Proteome-wide genetic study in East Asians and Europeans identified multiple therapeutic targets for ischemic stroke

Supplemental Material

Members of the China Kadoorie Biobank collaborative group

International Steering Committee: Junshi Chen, Zhengming Chen (PI), Robert Clarke, Rory Collins, Liming Li (PI), Chen Wang, Jun Lv, Richard Peto, Robin Walters.

International Co-ordinating Centre, Oxford: Daniel Avery, Derrick Bennett, Ruth Boxall, Sushila Burgess, Ka Hung Chan, Yiping Chen, Zhengming Chen, Johnathan Clarke, Robert Clarke, Huaidong Du, Ahmed Edris Mohamed, Hannah Fry, Simon Gilbert, Pek Kei Im, Andri Iona, Maria Kakkoura, Christiana Kartsonaki, Hubert Lam, Kuang Lin, James Liu, Mohsen Mazidi, Iona Millwood, Sam Morris, Qunhua Nie, Alfred Pozarickij, Paul Ryder, Saredo Said, Dan Schmidt, Becky Stevens, Iain Turnbull, Robin Walters, Baihan Wang, Lin Wang, Neil Wright, Ling Yang, Xiaoming Yang, Pang Yao.

National Co-ordinating Centre, Beijing: Xiao Han, Can Hou, Qingmei Xia, Chao Liu, Jun Lv, Pei Pei, Dianjanyi Sun, Canqing Yu.

10 Regional Co-ordinating Centres:

Guangxi Provincial CDC: Naying Chen, Duo Liu, Zhenzhu Tang. Liuzhou CDC: Ningyu Chen, Qilian Jiang, Jian Lan, Mingqiang Li, Yun Liu, Fanwen Meng, Jinhuai Meng, Rong Pan, Yulu Qin, Ping Wang, Sisi Wang, Liuping Wei, Liyuan Zhou. **Gansu** Provincial CDC: Caixia Dong, Pengfei Ge, Xiaolan Ren. Maiji CDC: Zhongxiao Li, Enke Mao, Tao Wang, Hui Zhang, Xi Zhang. **Hainan** Provincial CDC: Jinyan Chen, Ximin Hu, Xiaohuan Wang. Meilan CDC: Zhendong Guo, Huimei Li, Yilei Li, Min Weng, Shukuan Wu. **Heilongjiang** Provincial CDC: Shichun Yan, Mingyuan Zou, Xue Zhou. Nangang CDC: Ziyan Guo, Quan Kang, Yanjie Li, Bo Yu, Qinai Xu. **Henan** Provincial CDC: Liang Chang, Lei Fan, Shixian Feng, Ding Zhang, Gang Zhou. Huixian CDC: Yulian Gao, Tianyou He, Pan He, Chen Hu, Huarong Sun, Xukui Zhang. **Hunan** Provincial CDC: Biyun Chen, Zhongxi Fu, Yuelong Huang, Huilin Liu, Qiaohua Xu, Li Yin. Liuyang CDC: Huajun Long, Xin Xu, Hao Zhang, Libo Zhang. **Jiangsu** Provincial CDC: Jian Su, Ran Tao, Ming Wu, Jie Yang, Jinyi Zhou, Yonglin Zhou. Suzhou CDC: Yihe Hu, Yujie Hua, Jianrong Jin, Fang Liu, Jingchao Liu, Yan Lu, Liangcai Ma, Aiyu Tang, Jun Zhang. **Qingdao** CDC: Liang Cheng, Ranran Du, Ruqin Gao, Feifei Li, Shanpeng Li, Yongmei Liu, Feng Ning, Zengchang Pang, Xiaohui Sun, Xiaocao Tian, Shaojie Wang, Yaoming Zhai, Hua Zhang, Licang CDC: Wei Hou, Silu Lv, Junzheng Wang. **Sichuan** Provincial CDC: Xiaofang Chen, Xianping Wu, Ningmei Zhang, Weiwei Zhou. Pengzhou CDC: Xiaofang Chen, Jianguo Li, Jiaqiu Liu, Guojin Luo, Qiang Sun, Xunfu Zhong. **Zhejiang** Provincial CDC: Weiwei Gong, Ruying Hu, Hao Wang,Meng Wang, Min Yu. Tongxiang CDC: Lingli Chen, Qijun Gu, Dongxia Pan, Chunmei Wang, Kaixu Xie, Xiaoyi Zhang.

STOBE MR Checklist

Item No.	Section	Checklist item	Page No.	Relevant text from manuscript
1	**TITLE and ABSTRACT**	Indicate Mendelian randomization (MR) as the study's design in the title and/or the abstract if that is a main purpose of the study	2	Included in abstract methods section.
	INTRODUCTION			
2	**Background**	Explain the scientific background and rationale for the reported study. What is the exposure? Is a potential causal relationship between exposure and outcome plausible? Justify why MR is a helpful method to address the study question	5	MR can help to clarify the causal relevance of traits (e.g., circulating proteins) and diseases (e.g. stroke)
3	**Objectives**	State specific objectives clearly, including pre-specified causal hypotheses (if any). State that MR is a method that, under specific assumptions, intends to estimate causal effects	6, 9	Study aims outlined at the end of the introduction. MR estimating causal effects mentioned in the methods.
	METHODS			
4	**Study design and data sources**	Present key elements of the study design early in the article. Consider including a table listing sources of data for all phases of the study. For each data source contributing to the analysis, describe the following:		
	a)	Setting: Describe the study design and the underlying population, if possible. Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection, when available.	6-9	Populations used for two-sample MR outlined in methods section.
	b)	Participants: Give the eligibility criteria, and the sources and methods of selection of participants. Report the sample size, and whether any power or sample size calculations were carried out prior to the main analysis	6-8	Details of eligibility criteria, sample sizes of two-sample MR datasets outlined in the methods.
	c)	Describe measurement, quality control and selection of genetic variants	8	Genetic instrument for proteins (*cis*-pQTL) in the methods.
	d)	For each exposure, outcome, and other relevant variables, describe methods of assessment and diagnostic criteria for diseases	7-8	Proteomics assay, stroke types (any or ischemic) and stroke risk factors
	e)	Provide details of ethics committee approval and participant informed consent, if relevant	6-7	Ethics statement included in the methods
5	**Assumptions**	Explicitly state the three core IV assumptions for the main analysis (relevance, independence and exclusion restriction) as well assumptions for any additional or sensitivity analysis	8	cis-pQTL
6	**Statistical methods: main**	Describe statistical methods and statistics used		

	analysis				
		a)	Describe how quantitative variables were handled in the analyses (i.e., scale, units, model)	9	Outlined for two-sample MR analyses in the statistical analysis section
		b)	Describe how genetic variants were handled in the analyses and, if applicable, how their weights were selected	8	Leading cis-pQTL was selected
		c)	Describe the MR estimator (e.g. two-stage least squares, Wald ratio) and related statistics. Detail the included covariates and, in case of two-sample MR, whether the same covariate set was used for adjustment in the two samples	9	Outlined in statistical analysis section.
		d)	Explain how missing data were addressed	NA	
		e)	If applicable, indicate how multiple testing was addressed	10	Benjamini-Hochberg FDR (5%)
7	**Assessment of assumptions**		Describe any methods or prior knowledge used to assess the assumptions or justify their validity	8, 11	Use of cis-Pqtl definition and F statistics outlined.
8	**Sensitivity analyses and additional analyses**		Describe any sensitivity analyses or additional analyses performed (e.g. comparison of effect estimates from different approaches, independent replication, bias analytic techniques, validation of instruments, simulations)	9,	Use of colocalization analysis, trans-pQTL MR
9	**Software and pre-registration**				
		a)	Name statistical software and package(s), including version and settings used	10	
		b)	State whether the study protocol and details were pre-registered (as well as when and where)	NA	
	RESULTS				
10	**Descriptive data**				
		a)	Report the numbers of individuals at each stage of included studies and reasons for exclusion. Consider use of a flow diagram		Figure 1 – flow chart
		b)	Report summary statistics for phenotypic exposure(s), outcome(s), and other relevant variables (e.g. means, SDs, proportions)		Baseline characteristics of CKB and UKB in supplementary material
		c)	If the data sources include meta-analyses of previous studies, provide the assessments of heterogeneity across these studies	NA	
		d)	For two-sample MR: i. Provide justification of the similarity of the genetic variant-exposure associations	19	The limitations includes details sample overlap

			between the exposure and outcome samples ii. Provide information on the number of individuals who overlap between the exposure and outcome studies		
11	**Main results**				
		a)	Report the associations between genetic variant and exposure, and between genetic variant and outcome, preferably on an interpretable scale		Table S4
		b)	Report MR estimates of the relationship between exposure and outcome, and the measures of uncertainty from the MR analysis, on an interpretable scale, such as odds ratio or relative risk per SD difference	10-12	Results and Figure 3
		c)	If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	NA	
		d)	Consider plots to visualize results (e.g. forest plot, scatterplot of associations between genetic variants and outcome versus between genetic variants and exposure)		Figure 3, Table S5
12	**Assessment of assumptions**				
		a)	Report the assessment of the validity of the assumptions	10, 11	Assessments outlined in methods: F test for assumption 1 validation. Nature of *cis*-pQTL
		b)	Report any additional statistics (e.g., assessments of heterogeneity across genetic variants, such as I^2, Q statistic or E-value)	NA	
13	**Sensitivity analyses and additional analyses**				
		a)	Report any sensitivity analyses to assess the robustness of the main results to violations of the assumptions	10, 11	Assessments outlined in methods: F test for assumption 1 validation. Nature of cis-pQTL
		b)	Report results from other sensitivity analyses or additional analyses	11	Use other available stroke consortium with relative less number of stroke cases (MEGASTROKE)
		c)	Report any assessment of direction of causal relationship (e.g., bidirectional MR)		
		d)	When relevant, report and compare with estimates from non-MR analyses	14	Observational associations of MMP12

		e)	Consider additional plots to visualize results (e.g., leave-one-out analyses)		
	DISCUSSION				
14	**Key results**		Summarize key results with reference to study objectives	13	Discussion opening paragraph.
15	**Limitations**		Discuss limitations of the study, taking into account the validity of the IV assumptions, other sources of potential bias, and imprecision. Discuss both direction and magnitude of any potential bias and any efforts to address them	19-20	Limitations section.
16	**Interpretation**				
		a)	Meaning: Give a cautious overall interpretation of results in the context of their limitations and in comparison with other studies	13-18	Discussion
		b)	Mechanism: Discuss underlying biological mechanisms that could drive a potential causal relationship between the investigated exposure and the outcome, and whether the gene-environment equivalence assumption is reasonable. Use causal language carefully, clarifying that IV estimates may provide causal effects only under certain assumptions	13-18	Discussion
		c)	Clinical relevance: Discuss whether the results have clinical or public policy relevance, and to what extent they inform effect sizes of possible interventions	13-18	Drug targets development information
17	**Generalizability**		Discuss the generalizability of the study results (a) to other populations, (b) across other exposure periods/timings, and (c) across other levels of exposure	18	Ancestry-specific genetic analysis
	OTHER INFORMATION				
18	**Funding**		Describe sources of funding and the role of funders in the present study and, if applicable, sources of funding for the databases and original study or studies on which the present study is based	21	Role of funders section in methods
19	**Data and data sharing**		Provide the data used to perform all analyses or report where and how the data can be accessed, and reference these sources in the article. Provide the statistical code needed to reproduce the results in the article, or report whether the code is publicly accessible and if so, where	6, 21	Data Sharing Statement
20	**Conflicts of Interest**		All authors should declare all potential conflicts of interest	21	Disclosure section

1. Skrivankova VW, Richmond RC, Woolf BAR, Yarmolinsky J, Davies NM, Swanson SA, et al. Strengthening the Reporting of Observational Studies in Epidemiology using Mendelian Randomization (STROBE-MR) Statement. JAMA. 2021;under review.
2. Skrivankova VW, Richmond RC, Woolf BAR, Davies NM, Swanson SA, VanderWeele TJ, et al. Strengthening the Reporting of Observational Studies in Epidemiology using Mendelian Randomisation (STROBE-MR): Explanation and Elaboration. BMJ. 2021;375:n2233.

Table S1. Data sources for the two-sample MR analyses of stroke and stroke risk factors

Phenotype	EAS	EUR
OLINK proteomics (n=2923)	CKB (3977)	UKB (34,557)
Stroke types		
IS	BBJ (22,664/152,022) GIGA-EAS (19,032/237,242)	GIGA-EUR (62,100/1,234,808) MEGA-EUR (34,217/406,111)
AS	GIGA-EAS (27,413/237,242)	GIGA-EUR (73,652/1,234,808)
CES	GIGA-EAS (926/237,242)	GIGA-EUR (10804/1,234,808)
LAS	GIGA-EAS (1735/237,242)	GIGA-EUR (6399/1,234,808)
SVS	GIGA-EAS (5532/237,242)	GIGA-EUR (6811/1,234,808)
Stroke risk factors		
SBP	BBJ (145,505)	ICBP (1,006,863)
LDL-C	GLGC (146,500)	GLGC (1,320,000)
T2D	T2DGGI (88,109/339,395)	T2DGGI (242,283/1,569,734)
BMI	BBJ (173,430)	~700,000
AF	BBJ (4150/155,540)	60620/970,216

Sample size shown as a total number for quantitative traits and cases/controls for binary traits.
Abbreviations: AS=any stroke; EAS=East Asian; EUR=European; IS=ischemic stroke; SBP=Systolic blood pressure
BBJ=Biobank Japan; ICBP=International Consortium for Blood Pressure; GLGC=Global Lipids Genetic Consortium

Table S2. Baseline characteristics of participants in proteomic subset and genotyped cohort in CKB

Characteristics [a]	Proteomic subset (n=3977)	Genotyped CKB (n=79,159)
Demographic and lifestyle factors		
Age, years, mean (SD)	57.3 (11.6)	51.7 (10.6)
Women, %	53.7	60.2
Urban residents, %	48.8	45.4
≥6 years of education, %	45.1	21.5
Ever regular smoker in men, %	75.0	74.0
Ever regular smoker in women, %	5.8	3.1
Ever regular drinker in men, %	34.6	36.9
Ever regular drinker in women, %	3.0	2.4
Physical activity, MET-h/day, mean (SD)	17.3 (10.7)	21.3 (13.9)
Medical history and health status, %		
Self-rated poor health	16.6	10.2
Diabetes	11.2	8.2
Chronic kidney disease	1.4	1.3
Cancer	0.6	0.5
Anthropometry and blood pressure, mean (SD)		
BMI, kg/m^2	23.9 (3.3)	23.6 (3.4)
WC, cm	81.9 (9.1)	80.1 (9.7)
SBP, mmHg	138 (22)	130 (21)
RPG, mmol/L	6.1 (2.5)	6.0 (2.3)
Fasting time, hours	4.7 (4.1)	5.2 (4.2)

[a] Adjusted for age, sex and study area, as appropriate.
Abbreviations: SD=Standard deviation; BMI=Body mass index; SBP=Systolic blood pressure; MET=Metabolic equivalent of task; RPG=Random plasma glucose; WC=Waist circumference

Table S3. Baseline characteristics of participants in UKB

Characteristics	UKB-PPP Randomized baseline	UKB full cohort
Demographic and lifestyle factors		
Age, years, mean (SD)	56.7 (8.1)	56.5 (8.1)
Women, %	54.3	54.4
Townsend Deprivation Index	-1.23 (3.15)	-1.29 (3.09)
Smoking		
Never	54.3%	54.8%
Previous	34.8%	34.6%
Current	10.9%	10.6%
Ethnic background		
Asian/Asian British	2.0%	2.0%
Black/Black British	1.7%	1.6%
Chinese	0.3%	0.3%
Mixed	0.7%	0.6%
White	94.4%	94.6%
Other ethnic group	1.0%	0.9%
ABO blood group		
O	43.2%	43.3%
A	43.4%	43.5%
B	9.7%	9.6%
AB	3.7%	3.6%
Anthropometry, mean (SD)		
BMI, kg/m^2	27.4 (4.8)	27.4 (4.8)
Biochemistry		
Alanine aminotransferase (U/l)	23.41 (14.28)	23.55 (14.18)
Alkaline phosphatase (U/l)	83.74 (26.3)	83.67 (26.46)
Aspartate aminotransferase (U/l)	26.27 (11.09)	26.23 (10.66)
Cholesterol (mmol/l)	5.69 (1.15)	5.69 (1.14)
Creatinine (umol/l)	72.28 (18.91)	72.31 (18.55)
Gamma glutamyltransferase (U/l)	37.29 (41.11)	37.39 (42.09)
Glucose (mmol/l)	5.13 (1.25)	5.12 (1.24)
Glycated haemoglobin (HbA1c) (mmol/mol)	36.20 (6.86)	36.13 (6.78)
HDL-cholesterol (mmol/l)	1.45 (0.38)	1.45 (0.38)
Triglycerides (mmol/l)	1.75 (1.03)	1.75 (1.03)

Table S4. The *cis*-pQTLs in CKB and UKB

Exposure	Ancestry	Variant_ID	rsID	CHR	BP (b37)	EA	AA	EA_FREQ_protein	BETA_protein	SE_protein	P_protein	PIP_protein
ALDH2	EAS	12:112230019:C:G	rs4646776	12	112230019	C	G	0.214	-0.371	0.0277	6.30E-41	0.419
FGF5	EAS	4:81184341:A:T	rs16998073	4	81184341	T	A	0.406	0.634	0.0227	3.00E-172	0.998
ABO	EUR	9:136149229:C:T	rs505922	9	136149229	C	T	0.317	1.171	0.0082	0	1.000
F11	EUR	4:187207381:C:T	rs2289252	4	187207381	T	C	0.399	0.385	0.0070	0	0.910
FGF5	EUR	4:81182554:C:T	rs12509595	4	81182554	C	T	0.293	0.682	0.0078	0	0.999
FURIN	EUR	15:91420940:C:G	rs2071410	15	91420940	G	C	0.326	0.194	0.0079	3.52E-135	0.323
GRK5	EUR	10:121010256:A:G	rs10886430	10	121010256	G	A	0.128	-0.407	0.0118	3.38E-262	1.000
KIAA0319	EUR	6:24654443:C:T	rs3181238	6	24654443	T	C	0.640	0.308	0.0078	0	0.782
Apo(a)	EUR	6:161089307:C:T	rs56393506	6	161089307	T	C	0.172	0.761	0.0104	0	1.000
MMP12	EUR	11:102748695:A:G	rs17368814	11	102748695	G	A	0.125	-0.749	0.0106	0	0.638
PROCR	EUR	20:33764554:A:G	rs867186	20	33764554	G	A	0.086	0.730	0.0138	0	0.193
SCARA5	EUR	8:27805783:C:T	rs2726951	8	27805783	T	C	0.269	-0.337	0.0079	0	0.662
TMPRSS5	EUR	11:113561421:A:C	rs7114195	11	113561421	C	A	0.639	0.785	0.0078	0	0.849

Table S5. Causal effects of 12 proteins on stroke outcomes in EAS and EUR

	Ischemic stroke in EUR			Ischemic stroke in EAS			Any stroke in EUR			Any stroke in EAS		
Protein	**OR (95%CI)**	***P***	***P_FDR***	**OR (95%CI)**	***P***	***P_FDR***	**OR (95%CI)**	***P***	***P_FDR***	**OR (95%CI)**	***P***	***P_FDR***
ABO	1.04 (1.03-1.06)	3.6E-10	2.6E-07	1.00 (0.97-1.03)	9.2E-01	9.9E-01	1.03 (1.02-1.05)	9.9E-08	7.2E-05	1.00 (0.98-1.03)	7.1E-01	1.0E+00
F11	1.12 (1.08-1.16)	6.6E-09	3.2E-06	1.04 (0.96-1.12)	3.6E-01	9.2E-01	1.08 (1.05-1.12)	2.9E-06	8.5E-04	1.03 (0.97-1.10)	3.2E-01	9.5E-01
FGF5	1.04 (1.02-1.07)	2.1E-04	3.8E-02	1.14 (1.08-1.19)	2.1E-07	9.4E-05	1.04 (1.02-1.06)	3.9E-04	5.1E-02	1.15 (1.10-1.20)	9.0E-11	8.4E-08
FURIN	1.23 (1.14-1.34)	8.7E-07	2.5E-04	1.15 (0.96-1.39)	1.3E-01	8.4E-01	1.21 (1.12-1.30)	1.4E-06	5.0E-04	1.10 (0.94-1.29)	2.3E-01	9.2E-01
GRK5	0.83 (0.78-0.87)	6.5E-11	9.5E-08	-	-	-	0.84 (0.80-0.89)	4.6E-10	6.7E-07	-	-	-
KIAA0319	0.90 (0.86-0.94)	1.5E-05	3.7E-03	0.90 (0.79-1.02)	9.6E-02	8.4E-01	0.91 (0.87-0.95)	2.3E-05	4.9E-03	0.90 (0.80-1.00)	5.6E-02	7.9E-01
Apo(a)	1.05 (1.02-1.08)	4.8E-04	6.3E-02	-	-	-	1.05 (1.02-1.07)	1.9E-04	2.7E-02	-	-	-
MMP12	0.93 (0.91-0.96)	2.3E-07	8.2E-05	0.95 (0.86-1.05)	2.9E-01	8.8E-01	0.94 (0.91-0.96)	1.9E-07	9.1E-05	0.97 (0.89-1.05)	4.7E-01	9.9E-01
PROCR	0.92 (0.89-0.96)	2.7E-05	5.5E-03	0.97 (0.89-1.07)	5.5E-01	9.5E-01	0.93 (0.89-0.96)	1.0E-05	2.5E-03	0.97 (0.90-1.04)	3.5E-01	9.7E-01
SCARA5	0.92 (0.88-0.97)	8.3E-04	8.6E-02	0.96 (0.88-1.04)	2.8E-01	8.8E-01	0.92 (0.88-0.96)	1.4E-04	2.3E-02	0.99 (0.92-1.06)	7.5E-01	1.0E+00
TMPRSS5	1.04 (1.02-1.05)	2.4E-04	3.8E-02	1.01 (0.97-1.05)	5.6E-01	9.5E-01	1.04 (1.02-1.05)	6.6E-05	1.2E-02	1.02 (0.99-1.05)	2.2E-01	9.2E-01
ALDH2	-	-	-	1.60 (1.39-1.84)	8.2E-11	7.3E-08	-	-	-	1.36 (1.23-1.51)	9.8E-09	4.6E-06

Table S6. MR mediation results for 6 protein targets on ischaemic stroke outcomes via SBP

Population	Exposure	Mediator	Outcome	Mediation/Indirect effect		Total Effect*			Proportion of mediation effect, %
				beta	se	beta	se	pval	
EUR	FGF5	SBP	IS	0.033	0.003	0.042	0.011	2.1E-04	77
EUR	FURIN	SBP	IS	0.087	0.008	0.210	0.042	8.7E-07	41
EUR	KIAA0319	SBP	IS	-0.008	0.003	-0.107	0.025	1.5E-05	8
EUR	PROCR	SBP	IS	-0.001	0.002	-0.082	0.019	2.7E-05	1
EUR	TMPRSS5	SBP	IS	0.002	0.001	0.035	0.009	2.4E-04	6
EAS	FGF5	SBP	IS	0.006	0.002	0.129	0.024	2.1E-07	5
EAS	ALDH2	SBP	IS	0.005	0.004	0.470	0.071	8.2E-11	1

* Total effect is the effect of exposure on outcome derived from Two-sample MR

Table S7. Summary of MR using *trans*-pQTL for 11 proteins associated with stroke types in EUR

Assay target	No. of *cis*-pQTL	No. of *trans*-pQTL (annotated gene)	MR using *trans*-pQTL
ABO	1	1 (FUT6)	
F11	1	4 (KNG1, C4B, ABCA6, GCKR)	KNG1
FGF5	1	1 (SERPINA1)	
FURIN	1	2 (MLXIPL, GCKR)	
GRK5	1	0	
KIAA0319	1	0	
Apo(a)	1	1 (APOC1)	
MMP12	1	4 (CD163, LILRB5, MIR6891, SH2B3, MSR1)	SH2B3
PROCR	1	0	
SCARA5	1	5 (LINC02356, ABO, ST3GAL4)	LINC02356
TMPRSS5	1	19 (ARHGAP24, LINC02474, MERTK, LINC02026, ZNF568, FUT9, SPPL2A, SOS2, MLH3, B4GALT1, MRC1, ETS2, ST3GAL4, DLEU1, PHLDA3)	

Table S8. PheWAS results of 10 IS-related proteins

Gene	Phenotype	Huge score	Group	Evidence Range
ABO	ALP	345	HEPATIC	Extreme
ABO	DBP	46	CARDIOVASCULAR	Very strong
ABO	BSadjFastingTime	45	GLYCEMIC	Very strong
ABO	Pancreas_volume	45	GLYCEMIC	Very strong
ABO	vitDBP_GChapAdj	45	METABOLITE	Very strong
ABO	NeutroPerc	45	HEMATOLOGICAL	Very strong
ABO	HR	45	CARDIOVASCULAR	Very strong
ABO	CRP	45	CARDIOVASCULAR	Very strong
ABO	DirectBilirubin	45	HEPATIC	Very strong
ABO	TG	45	LIPIDS	Very strong
ABO	VitD	45	METABOLITE	Very strong
ABO	Hematocrit	45	HEMATOLOGICAL	Very strong
ABO	ALT	45	HEPATIC	Very strong
ABO	nonHDL	45	LIPIDS	Very strong
ABO	Stroke	45	STROKE	Very strong
ABO	AST_ALT_ratio	45	HEPATIC	Very strong
ABO	DI_Kadowaki_Matsuda	45	GLYCEMIC	Very strong
ABO	HYPERTENSION	45	CARDIOVASCULAR	Very strong
ABO	MonoPerc	45	HEMATOLOGICAL	Very strong
ABO	HypoThyroidNOS	45	GLYCEMIC	Very strong
ABO	CAD	45	CARDIOVASCULAR	Very strong
ABO	Ferritin	45	METABOLITE	Very strong
ABO	LymphoPerc	45	HEMATOLOGICAL	Very strong
ABO	POAG	45	OCULAR	Very strong
ABO	Dyslipid	45	LIPIDS	Very strong
ABO	ApoB	45	LIPIDS	Very strong
ABO	TSH	45	GLYCEMIC	Very strong
ABO	BS	45	GLYCEMIC	Very strong
ABO	UMOD_aptamer	45	RENAL	Very strong
ABO	GGT	45	HEPATIC	Very strong
ABO	T2DwNM	45	GLYCEMIC	Very strong
ABO	Stroke_ischemic	45	STROKE	Very strong
ABO	TGnonT2D	45	LIPIDS	Very strong
ABO	toastCE	45	STROKE	Very strong
ABO	ReticuloCount	45	HEMATOLOGICAL	Very strong
ABO	tLVSV	45	CARDIOVASCULAR	Very strong
ABO	MeanCorpHbConc	45	HEMATOLOGICAL	Very strong
ABO	MeanReticuloVol	45	HEMATOLOGICAL	Very strong
ABO	eGFRcrcys	45	RENAL	Very strong
ABO	HF	45	CARDIOVASCULAR	Very strong
ABO	ProteinIntake	45	NUTRITIONAL	Very strong
ABO	Intraocular_pressure	45	OCULAR	Very strong
ABO	HbConc	45	HEMATOLOGICAL	Very strong
ABO	IGF1	45	OTHER	Very strong
ABO	MI	45	CARDIOVASCULAR	Very strong

ABO	NeutCount	45	HEMATOLOGICAL	Very strong
ABO	PVD	45	CARDIOVASCULAR	Very strong
ABO	Thyroid	45	GLYCEMIC	Very strong
ABO	eGFRcrea	45	RENAL	Very strong
ABO	HBA1C	45	GLYCEMIC	Very strong
ABO	LDH	45	METABOLITE	Very strong
ABO	2hrGadjBMI	45	GLYCEMIC	Very strong
ABO	MonoCount	45	HEMATOLOGICAL	Very strong
ABO	DiabeticRetino	45	DIABETIC COMPLICATIONS	Very strong
ABO	FG	45	GLYCEMIC	Very strong
ABO	TotalProtein	45	HEMATOLOGICAL	Very strong
ABO	SHBGadjBMI	45	REPRODUCTIVE TRAITS	Very strong
ABO	PancreasFat	45	GLYCEMIC	Very strong
ABO	BILIRUBIN	45	HEPATIC	Very strong
ABO	VaricoseVeins	45	CARDIOVASCULAR	Very strong
ABO	PulsePress	45	CARDIOVASCULAR	Very strong
ABO	AST	45	HEPATIC	Very strong
ABO	PlatCrit	45	HEMATOLOGICAL	Very strong
ABO	FGadjBMI	45	GLYCEMIC	Very strong
ABO	Testosterone_total	45	REPRODUCTIVE TRAITS	Very strong
ABO	NucleatedRedPerc	45	HEMATOLOGICAL	Very strong
ABO	RedCount	45	HEMATOLOGICAL	Very strong
ABO	T2DwOM	45	GLYCEMIC	Very strong
ABO	Estradiol	45	REPRODUCTIVE TRAITS	Very strong
ABO	Hb	45	HEMATOLOGICAL	Very strong
ABO	BSandFG	45	GLYCEMIC	Very strong
ABO	BUN	45	RENAL	Very strong
ABO	SHBG	45	REPRODUCTIVE TRAITS	Very strong
ABO	Ca	45	METABOLITE	Very strong
ABO	HBA1CadjBMI	45	GLYCEMIC	Very strong
ABO	TGtoHDL	45	LIPIDS	Very strong
ABO	pancreasT1time	45	GLYCEMIC	Very strong
ABO	MeanSpheredVol	45	HEMATOLOGICAL	Very strong
ABO	liverT1time	45	HEPATIC	Very strong
ABO	CHOL	45	LIPIDS	Very strong
ABO	Alb	45	HEPATIC	Very strong
ABO	LVEDVI	45	CARDIOVASCULAR	Very strong
ABO	HighScatReticuloCount	45	HEMATOLOGICAL	Very strong
ABO	NucleatedRedCount	45	HEMATOLOGICAL	Very strong
ABO	xinsG30	45	GLYCEMIC	Very strong
ABO	T2DadjBMI	45	GLYCEMIC	Very strong
ABO	EO_Stroke_ischemic	45	STROKE	Very strong
ABO	EosinCount	45	HEMATOLOGICAL	Very strong
ABO	Creatinine	45	RENAL	Very strong
ABO	ApoA	45	LIPIDS	Very strong
ABO	T2D	45	GLYCEMIC	Very strong
ABO	SBP	45	CARDIOVASCULAR	Very strong

ABO	WBC	45	HEMATOLOGICAL	Very strong
ABO	eGFRcreaNoDiabetes	45	RENAL	Very strong
ABO	ReticuloPerc	45	HEMATOLOGICAL	Very strong
ABO	HmPerc	45	HEMATOLOGICAL	Very strong
ABO	PulseRate	45	CARDIOVASCULAR	Very strong
ABO	FIadjBMI	45	GLYCEMIC	Very strong
ABO	SVI	45	CARDIOVASCULAR	Very strong
ABO	ProteinIntakeBMIadj	45	NUTRITIONAL	Very strong
ABO	LymphoCount	45	HEMATOLOGICAL	Very strong
ABO	HDL	45	LIPIDS	Very strong
ABO	CK	45	CARDIOVASCULAR	Very strong
ABO	LDL	45	LIPIDS	Very strong
F11	Stroke_ischemic	45	STROKE	Very strong
ALDH2	GGT	350	HEPATIC	Compelling
ALDH2	HEIGHT	114	ANTHROPOMETRIC	Extreme
ALDH2	DescAortaDiam	45	CARDIOVASCULAR	Very strong
ALDH2	AST	45	HEPATIC	Very strong
ALDH2	ALP	45	HEPATIC	Very strong
ALDH2	PlatVol	45	HEMATOLOGICAL	Very strong
ALDH2	UA	45	RENAL	Very strong
ALDH2	T2D	20	GLYCEMIC	Strong
ALDH2	ALT	20	HEPATIC	Strong
ALDH2	PlatCount	20	HEMATOLOGICAL	Strong
ALDH2	FG	20	GLYCEMIC	Strong
ALDH2	Stroke_ischemic	20	STROKE	Strong
ALDH2	SBP	20	CARDIOVASCULAR	Strong
ALDH2	MeanCorpHb	20	HEMATOLOGICAL	Strong
ALDH2	HYPERTENSION	20	CARDIOVASCULAR	Strong
ALDH2	CAD	20	CARDIOVASCULAR	Strong
ALDH2	MeanCorpVol	20	HEMATOLOGICAL	Strong
ALDH2	HDL	20	LIPIDS	Strong
ALDH2	MeanCorpHbConc	20	HEMATOLOGICAL	Strong
ALDH2	Stroke	20	STROKE	Strong
ALDH2	DBP	20	CARDIOVASCULAR	Strong
ALDH2	LDL	20	LIPIDS	Strong
MMP12	HBA1C	45	GLYCEMIC	Very strong
PROCR	MI	350	CARDIOVASCULAR	Compelling
PROCR	CAD	350	CARDIOVASCULAR	Compelling
PROCR	Creatinine	45	RENAL	Very strong
PROCR	eGFRcreaNoDiabetes	45	RENAL	Very strong
PROCR	HEIGHT	28	ANTHROPOMETRIC	Strong
GRK5	HEIGHT	47	ANTHROPOMETRIC	Very strong
GRK5	PlatDistWidth	45	HEMATOLOGICAL	Very strong
GRK5	LymphoCount	45	HEMATOLOGICAL	Very strong
GRK5	Stroke_ischemic	45	STROKE	Very strong
GRK5	BUN	45	RENAL	Very strong
GRK5	FT4	45	GLYCEMIC	Very strong
GRK5	PlatVol	45	HEMATOLOGICAL	Very strong

GRK5	Cl	45	METABOLITE	Very strong
GRK5	PlatCount	45	HEMATOLOGICAL	Very strong
GRK5	LymphoPerc	45	HEMATOLOGICAL	Very strong
GRK5	NeutroPerc	45	HEMATOLOGICAL	Very strong
GRK5	Ca	45	METABOLITE	Very strong
FGF5	Creatinine	45	RENAL	Very strong
FGF5	eGFRcys	45	RENAL	Very strong
FGF5	Hematocrit	45	HEMATOLOGICAL	Very strong
FGF5	RedCount	45	HEMATOLOGICAL	Very strong
FGF5	HEIGHT	45	ANTHROPOMETRIC	Very strong
FGF5	DBP	45	CARDIOVASCULAR	Very strong
FGF5	Cl	45	METABOLITE	Very strong
FGF5	K	45	METABOLITE	Very strong
FGF5	eGFRcreaNoDiabetes	45	RENAL	Very strong
FGF5	eGFRcrea_med_DeclineAdjBL	45	RENAL	Very strong
FGF5	ALP	45	HEPATIC	Very strong
FGF5	MAP	45	CARDIOVASCULAR	Very strong
FGF5	PulsePress	45	CARDIOVASCULAR	Very strong
FGF5	CystatinC	45	RENAL	Very strong
FGF5	MI	45	CARDIOVASCULAR	Very strong
FGF5	GestationalHypertension	45	REPRODUCTIVE TRAITS	Very strong
FGF5	eGFRcrcys	45	RENAL	Very strong
FGF5	BUN	45	RENAL	Very strong
FGF5	NucleatedRedPerc	45	HEMATOLOGICAL	Very strong
FGF5	MeanCorpVol	45	HEMATOLOGICAL	Very strong
FGF5	PRI	45	ECG TRAITS	Very strong
FGF5	CKD	45	RENAL	Very strong
FGF5	DescAortaDiam	45	CARDIOVASCULAR	Very strong
FGF5	PreEclampsia	45	REPRODUCTIVE TRAITS	Very strong
FGF5	CAD	45	CARDIOVASCULAR	Very strong
FGF5	AF	45	ATRIAL FIBRILLATION	Very strong
FGF5	NucleatedRedCount	45	HEMATOLOGICAL	Very strong
FGF5	Stroke	45	STROKE	Very strong
FGF5	AfibFlutter	45	ATRIAL FIBRILLATION	Very strong
FGF5	Bicarbonate	45	METABOLITE	Very strong
FGF5	HbConc	45	HEMATOLOGICAL	Very strong
FGF5	eGFRcrea	45	RENAL	Very strong
FGF5	SBP	45	CARDIOVASCULAR	Very strong
FGF5	Stroke_ischemic	45	STROKE	Very strong
FGF5	LVESAS	45	CARDIOVASCULAR	Very strong
FGF5	Urate	45	RENAL	Very strong
FGF5	HYPERTENSION	45	CARDIOVASCULAR	Very strong
FGF5	HmPerc	45	HEMATOLOGICAL	Very strong
FGF5	BMI	45	ANTHROPOMETRIC	Very strong
FGF5	UA	45	RENAL	Very strong
FGF5	eGFRcreaInDiabetes	45	RENAL	Very strong
FURIN	HYPERTENSION	45	CARDIOVASCULAR	Very strong
FURIN	NeutroPerc	45	HEMATOLOGICAL	Very strong

FURIN	BasoPerc	45	HEMATOLOGICAL	Very strong
FURIN	MBW	45	REPRODUCTIVE TRAITS	Very strong
FURIN	HDL	45	LIPIDS	Very strong
FURIN	Stroke	45	STROKE	Very strong
FURIN	PulsePress	45	CARDIOVASCULAR	Very strong
FURIN	MeanSpheredVol	45	HEMATOLOGICAL	Very strong
KIAA0319	MeanCorpVol	45	HEMATOLOGICAL	Very strong
KIAA0319	RBCDistWidth	45	HEMATOLOGICAL	Very strong
KIAA0319	LymphoCount	45	HEMATOLOGICAL	Very strong
KIAA0319	PlatCount	45	HEMATOLOGICAL	Very strong
KIAA0319	HEIGHT	45	ANTHROPOMETRIC	Very strong
KIAA0319	MeanSpheredVol	45	HEMATOLOGICAL	Very strong
KIAA0319	PlatVol	45	HEMATOLOGICAL	Very strong
KIAA0319	ALP	20	HEPATIC	Strong
KIAA0319	MeanCorpHb	20	HEMATOLOGICAL	Strong
TMPRSS5	SHBG	45	REPRODUCTIVE TRAITS	Very strong
TMPRSS5	TotalProtein	45	HEMATOLOGICAL	Very strong
TMPRSS5	EosinPerc	45	HEMATOLOGICAL	Very strong
TMPRSS5	BILIRUBIN	20	HEPATIC	Strong
TMPRSS5	SHBGadjBMI	20	REPRODUCTIVE TRAITS	Strong
TMPRSS5	PlatVol	20	HEMATOLOGICAL	Strong
TMPRSS5	PlatCount	20	HEMATOLOGICAL	Strong

PheWAS results were obtained from CMDKP (https://hugeamp.org/) using HuGE Scores to evaluate the extent to which human genetic evidence supports the role of a gene in a disease or phenotype. Phenotypes with supporting evidence stronger than "strong" (compelling, extreme, very strong and strong) were included.

Figure S1. Colocalization analysis of 2 proteins and IS GWAS in EAS

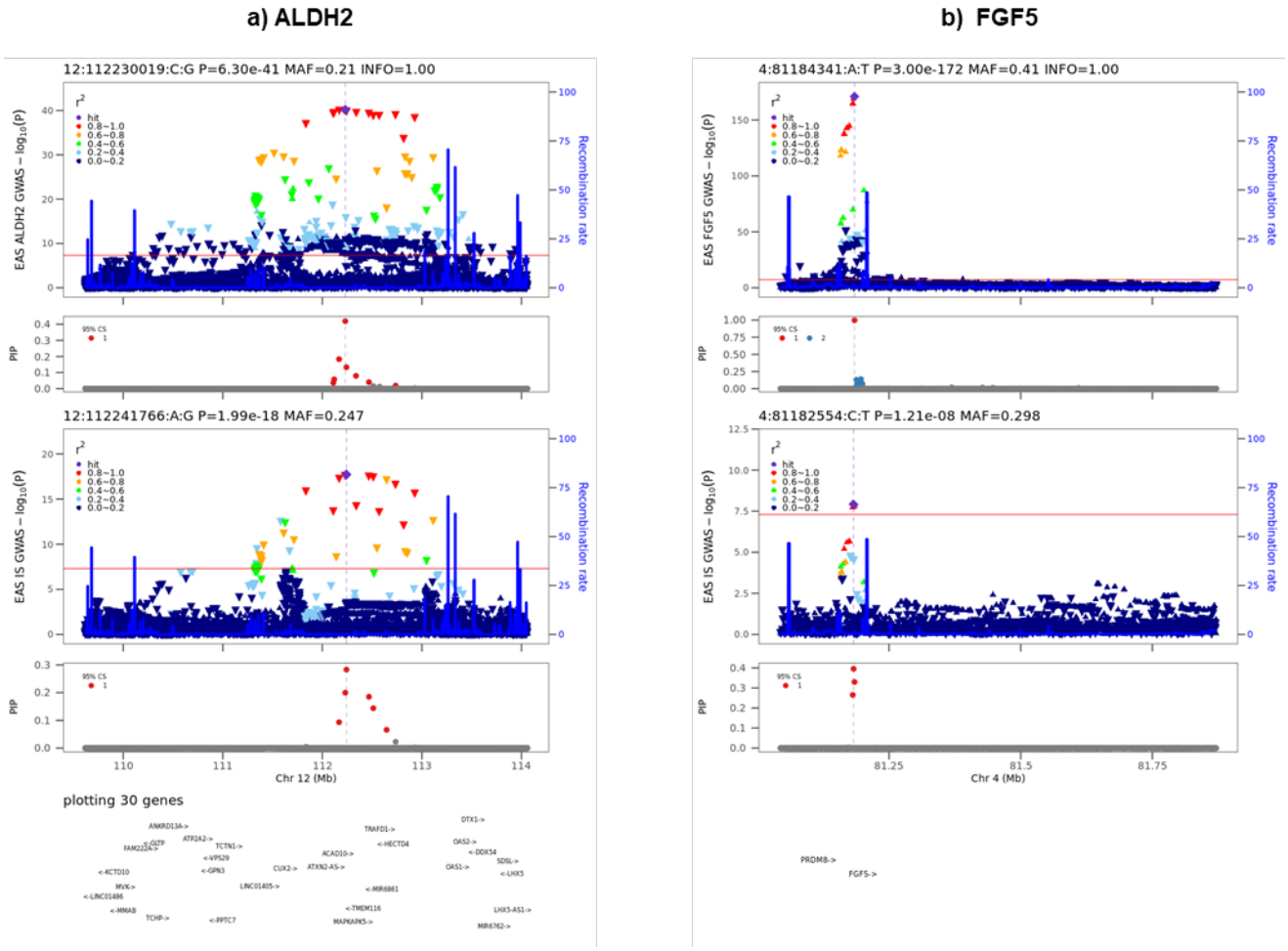

Figure S2. Causal effects of 12 proteins on stroke outcomes and several risk factors for stroke

Red represents significant positive associations and blue represents significant inverse associations. Abbreviations: AS=any stroke; BP=blood pressure; IS=ischemic stroke; CES=cardioembolic stroke; LAS=large artery stroke; SVS=small vessel stroke; T2D=type 2 diabetes; LDL=low-density lipoprotein cholesterol; BMI=body mass index; AF=atrial fibrillation.

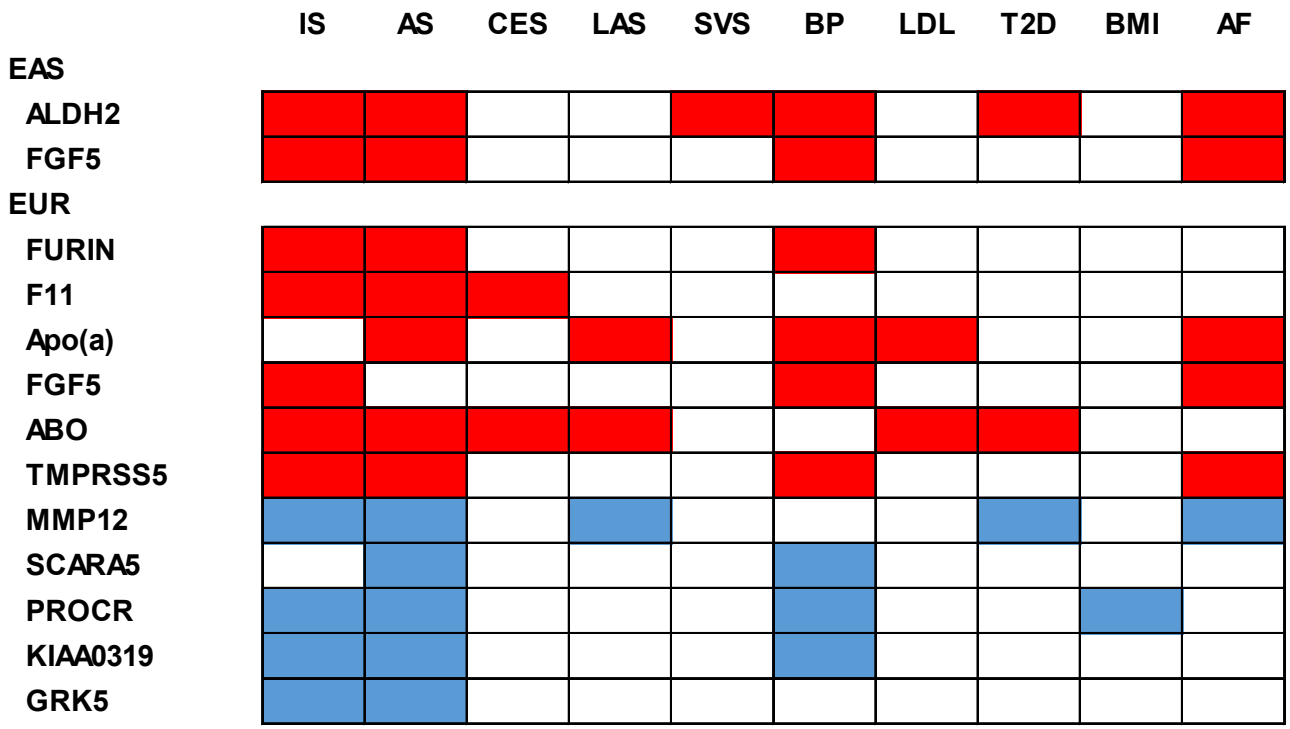

Figure S3. Colocalization analysis of 9 proteins and IS GWAS in EUR