Journal of the American Heart Association

ORIGINAL RESEARCH

Causal Effects of YKL-40 on Ischemic Stroke and Its Subtypes: A 2-Sample Mendelian Randomization Study

Qingyun Xu, MD;* Lulu Sun, MD;* Yinan Wang, MD; Ruirui Wang, MD; Yiming Jia, MD; Daoxia Guo, MD, PhD; Mengyao Shi, MD, PhD; Pinni Yang, MD; Yonghong Zhang, MD, PhD; Zhengbao Zhu, MD, PhD

BACKGROUND: Chitinase-3 like protein 1 (CHI3L1, YKL-40) was reported to be implicated in the development of ischemic stroke, but whether the association between them was causal remained unclear. We conducted a 2-sample Mendelian randomization study to explore the associations of genetically determined plasma YKL-40 with ischemic stroke and its subtypes (large artery stroke, small vessel stroke, and cardioembolic stroke).

METHODS AND RESULTS: Based on genome-wide association study data of 3394 European-descent individuals, we selected 13 single-nucleotide polymorphisms associated with plasma YKL-40 as genetic instruments. Summary data about ischemic stroke and its subtypes were obtained from the Multiancestry Genome-wide Association Study of Stroke Consortium, involving 34 217 ischemic stroke cases and 406 111 controls of European ancestry. We used the inverse-variance weighted method followed by a series of sensitivity analyses to assess the causal associations of plasma YKL-40 with ischemic stroke and its subtypes. The primary analysis showed that genetically determined high YKL-40 levels were associated with increased risks of large artery stroke (odds ratio [OR], 1.08 [95% CI, 1.04–1.12]; P=1.73×10^{-4}) and small vessel stroke (OR, 1.05 [95% CI, 1.01–1.09]; P=7.96×10^{-3}) but not with ischemic stroke or cardioembolic stroke. Sensitivity analyses further confirmed these associations, and Mendelian randomization-Egger indicated no evidence of genetic pleiotropy. In addition, supplementary analysis based on the summary data from the Olink proximity extension assay cardiovascular I (Olink CVD-I) panel showed that high YKL-40 levels were positively associated with the risks of large artery stroke (OR, 1.15 [95% CI, 1.08–1.22]; P=4.16×10^{-6}) but not with small vessel stroke.

CONCLUSIONS: Genetically determined high plasma YKL-40 levels were causal associated with increased risks of large artery stroke.

Key Words: large artery stroke ■ Mendelian randomization ■ small vessel stroke ■ YKL-40

In 2019, there were 101 million stroke cases globally, in which 77.19 million were ischemic stroke.[1] In the past decades, intervention strategies for ischemic stroke have made great progress, but ischemic stroke remains a major public health challenge.[2] Ischemic stroke is a complex disease with multiple pathophysiological pathways, and the established traditional risk factors can only explain part of the risk of ischemic stroke.[3] Therefore, it is urgent for us to accurately identify novel biomarkers for the prediction and prevention of ischemic stroke.

Chitinase-3 like protein 1 (CHI3L1, YKL-40), a 40 kDa heparin and chitin binding glycoprotein, is a member

Correspondence to: Zhengbao Zhu, MD, PhD and Yonghong Zhang, MD, PhD, Department of Epidemiology, School of Public Health and Jiangsu Key Laboratory of Preventive and Translational Medicine for Geriatric Diseases, Suzhou Medical College of Soochow University, 199 Renai Rd, Industrial Park District, Suzhou, Jiangsu Province 215123, China. Email: zbzhu@suda.edu.cn; yhzhang@suda.edu.cn

*Q. Xu and L. Sun contributed equally.

This manuscript was sent to Jose R. Romero, MD, Associate Editor, for review by expert referees, editorial decision, and final disposition.

Supplemental Material is available at https://www.ahajournals.org/doi/suppl/10.1161/JAHA.122.029000

For Sources of Funding and Disclosures, see page 7.

JAHA is available at: www.ahajournals.org/journal/jaha

CLINICAL PERSPECTIVE

What Is New?

- To our knowledge, this is the first study to investigate the associations of genetically determined plasma YKL-40 with ischemic stroke and its subtypes via the 2-sample Mendelian randomization analysis.
- Based on well-designed large-scale genome-wide association studies, we found that genetically determined high YKL-40 levels were associated with increased risks of large artery stroke.

What Are the Clinical Implications?

- The present Mendelian randomization study demonstrated the potential causal relationships between plasma YKL-40 and the risk of large artery stroke from the genetic insights, suggesting that YKL-40 might be a promising biomarker to identify high-risk individuals for active monitoring and early intervention of large artery stroke.

Nonstandard Abbreviations and Acronyms

IVW	inverse-variance weighted
MEGASTROKE	Multiancestry Genome-Wide Association Study of Stroke Consortium Mendelian randomization
MR-PRESSO	Mendelian Randomization Pleiotropy Residual Sum and Outlier

of the mammalian chitinase-like proteins produced by lipid-laden macrophages inside the vessel wall.[4,5] YKL-40 has been widely suggested to play a major role in the progress of inflammation and endothelial dysfunction.[6,7] Several previous observational studies have reported a significant association between high YKL-40 and increased risk of ischemic stroke.[8,9] However, it is difficult to draw definitive conclusions about the causality of the association in observational studies because of residual confounding and reverse causation biases. Therefore, further studies that may provide novel clues for the prevention strategies of ischemic stroke are needed to explore the potential causality for the relationship between YKL-40 and ischemic stroke.

Mendelian randomization (MR) is an emerging method that uses genetic variants associated with the exposure as instrumental variables to estimate the potential causal associations between the exposures and the outcomes.[10] Potential unmeasured confounders and reverse causation can be minimized in the MR study because of random inheritance of parental genetic variants at conception.[10–12] Two-sample MR uses genetic variants that are associated with the exposure and outcome from 2 separate population studies and can increase the scope for application of the MR methodology.[13] Herein, we performed 2-sample MR analyses to explore the potential causal effects of genetically determined plasma YKL-40 levels on the risks of ischemic stroke and its 3 subtypes (large artery stroke [LAS], small vessel stroke [SVS], and cardioembolic stroke [CES]).

METHODS

Study Design

The data and methods that support the findings of this study are available from the corresponding author upon reasonable request. The present study followed the STROBE-MR (Strengthening the Reporting of Observational Studies in Epidemiology using MR) guidelines.[14] As shown in Figure 1, we designed a 2-sample MR study to investigate the associations of genetically determined plasma YKL-40 levels with the risks of ischemic stroke and its subtypes. There are 3 assumptions for MR design: (1) genetic variants directly affect exposures, (2) genetic variants are independent of potential confounders, and (3) genetic variants affect outcomes only through the effects on exposures. The summary-level data for plasma YKL-40 levels were obtained from the genome-wide association studies (GWASs) in the IMPROVE study (Carotid Intima Media Thickness and Intima Media Thickness-Progression as Predictors of Vascular Events in a High-Risk

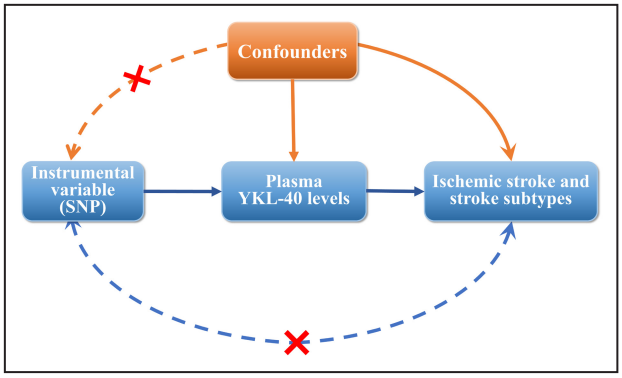

Figure 1. Conceptual framework for Mendelian randomization of plasma YKL-40 levels and the risk of ischemic stroke.
The assumption is that (1) instrumental variables are associated with plasma YKL-40 levels, (2) instrumental variables are not associated with confounders, (3) and instrumental variables affect ischemic stroke or its subtypes only through the effects on plasma YKL-40 levels. SNP indicates single-nucleotide polymorphism.

European Population).[15] Summary statistics about ischemic stroke and its subtypes were obtained from GWAS launched by the International Stroke Genetics Consortium (Multiancestry Genome-Wide Association Study of Stroke Consortium Mendelian randomization [MEGASTROKE]).[16] All participants in the present MR analysis were subjects of European ancestry. The protocol and data were approved by the ethics committee of the original GWASs, and written informed consent was obtained from each participant before data collection.

Selection of Instrumental Variables for Plasma YKL-40

The genetic data of plasma YKL-40 were obtained from the GWASs of 83 proteins with ≈5 million single-nucleotide polymorphisms (SNPs) among 3394 European-descent individuals in the IMPROVE study (available from the IEU GWAS database: https://gwas.mrcieu.ac.uk/).[15] In the present study, the SNPs that were identified to be associated with plasma YKL-40 levels at the genome-wide significance level ($P<5.0\times10^{-8}$) and were not in linkage disequilibrium with other SNPs ($r^2<0.3$ within a clumping window of 10 000 kb) were selected as genetic instruments for plasma YKL-40 levels. To better control the pleiotropy between SNPs, only cis-protein quantitative trait loci (limits SNP-gene distances to 1 Mb or less of *YKL-40/CHI3L1*) were screened as final instrumental variables. Overall, a total of 13 SNPs (rs10920579, rs17511046, rs2993655, rs61821143, rs6696205, rs2071579, rs2486961, rs883125, rs7551263, rs1845466, rs12410110, rs4950877, rs12033162) were selected as genetic instruments for plasma YKL-40 levels in this MR study, and all were located on chromosome 1 (Table 1).

Data Source for Ischemic Stroke and Its Subtypes

In the present study, summary data for the associations of each SNP with ischemic stroke and its 3 main subtypes (LAS, SVS, and CES) were derived from the previously published GWAS released by the MEGASTROKE project (https://gwas.mrcieu.ac.uk/).[16] Briefly, the MEGASTROKE was the largest international collaboration of stroke launched by the International Stroke Genetics consortium. This IMPROVE study data set was a meta-analysis of 29 European-ancestry GWASs with 8 million SNPs for associations with ischemic stroke in 440 328 European participants, including 34 217 cases and 406 111 controls.[16] Among these ischemic stroke cases, 7193 cases were CES, 4373 cases were LAS, 5386 cases were SVS. The classification of ischemic stroke is based on Trial of Org 10 172 in Acute Stroke Treatment criteria,[17] of which LAS caused by stenosis or occlusion of a major brain artery or branch cortical artery because of atherosclerosis,[18] SVS referring to lesions of the small perforating arteries and arterioles mainly caused by cerebral microvessel endothelial dysfunction,[19] and CES caused by arterial occlusions because of thrombogenic atrial substrate.[20]

Statistical Analysis

To minimize the influence of winner-curse bias on our study, False Discovery Rate Inverse Quantile Transformation, with more accurate and computationally efficient orders of magnitude, was used to obtain the adjusted beta by using "winnerscurse" package of R software.[21] The strength of the instrumental variables for plasma YKL-40 was assessed using the F statistic with the formula $F=\left(\frac{N-K-1}{K}\right)\left(\frac{R^2}{1-R^2}\right)$, where R^2 was the proportion of variation in genetically determined plasma YKL-40 explained by the SNPs, N was the sample size, and K was the number of SNPs in genetically determined plasma YKL-40.[22] An F statistic greater than 10 indicates the genetic variants being valid instrumental variables in MR study.[13] Additionally, the statistical power of this MR study was calculated using an online web tool named mRnd (https://shiny.cnsgenomics.com/mRnd/). We used the inverse-variance weighted (IVW) method to estimate the associations of genetically determined plasma YKL-40 with the risks of ischemic stroke and its subtypes in the main analysis.[23] Cochran Q statistic was calculated to quantify the heterogeneity among the genetic instruments.[24] To avoid possible causal estimates bias that might be caused by fitting different models, we conducted the fixed-effect IVW model and multiplicative random-effect IVW model to further explore the association between YKL-40 and ischemic stroke and its subtypes.[25,26]

A series of sensitivity analyses, including maximum likelihood, MR-Robust Adjusted Profile Score analysis, MR-Pleiotropy Residual Sum and Outlier (MR-PRESSO) analysis, MR-Egger regression analysis, and leave-one-out analysis, were performed to assess the robustness of our findings because of their resilience to violations of certain assumptions underlying the MR study. In the presence of measurement error in the SNP-exposure effect, the maximum likelihood method helps provide more reliable estimates.[27] The MR-Robust Adjusted Profile Score analysis models the pleiotropic effects of genetic variants directly using a random effects distribution and might solve the bias of horizontal pleiotropy and weak instruments.[28,29] The MR-PRESSO is used to identify potential outliers and might provide a robust estimate with outlier correction.[30] In addition, we used the MR-Egger regression method to evaluate the average pleiotropic effects across all SNPs via the intercept term.[31,32] Finally, the iterative leave-one-out analysis was conducted to

Table 1. Characterization of 13 SNPs Associated With Plasma YKL-40 Levels

SNP*	Chromosome	Position	Nearest gene	EA	OA	MAF	Beta_MAF	Beta	Beta_FIQT	SE	*P* value
rs10920579	1	203 158 972	*CHI3L1*	G	A	0.2116	−0.6325	0.6325	0.6292	0.028	8.73×10^{-8}
rs17511046	1	203 110 045	*ADORA1*	T	C	0.0641	−1.1201	−1.1201	−1.1092	0.078	1.08×10^{-45}
rs2993655	1	203 170 934	*CHI3L1*	A	G	0.4512	−0.4693	0.4693	0.4652	0.035	6.70×10^{-41}
rs61821143	1	203 145 409	*MYBPH*	G	C	0.2170	−0.4634	−0.4634	−0.4590	0.040	3.50×10^{-31}
rs6696205	1	203 088 614	*MYOPARR*	A	G	0.2532	−0.4400	−0.4400	−0.4354	0.044	1.58×10^{-23}
rs2071579	1	203 153 634	*CHI3L1*	G	C	0.4931	0.3315	0.3315	0.3290	0.033	3.07×10^{-23}
rs2486961	1	203 191 904	*CHIT1*	T	C	0.2229	0.4674	0.4674	0.4641	0.050	1.49×10^{-20}
rs883125	1	203 146 166	*MYBPH*	G	C	0.1905	0.4122	0.4122	0.4097	0.045	9.22×10^{-20}
rs7551263	1	203 150 756	*CHI3L1*	C	T	0.1908	0.3582	0.3582	0.3562	0.042	3.92×10^{-17}
rs1845466	1	203 120 143	*ADORA1*	T	G	0.3264	0.2626	−0.2626	−0.2609	0.039	1.18×10^{-11}
rs12410110	1	203 152 975	*CHI3L1*	C	G	0.0230	−0.8765	−0.8765	−0.8717	0.138	2.51×10^{-10}
rs4950877	1	203 055 101	*MYOG*	G	A	0.4440	−0.2403	0.2403	0.2398	0.041	5.37×10^{-9}
rs12033162	1	203 176 509	*CHIT1*	G	A	0.1703	0.2744	0.2744	0.2744	0.048	1.10×10^{-8}

Beta_FIQT indicates adjusted beta by FDR Inverse Quantile Transformation; EA, effect allele; EAF, effect allele frequency; MAF, minor allele frequencies; OA, other allele; and SNP, single-nucleotide polymorphism.
*The single-nucleotide polymorphisms associated with plasma YKL-40 levels were obtained from Folkersen et al's genome-wide association study.[15]

identify outlying or pleiotropic genetic variants by leaving each of them out of the MR analysis in turn.[30,31] To verify the robustness of our findings, we further used summary data of YKL-40 from GWAS data set of the Olink proximity extension assay cardiovascular I (Olink CVD-I) panel (https://pubmed.ncbi.nlm.nih.gov/33067605) to assess the associations of YKL-40 with ischemic stroke and its subtypes as supplementary analysis. Given that splice-site mutation could change protein conformation and function,[33,34] we also performed supplementary analysis by removing SNPs that were in the range of 500 base pairs (bp) of the splice site based on the location information in the National Center for Biotechnology Imformation website (http://www.ncbi.nih.gov/).

Results in this MR study were presented as odds ratios (ORs) with their 95% CIs of the outcomes (ischemic stroke, LAS, SVS, and CES) per SD increase in genetically determined log-transformed plasma YKL-40 levels. For all outcomes, a Bonferroni-corrected significance level of 2-sided $P<0.0125$ (0.05/4 [ischemic stroke and its 3 subtypes]) was considered statistically significant. All analyses were performed using R packages named winnerscurse, gtx, MR-PRESSO, MendelianRandomization, and TwoSampleMR in R software (version 4.1.0; R Development Core Team).

RESULTS

Genetic Instruments

A total of 13 SNPs were selected as instrumental variables for plasma YKL-40 in this study (Table 1), and the phenotypic variance explained by these 13 SNPs was ≈38.05% (Table S1). The *F* statistic for the genetic instruments of plasma YKL-40 was 159.73, suggesting that there was little weak instrument bias in the present study.

Associations Between Genetically Determined Plasma YKL-40 and Ischemic Stroke

Given the Q statistic and Q statistic associated *P* value of Cochran Q test, we inferred no heterogeneity across the genetic instruments (all $P>0.05$) (Table S2). As shown in Figure 2, the IVW analysis with fixed effect model showed that genetically determined high plasma YKL-40 levels were positively associated with the risks of LAS (OR per SD increase, 1.08 [95% CI, 1.04–1.12]; $P=1.73\times10^{-4}$) and SVS (OR per SD increase, 1.05 [95% CI, 1.01–1.09]; $P=7.96\times10^{-3}$). In contrast, no statistical evidence was found for the association of genetically determined YKL-40 levels with ischemic stroke (OR per SD increase, 1.00 [95% CI, 0.98–1.01]; $P=0.753$) or CES (OR per SD increase, 0.97 [95% CI, 0.94–0.99]; $P=0.020$). In the IVW analysis fitting multiplicative random effects

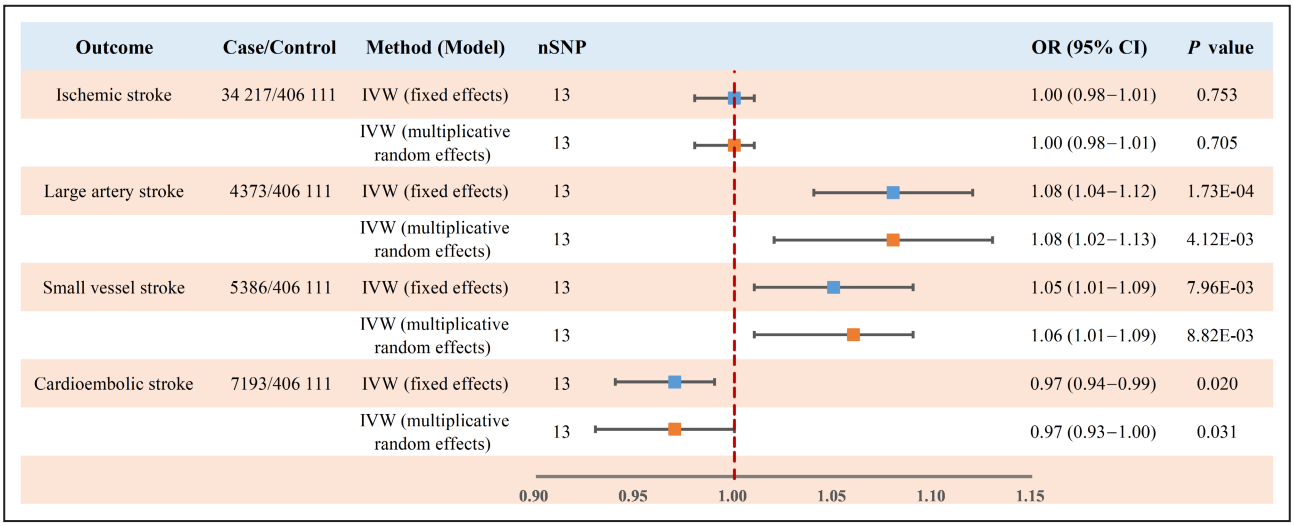

Figure 2. Associations of genetically determined plasma YKL-40 levels with the risks of ischemic stroke and its subtypes in the inverse-variance weighted Mendelian randomization analysis with fixed effects model and the multiplicative random effects model.

Effect estimates are presented as odds ratio (95% CI) per SD increase in genetically determined log-transformed plasma YKL-40 levels. IVW indicates inverse-variance weighted; nSNP, number of single-nucleotide polymorphism; and OR, odds ratio.

model, genetically determined high plasma YKL-40 levels were significantly associated with the risks of LAS (OR per SD increase, 1.08 [95% CI, 1.02–1.13]; P=4.12×10^{-3}) and SVS (OR per SD increase, 1.06 [95% CI, 1.01–1.09]; P=8.82×10^{-3}) but not with ischemic stroke (OR per SD increase, 1.00 [95% CI, 0.98–1.01]; P=0.705) or CES (OR per SD increase, 0.97 [95% CI, 0.93–1.00]; P=0.031). Associations between each instrumental variables for plasma YKL-40 levels and the risks of LAS and SVS are presented in Figure S1. Table S3 reveals that removing 2 possible outlier SNPs (rs17511046 and rs12410110) according to Figure S1 did not mask our findings.

Sensitivity Analyses

In the sensitivity analyses indicated in Table 2, genetically determined plasma YKL-40 levels are shown to be associated with LAS using the maximum likelihood method (OR, 1.08 [95% CI, 1.04–1.12]; P=1.82×10^{-4}) and MR Robust Adjusted Profile Score method (OR, 1.08 [95% CI, 1.04–1.12]; P=1.82×10^{-4}). In addition, genetically determined high plasma YKL-40 levels were also associated with the increased risk of SVS in the sensitivity analyses with the maximum likelihood method (OR, 1.05 [95% CI, 1.01–1.09]; P=7.95×10^{-3}) and MR Robust Adjusted Profile Score (OR, 1.05 [95% CI, 1.01–1.09]; P=8.16×10^{-3}) methods. Notably, the risk of LAS (OR, 1.08 [95% CI, 1.02–1.13]; P=0.014) and SVS (OR, 1.05 [95% CI, 1.01–1.09]; P=0.022) increased with a gradual increase of genetically determined plasma YKL-40 levels in MR-PRESSO method, while the associations did not reach the Bonferroni-corrected significance levels. Besides, because there was no difference between MR-PRESSO global test and MR-PRESSO outlier test, we indicated that horizontal pleiotropic outlier variants are unlikely to exist in these instrumental variables.[28] Furthermore, the intercept term of MR-Egger regression showed that our findings were not susceptible to the effects of the horizontal pleiotropy on the associations of plasma YKL-40 with ischemic stroke (odds [intercept]: 1.02 [95% CI, 1.00–1.04]; P=0.049), LAS (odds [intercept]: 1.07 [95% CI, 1.02–1.13]; P=0.025), SVS (odds [intercept]: 1.05 [95% CI, 1.00–1.10]; P=0.033) and CES (odds [intercept]: 1.02 [95% CI, 0.98–1.06]; P=0.414). Leave-one-out analysis showed that the associations of genetically determined plasma YKL-40 levels with ischemic stroke and its subtypes were not driven by any individual SNP (Figures S2 through S5).

In the supplementary analysis based on the summary data from the Olink CVD-I panel, a total of 40 SNPs were selected as instrumental variants of YKL-40 (Table S4). Our results showed that genetically determined high levels of YKL-40 were associated with increased risks of LAS (OR, 1.15 [95% CI, 1.08–1.22]; P=4.16×10^{-6}; Table 3). In further supplementary analysis by removing 2 SNPs within 500 bp of splice site (rs2486961 and rs4950877), our results still showed a significant association of genetically determined YKL-40 levels with LAS (OR, 1.08 [95% CI, 1.03–1.12]; P=3.26×10^{-4}; Table 3; Table S5).

DISCUSSION

To the best of our knowledge, this is the first study to investigate the associations of genetically determined plasma YKL-40 with ischemic stroke and its subtypes via the 2-sample MR analysis. Based on the

Table 2. Summary Mendelian Randomization Statistical Analyses for the Effect of Genetically Determined Plasma YKL-40 Levels on the Risks of Ischemic Stroke and Its Subtypes

Outcome	Parameter	OR (95% CI)	*P* value
IS			
Maximum likelihood	OR	1.00 (0.98–1.01)	0.751
MR-RAPS	OR	1.00 (0.98–1.01)	0.755
MR PRESSO	OR	1.00 (0.98–1.01)	0.711
	Outlier-corrected OR	1.00 (0.98–1.01)	0.711
MR-Egger	OR	0.97 (0.92–1.01)	0.138
	Odds (intercept)	1.02 (1.00–1.04)	0.049
LAS			
Maximum likelihood	OR	1.08 (1.04–1.12)	1.82×10^{-4}
MR-RAPS	OR	1.08 (1.04–1.12)	1.66×10^{-4}
MR PRESSO	OR	1.08 (1.02–1.13)	0.014
	Outlier-corrected OR	1.08 (1.02–1.13)	0.014
MR-Egger	OR	0.93 (0.83–1.05)	0.261
	Odds (intercept)	1.07 (1.02–1.13)	0.025
SVS			
Maximum likelihood	OR	1.05 (1.01–1.09)	7.95×10^{-3}
MR-RAPS	OR	1.05 (1.01–1.09)	8.16×10^{-3}
MR PRESSO	OR	1.05 (1.01–1.09)	0.022
	Outlier-corrected OR	1.05 (1.01–1.09)	0.022
MR-Egger	OR	0.95 (0.86–1.05)	0.336
	Odds (intercept)	1.05 (1.00–1.10)	0.033
CES			
Maximum likelihood	OR	0.96 (0.94–0.99)	0.020
MR-RAPS	OR	0.96 (0.94–0.99)	0.020
MR PRESSO	OR	0.97 (0.93–1.00)	0.052
	Outlier-corrected OR	0.97 (0.93–1.00)	0.052
MR-Egger	OR	0.93 (0.85–1.02)	0.148
	Odds (intercept)	1.02 (0.98–1.06)	0.414

CES indicates cardioembolic stroke; IS, ischemic stroke; IVW, inverse-variance weighted method; LAS, large artery stroke; MR-PRESSO, MR Pleiotropy Residual Sum and Outlier; MR-RAPS, Robust adjusted profile score; OR, odds ratio; and SVS, small vessel stroke.

large-scale GWAS from the MEGASTROKE consortium, we found that genetically determined high YKL-40 levels were associated with increased risks of LAS and SVS but not with ischemic stroke or CES with a series of sensitivity analyses confirming. Further supplementary analysis showed that genetically determined high YKL-40 levels was positively associated with the risk of LAS but not SVS, which suggested that plasma YKL-40 levels might play a critical role in the development of LAS.

Table 3. Supplementary Analysis for the Associations of Genetically Determined Plasma YKL-40 Levels With the Risks of Ischemic Stroke and Its Subtypes

Disease	OR (95% CI)	*P* value
Supplementary analysis based on the GWAS of Olink CVD-I panel		
IS	1.01 (0.98,1.03)	0.514
LAS	1.15 (1.08,1.22)	4.16×10^{-6}
SVS	1.07 (1.01,1.12)	0.021
CES	0.989 (0.93, 1.04)	0.602
Supplementary analysis by removing 2 SNPs* within 500 bp of splice site		
IS	1.00(0.98, 1.02)	0.949
LAS	1.08(1.03, 1.12)	3.26×10^{-4}
SVS	1.04(1.00, 1.08)	0.042
CES	0.96(0.93, 0.99)	0.014

CES indicates cardioembolic stroke; IS, ischemic stroke; LAS, large artery stroke; Olink CVD-I panel, the Olink proximity extension assay cardiovascular I panel; OR, odds ratio; and SVS, small vessel stroke.

*The 2 SNPs within 500 bp of splice site that we removed were rs2486961 and rs4950877.

In the past decades, there have been several observational studies suggesting the significant associations of YKL-40 with the risks of ischemic stroke and its subtypes.[8,9,35] For example, in an analysis based on the 141 patients with LAS and 85 controls, serum YKL-40 level was significantly higher among patients with LAS than that among controls.[35] In a cohort study with 8899 participants from the Copenhagen City Heart Study, high plasma YKL-40 levels were reported to be associated with increased risk of ischemic stroke.[8] Another cohort study conducted in the Danish population revealed that elevated YKL-40 levels were associated with increased risk of first-time stroke during the 15-year follow-up period.[9] Therefore, it was hypothesized that YKL-40 might be a causal risk factor for ischemic stroke. However, the association between YKL-40 and ischemic stroke might also be attributable to shared underlying risk factors (confounding) or ischemic stroke might causally influence the YKL-40 levels (reverse causation). Compared with traditional observational studies, our MR study used the genetic variants associated with YKL-40 as instruments to estimate the association between genetically determined plasma YKL-40 levels and ischemic stroke without confounding and reverse causation biases. In this large-scale MR study, we found YKL-40 to be associated with LAS but not ischemic stroke. This may be because of the relatively low proportion of LAS in GWAS of ischemic stroke, and other subtypes of ischemic stroke attenuated the effect of YKL-40 on overall ischemic stroke.

Several possible pathophysiological mechanisms underlying the detrimental effect of YKL-40 on the risks of LAS have been suggested by previous studies, including endothelial dysfunction, accelerated

growth, and instability of arteriosclerotic plaques.[36–38] As a pleiotropic inflammatory factor that affects vascular endothelial cells and macrophage, YKL-40 can activate endothelial cells to express vascular adhesion molecule-1 and intercellular adhesion molecule-1, thereby injuring the vascular endothelial cells and causing endothelial dysfunction.[36] In the early stage of arterial plaque formation, elevated expression of YKL-40 aggravates atherosclerotic lesions by inhibiting macrophage apoptosis in plaques.[39] In addition, high YKL-40 levels may specifically cause a loss of stably differentiated smooth muscle cell content, leading to thinning and rupture of arteriosclerotic plaque caps.[40] Therefore, the detailed mechanism underlying the different associations of YKL-40 with ischemic stroke and its subtypes warrants further study.

Our findings have several important public health significances and clinical implications. In the present MR study, we demonstrated the potential causal relationships between plasma YKL-40 and incidence of LAS from the genetic insights, which might provide novel clues for the prevention of LAS. From findings of our study, YKL-40 might act as a promising biomarker to identify high-risk individuals for active monitoring and early intervention of LAS. In addition, it is of clinical interest to explore whether targeting YKL-40 could reduce the risk of LAS.

The present study has several important strengths. First, the MR method used genetic variants as instrumental variables to make reasonable inferences for the potential causality, which avoided reverse causation and confounding biases.[41] Second, the genetic instrument for YKL-40 used in our MR study accounted for 38.47% of the variances in plasma YKL-40 with a high F statistic, indicating that the 13 SNPs used as proxies for plasma YKL-40 levels in the present MR study were valid genetic instruments. Third, the present study was conducted based on the largest available GWASs about YKL-40, ischemic stroke, and its subtypes, which enabled us to provide a valid appraisal of the association between genetically determined plasma YKL-40 levels and ischemic stroke with high statistical power.

Our study also has some limitations. First, it was difficult to completely avoid the impact of the potential pleiotropy on the results in the MR study.[32] However, MR-Egger regression analysis indicated that these associations were not susceptible to the effects of the horizontal pleiotropy. Second, this MR study estimated the lifetime effect of plasma YKL-40 levels on the risk of ischemic stroke and its subtypes, so the results should not be directly extrapolated to assess the effect of any potential clinical intervention targeting YKL-40. Third, there was potential heterogeneity between participants of the MEGASTROKE and the IMPROVE cohorts (eg, population health status and ethnicity), which possibly had an impact on the results of our 2-sample MR study.[42] Fourth, the participants included in the present study were of European ancestry, which decreased the possibility of spurious associations because of population stratification bias but restricted the extrapolation of our findings to non-European populations. Thus, further studies with multiethnic samples should be performed to confirm our findings. Fifth, because of the limitations of MR methodology, we could only detect weak instruments, horizontal pleiotropy, or heterogeneity via some statistical models, but it was hard to quantify these biases. Further studies without horizontal pleiotropy and heterogeneity are needed to verify our findings. Finally, as a biomarker of cerebrospinal fluid neuroinflammation that is encoded by the *CHI3L1* gene, YKL-40 is expressed primarily in astrocytes in the brain and macrophages in the periphery.[43] However, summary data of YKL-40/*CHI3L1* from cerebrospinal fluid are still lacking, and the potential causal associations between YKL-40 and the risks of ischemic stroke and its subtypes merit further investigation if relevant GWAS data sets about YKL-40/*CHI3L1* from cerebrospinal fluid are available in the future.

CONCLUSIONS

We found that genetically determined high plasma YKL-40 levels were associated with increased risks of LAS, suggesting that plasma YKL-40 levels might play a critical role in the development of LAS. Further studies are needed to verify our findings and explore the detailed mechanism underlying the detrimental effects of YKL-40 on the risk of LAS.

ARTICLE INFORMATION

Received November 27, 2022; accepted July 26, 2023.

Affiliations

Department of Epidemiology, School of Public Health and Jiangsu Key Laboratory of Preventive and Translational Medicine for Geriatric Diseases, Suzhou Medical College of Soochow University, Suzhou, China (Q.X., L.S., Y.W., R.W., Y.J., M.S., P.Y., Y.Z., Z.Z.); and School of Nursing (D.G.), Medical College of Soochow University, Suzhou, China.

Acknowledgments

We thank the investigators of IMPROVE study and MEGASTROKE consortia for making their results publicly available. Full acknowledgment and funding statements for each of these resources are available via the relevant cited reports. The MEGASTROKE project received funding from sources specified at http://www.megastroke.org/acknowledgments.html.

Sources of Funding

This study was supported by the National Natural Science Foundation of China (grant: 82103921 and 82020108028) and the Natural Science Research Project of Jiangsu Provincial Higher Education (grant: 21KJB330006).

Disclosures

Nothing to report.

Supplemental Material

Table S1–S5
Figure S1–S5

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. Montellano FA, Ungethüm K, Ramiro L, Nacu A, Hellwig S, Fluri F, Whiteley WN, Bustamante A, Montaner J, Heuschmann PU. Role of blood-based biomarkers in ischemic stroke prognosis: a systematic review. *Stroke*. 2021;52:543–551. doi: 10.1161/STROKEAHA.120.029232
3. Makris K, Haliassos A, Chondrogianni M, Tsivgoulis G. Blood biomarkers in ischemic stroke: potential role and challenges in clinical practice and research. *Crit Rev Clin Lab Sci*. 2018;55:294–328. doi: 10.1080/10408363.2018.1461190
4. Johansen JS. Studies on serum YKL-40 as a biomarker in diseases with inflammation, tissue remodelling, fibroses and cancer. *Dan Med Bull*. 2006;53:172–209.
5. Rehli M, Krause SW, Andreesen R. Molecular characterization of the gene for human cartilage gp-39 (CHI3L1), a member of the chitinase protein family and marker for late stages of macrophage differentiation. *Genomics*. 1997;43:221–225. doi: 10.1006/geno.1997.4778
6. Rathcke CN, Vestergaard H. YKL-40—an emerging biomarker in cardiovascular disease and diabetes. *Cardiovasc Diabetol*. 2009;8:61. doi: 10.1186/1475-2840-8-61
7. Xu T, Zhong CK, Wang AL, Guo ZR, Bu XQ, Zhou YP, Tian YF, HuangFu XF, Zhu ZB, Zhang YH. YKL-40 level and hypertension incidence: a population-based nested case-control study in China. *J Am Heart Assoc*. 2016;5:e004534. doi: 10.1161/JAHA.116.004534
8. Kjaergaard AD, Bojesen SE, Johansen JS, Nordestgaard BG. Elevated plasma YKL-40 levels and ischemic stroke in the general population. *Ann Neurol*. 2010;68:672–680. doi: 10.1002/ana.22220
9. Rathcke CN, Thomsen SB, Linneberg A, Vestergaard H. Variations of *CHI3L1*, levels of the encoded glycoprotein YKL-40 and prediction of fatal and non-fatal ischemic stroke. *PLoS One*. 2012;7:e43498. doi: 10.1371/journal.pone.0043498
10. Lawlor DA, Harbord RM, Sterne JA, Timpson N, Davey SG. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Stat Med*. 2008;27:1133–1163. doi: 10.1002/sim.3034
11. Davies NM, Holmes MV, Davey SG. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. *BMJ (Clin Res Ed.)*. 2018;362:k601. doi: 10.1136/bmj.k601
12. Gaspar FW, Castorina R, Maddalena RL, Nishioka MG, McKone TE, Bradman A. Phthalate exposure and risk assessment in California child care facilities. *Environ Sci Technol*. 2014;48:7593–7601. doi: 10.1021/es501189t
13. Zheng J, Baird D, Borges MC, Bowden J, Hemani G, Haycock P, Evans DM, Smith GD. Recent developments in Mendelian randomization studies. *Curr Epidemiol Rep*. 2017;4:330–345. doi: 10.1007/s40471-017-0128-6
14. Skrivankova VW, Richmond RC, Woolf BAR, Yarmolinsky J, Davies NM, Swanson SA, VanderWeele TJ, Higgins JPT, Timpson NJ, Dimou N, et al. Strengthening the reporting of observational studies in epidemiology using Mendelian randomization: the strobe-MR statement. *JAMA*. 2021;326:1614–1621. doi: 10.1001/jama.2021.18236
15. Folkersen L, Fauman E, Sabater-Lleal M, Strawbridge RJ, Frånberg M, Sennblad B, Baldassarre D, Veglia F, Humphries SE, Rauramaa R, et al. Mapping of 79 loci for 83 plasma protein biomarkers in cardiovascular disease. *PLoS Genet*. 2017;13:e1006706. doi: 10.1371/journal.pgen.1006706
16. Malik R, Chauhan G, Traylor M, Sargurupremraj M, Okada Y, Mishra A, Rutten-Jacobs L, Giese AK, van der Laan SW, Gretarsdottir S, et al. Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nat Genet*. 2018;50:524–537. doi: 10.1038/s41588-018-0058-3
17. Adams HP Jr, Bendixen BH, Kappelle LJ, Biller J, Love BB, Gordon DL, Marsh EE III. Classification of subtype of acute ischemic stroke. Definitions for use in a multicenter clinical trial. Toast. Trial of org 10172 in acute stroke treatment. *Stroke*. 1993;24:35–41. doi: 10.1161/01.STR.24.1.35
18. Banerjee C, Turan TN. Large artery atherosclerosis: extracranial and intracranial. *Semin Neurol*. 2017;37:307–315. doi: 10.1055/s-0037-1603588
19. Wardlaw JM, Smith C, Dichgans M. Small vessel disease: mechanisms and clinical implications. *Lancet Neurol*. 2019;18:684–696. doi: 10.1016/S1474-4422(19)30079-1
20. Kamel H, Healey JS. Cardioembolic stroke. *Circ Res*. 2017;120:514–526. doi: 10.1161/CIRCRESAHA.116.308407
21. Bigdeli TB, Lee D, Webb BT, Riley BP, Vladimirov VI, Fanous AH, Kendler KS, Bacanu SA. A simple yet accurate correction for winner's curse can predict signals discovered in much larger genome scans. *Bioinformatics (Oxford, England)*. 2016;32:2598–2603. doi: 10.1093/bioinformatics/btw303
22. Burgess S, Thompson SG. Avoiding bias from weak instruments in Mendelian randomization studies. *Int J Epidemiol*. 2011;40:755–764. doi: 10.1093/ije/dyr036
23. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet Epidemiol*. 2013;37:658–665. doi: 10.1002/gepi.21758
24. Bowden J, Holmes MV. Meta-analysis and Mendelian randomization: a review. *Res Synth Methods*. 2019;10:486–496. doi: 10.1002/jrsm.1346
25. Burgess S, Thompson SG. *Mendelian Randomization: Methods for Causal Inference Using Genetic Variants*. 2nd ed. Chapman and Hall/CRC; 2021.
26. Burgess S, Bowden J. Integrating summarized data from multiple genetic variants in Mendelian randomization: bias and coverage properties of inverse-variance weighted methods. *arXiv*. 2015. https://arxiv.org/abs/1512.04486.
27. Davies NM, Von Hinke Kessler Scholder S, Farbmacher H, Burgess S, Windmeijer F, Smith GD. The many weak instruments problem and Mendelian randomization. *Stat Med*. 2015;34:454–468. doi: 10.1002/sim.6358
28. Verbanck M, Chen CY, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat Genet*. 2018;50:693–698. doi: 10.1038/s41588-018-0099-7
29. Zhao QY, Wang JS, Hemani G, Bowden J, Small DS. Statistical inference in two-sample summary-data Mendelian randomization using robust adjusted profile score. *Ann Stat*. 2020;48:1742–1769. doi: 10.1214/19-AOS1866
30. Ong JS, MacGregor S. Implementing MR-PRESSO and GCTA-GSMR for pleiotropy assessment in Mendelian randomization studies from a practitioner's perspective. *Genet Epidemiol*. 2019;43:609–616. doi: 10.1002/gepi.22207
31. Burgess S, Thompson SG. Interpreting findings from Mendelian randomization using the MR-Egger method. *Eur J Epidemiol*. 2017;32:377–389. doi: 10.1007/s10654-017-0255-x
32. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol*. 2015;44:512–525. doi: 10.1093/ije/dyv080
33. Braunschweig U, Gueroussov S, Plocik AM, Graveley BR, Blencowe BJ. Dynamic integration of splicing within gene regulatory pathways. *Cell*. 2013;152:1252–1269. doi: 10.1016/j.cell.2013.02.034
34. Barbosa-Morais NL, Irimia M, Pan Q, Xiong HY, Gueroussov S, Lee LJ, Slobodeniuc V, Kutter C, Watt S, Colak R, et al. The evolutionary landscape of alternative splicing in vertebrate species. *Science (New York, N.Y.)*. 2012;338:1587–1593.
35. Chen XL, Li Q, Huang WS, Lin YS, Xue J, Wang B, Jin KL, Shao B. Serum YKL-40, a prognostic marker in patients with large-artery atherosclerotic stroke. *Acta Neurol Scand*. 2017;136:97–102. doi: 10.1111/ane.12688
36. Malinda KM, Ponce L, Kleinman HK, Shackelton LM, Millis AJ. Gp38k, a protein synthesized by vascular smooth muscle cells, stimulates directional migration of human umbilical vein endothelial cells. *Exp Cell Res*. 1999;250:168–173. doi: 10.1006/excr.1999.4511
37. Badimon L, Vilahur G. Thrombosis formation on atherosclerotic lesions and plaque rupture. *J Intern Med*. 2014;276:618–632. doi: 10.1111/joim.12296
38. Derdeyn CP. Mechanisms of ischemic stroke secondary to large artery atherosclerotic disease. *Neuroimaging Clin N Am*. 2007;17:303–311. doi: 10.1016/j.nic.2007.03.001
39. Huan W, Yandong L, Chao W, Sili Z, Jun B, Mingfang L, Yu C, Lefeng Q. YKL-40 aggravates early-stage atherosclerosis by inhibiting macrophage apoptosis in an aven-dependent way. *Front Cell Dev Biol*. 2021;9:752773. doi: 10.3389/fcell.2021.752773

40. Tsantilas P, Lao S, Wu ZY, Eberhard A, Winski G, Vaerst M, Nanda V, Wang Y, Kojima Y, Ye JQ, et al. Chitinase 3 like 1 is a regulator of smooth muscle cell physiology and atherosclerotic lesion stability. *Cardiovasc Res*. 2021;117:2767–2780. doi: 10.1093/cvr/cvab014
41. Ebrahim S, Davey Smith G. Mendelian randomization: can genetic epidemiology help redress the failures of observational epidemiology? *Hum Genet*. 2008;123:15–33. doi: 10.1007/s00439-007-0448-6
42. Burgess S, Davey Smith G, Davies NM, Dudbridge F, Gill D, Glymour MM, Hartwig FP, Holmes MV, Minelli C, Relton CL, et al. Guidelines for performing Mendelian randomization investigations. *Wellcome Open Res*. 2020;4:186. doi: 10.12688/wellcomeopenres.15555.2
43. Bhardwaj R, Yester JW, Singh SK, Biswas DD, Surace MJ, Waters MR, Hauser KF, Yao Z, Boyce BF, Kordula T. Relb/p50 complexes regulate cytokine-induced YKL-40 expression. *J Immunol*. 2015;194:2862–2870.