pone.0095192
18. Said S, Pozarickij A, Lin K, Morris S, Kartsonaki C, Wright N, Fry H, Chen Y, Du H, Bennett D, et al. Ancestry diversity in the genetic determinants of the human plasma proteome and associated new drug targets. *medRxiv*. 2023;2023.2011.2013.23298365. doi: 10.1101/2023.11.13.23298365
19. Sakaue S, Kanai M, Tanigawa Y, Karjalainen J, Kurki M, Koshiba S, Narita A, Konuma T, Yamamoto K, Akiyama M, et al; FinnGen. A cross-population atlas of genetic associations for 220 human phenotypes. *Nat Genet*. 2021;53:1415–1424. doi: 10.1038/s41588-021-00931-x
20. Evangelou E, Warren HR, Mosen-Ansorena D, Mifsud B, Pazoki R, Gao H, Ntritsos G, Dimou N, Cabrera CP, Karaman I, et al; Million Veteran Program. Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. *Nat Genet*. 2018;50:1412–1425. doi: 10.1038/s41588-018-0205-x
21. Graham SE, Clarke SL, Wu K-HH, Kanoni S, Zajac GJM, Ramdas S, Surakka I, Ntalla I, Vedantam S, Winkler TW, et al; VA Million Veteran Program. The power of genetic diversity in genome-wide association studies of lipids. *Nature*. 2021;600:675–679. doi: 10.1038/s41586-021-04064-3
22. Wang G, Sarkar A, Carbonetto P, Stephens M. A simple new approach to variable selection in regression, with application to genetic fine mapping. *J R Stat Soc Series B Stat Methodol*. 2020;82:1273–1300. doi: 10.1111/rssb.12388
23. Carter AR, Sanderson E, Hammerton G, Richmond RC, Davey Smith G, Heron J, Taylor AE, Davies NM, Howe LD. Mendelian randomisation for mediation analysis: current methods and challenges for implementation. *Eur J Epidemiol*. 2021;36:465–478. doi: 10.1007/s10654-021-00757-1
24. Clarke R, Wright N, Walters R, Gan W, Guo Y, Millwood IY, Yang L, Chen Y, Lewington S, Lv J, et al; China Kadoorie Biobank Collaborative Group‡. Genetically predicted differences in systolic blood pressure and risk of cardiovascular and noncardiovascular diseases: a Mendelian randomization study in Chinese adults. *Hypertension*. 2023;80:566–576. doi: 10.1161/HYPERTENSIONAHA.122.20120
25. Costanzo MC, von Grotthuss M, Massung J, Jang D, Caulkins L, Koesterer R, Gilbert C, Welch RP, Kudtarkar P, Hoang Q, et al; AMP-T2D Consortium. The type 2 diabetes knowledge portal: an open access genetic resource dedicated to type 2 diabetes and related traits. *Cell Metab*. 2023;35:695–710.e6. doi: 10.1016/j.cmet.2023.03.001
26. Lilova Z, Hassan F, Riaz M, Ironside J, Ken-Dror G, Han T, Sharma P. Blood group and ischemic stroke, myocardial infarction, and peripheral vascular disease: a meta-analysis of over 145,000 cases and 2,000,000 controls. *J Stroke Cerebrovasc Dis*. 2023;32:107215. doi: 10.1016/j.jstrokecerebrovasdis.2023.107215
27. De Caterina R, Prisco D, Eikelboom JW. Factor XI inhibitors: cardiovascular perspectives. *Eur Heart J*. 2022;44:280–292. doi: 10.1093/eurheartj/ehac464
28. Zhang H, Löwenberg EC, Crosby JR, MacLeod AR, Zhao C, Gao D, Black C, Revenko AS, Meijers JC, Stroes ES, et al. Inhibition of the intrinsic coagulation pathway factor XI by antisense oligonucleotides: a novel antithrombotic strategy with lowered bleeding risk. *Blood*. 2010;116:4684–4692. doi: 10.1182/blood-2010-04-277798
29. Chan NC, Weitz JI. New therapeutic targets for the prevention and treatment of venous thromboembolism with a focus on factor XI inhibitors. *Arterioscler Thromb Vasc Biol*. 2023;43:1755–1763. doi: 10.1161/ATVBAHA.123.318781
30. Büller HR, Bethune C, Bhanot S, Gailani D, Monia BP, Raskob GE, Segers A, Verhamme P, Weitz JI; FXI-ASO TKA Investigators. Factor XI antisense oligonucleotide for prevention of venous thrombosis. *N Engl J Med*. 2014;372:232–240. doi: 10.1056/NEJMoa1405760
31. Gill D, Georgakis MK, Laffan M, Sabater-Lleal M, Malik R, Tzoulaki I, Veltkamp R, Dehghan A. Genetically determined FXI (Factor XI) levels and risk of stroke. *Stroke*. 2018;49:2761–2763. doi: 10.1161/STROKEAHA.118.022792
32. Kalani R, Bartz TM, Psaty BM, Elkind MSV, Floyd JS, Gerszten RE, Shojaie A, Heckbert SR, Bis JC, Austin TR, et al. Plasma proteomic associations with incident ischemic stroke in older adults the cardiovascular health study. *Neurology*. 2023;100:e2182–e2190. doi: 10.1212/WNL.0000000000207242
33. Veeravalli KK. Implications of MMP-12 in the pathophysiology of ischaemic stroke. *Stroke Vasc Neurol*. 2023;9:97–107. doi: 10.1136/svn-2023-002363
34. Amor M, Bianco V, Buerger M, Lechleitner M, Vujić N, Dobrijević A, Akhmetshina A, Pirchheim A, Schwarz B, Pessentheiner AR, et al. Genetic deletion of MMP12 ameliorates cardiometabolic disease by improving insulin sensitivity, systemic inflammation, and atherosclerotic features in mice. *Cardiovasc Diabetol*. 2023;22:327. doi: 10.1186/s12933-023-02064-3
35. Lyden P, Pryor KE, Coffey CS, Cudkowicz M, Conwit R, Jadhav A, Sawyer RN Jr, Claassen J, Adeoye O, Song S, et al; NeuroNEXT Clinical Trials Network NN104 Investigators. Final results of the RHAPSODY trial: a multi-center, phase 2 trial using a continual reassessment method to determine the safety and tolerability of 3K3A-APC, a recombinant variant of human activated protein C, in combination with tissue plasminogen activator, mechanical thrombectomy or both in moderate to severe acute ischemic stroke. *Ann Neurol*. 2019;85:125–136. doi: 10.1002/ana.25383
36. Johnson LR, Scott MG, Pitcher JA. G protein-coupled receptor kinase 5 contains a DNA-binding nuclear localization sequence. *Mol Cell Biol*. 2004;24:10169–10179. doi: 10.1128/MCB.24.23.10169-10179.2004
37. Wu J-H, Zhang L, Fanaroff AC, Cai X, Sharma KC, Brian L, Exum ST, Shenoy SK, Peppel K, Freedman NJ. G protein–coupled receptor kinase-5 attenuates atherosclerosis by regulating receptor tyrosine kinases and

7-transmembrane receptors. *Arterioscler Thromb Vasc Biol*. 2012;32:308–316. doi: 10.1161/ATVBAHA.111.239608
38. Rodriguez BAT, Bhan A, Beswick A, Elwood PC, Niiranen TJ, Salomaa V, Trégouët DA, Morange PE, Civelek M, Ben-Shlomo Y, et al; FinnGen Study. A platelet function modulator of thrombin activation is causally linked to cardiovascular disease and affects PAR4 receptor signaling. *Am J Hum Genet*. 2020;107:211–221. doi: 10.1016/j.ajhg.2020.06.008
39. Itoh N, Ornitz DM. Fibroblast growth factors: from molecular evolution to roles in development, metabolism and disease. *J Biochem*. 2011;149:121–130. doi: 10.1093/jb/mvq121
40. Giordano FJ, Ping P, McKirnan MD, Nozaki S, Demaria AN, Dillmann WH, Mathieu-Costello O, Hammond HK. Intracoronary gene transfer of fibroblast growth factor–5 increases blood flow and contractile function in an ischemic region of the heart. *Nat Med*. 1996;2:534–539. doi: 10.1038/nm0596-534
41. Lynch P, Lee TC, Fallavollita JA, Canty JM Jr, Suzuki G. Intracoronary administration of AdvFGF-5 (fibroblast growth factor-5) ameliorates left ventricular dysfunction and prevents myocyte loss in swine with developing collaterals and ischemic cardiomyopathy. *Circulation*. 2007;116:I-71–I-76. doi: 10.1161/CIRCULATIONAHA.106.681866
42. Domouzoglou EM, Naka KK, Vlahos AP, Papafaklis MI, Michalis LK, Tsatsoulis A, Maratos-Flier E. Fibroblast growth factors in cardiovascular disease: the emerging role of FGF21. *Am J Physiol Heart Circ Physiol*. 2015;309:H1029–H1038. doi: 10.1152/ajpheart.00527.2015
43. Franquinho F, Nogueira-Rodrigues J, Duarte JM, Esteves SS, Carter-Su C, Monaco AP, Molnár Z, Velayos-Baeza A, Brites P, Sousa MM. The dyslexia-susceptibility protein KIAA0319 inhibits axon growth through Smad2 signaling. *Cereb Cortex*. 2017;27:1732–1747. doi: 10.1093/cercor/bhx023
44. Wang H, Davison M, Wang K, Xia T-H, Kramer M, Call K, Luo J, Wu X, Zuccarino R, Bacon C, et al. Transmembrane protease serine 5: a novel Schwann cell plasma marker for CMT1A. *Ann Clin Transl Neurol*. 2020;7:69–82. doi: 10.1002/acn3.50965
45. Yamaguchi N, Okui A, Yamada T, Nakazato H, Mitsui S. Spinesin/TMPRSS5, a novel transmembrane serine protease, cloned from human spinal cord. *J Biol Chem*. 2002;277:6806–6812. doi: 10.1074/jbc.M103645200
46. Guipponi M, Toh MY, Tan J, Park D, Hanson K, Ballana E, Kwong D, Cannon PZ, Wu Q, Gout A, et al. An integrated genetic and functional analysis of the role of type II transmembrane serine proteases (TMPRSSs) in hearing loss. *Hum Mutat*. 2008;29:130–141. doi: 10.1002/humu.20617
47. Liu ZW, Ma Q, Liu J, Li JW, Chen YD. The association between plasma furin and cardiovascular events after acute myocardial infarction. *BMC Cardiovasc Disord*. 2021;21:468. doi: 10.1186/s12872-021-02029-y
48. Fry H, Mazidi M, Kartsonaki C, Clarke R, Walters RG, Chen Z, Millwood IY. The role of furin and its therapeutic potential in cardiovascular disease risk. *Int J Mol Sci*. 2024;25:9237. doi: 10.3390/ijms25179237
49. Li JY, Paragas N, Ned RM, Qiu A, Viltard M, Leete T, Drexler IR, Chen X, Sanna-Cherchi S, Mohammed F, et al. Scara5 is a ferritin receptor mediating non-transferrin iron delivery. *Dev Cell*. 2009;16:35–46. doi: 10.1016/j.devcel.2008.12.002
50. Helgadottir A, Gretarsdottir S, Thorleifsson G, Holm H, Patel RS, Gudnason T, Jones GT, van Rij AM, Eapen DJ, Baas AF, et al. Apolipoprotein(a) genetic sequence variants associated with systemic atherosclerosis and coronary atherosclerotic burden but not with venous thromboembolism. *J Am Coll Cardiol*. 2012;60:722–729. doi: 10.1016/j.jacc.2012.01.078
51. Erqou S, Thompson A, Di Angelantonio E, Saleheen D, Kaptoge S, Marcovina S, Danesh J. Apolipoprotein(a) isoforms and the risk of vascular disease: systematic review of 40 studies involving 58,000 participants. *J Am Coll Cardiol*. 2010;55:2160–2167. doi: 10.1016/j.jacc.2009.10.080