Genetic Architecture of Stroke of Undetermined Source: Overlap with Known Stroke Etiologies and Associations with Modifiable Risk Factors

Marios K. Georgakis, MD, PhD,[1,2,3] Livia Parodi, PhD,[1,2] Simon Frerich, MSc,[3] Ernst Mayerhofer, MD,[1,2] Georgios Tsivgoulis, MD,[4,5] James P. Pirruccello, MD,[6,7,8] Agnieszka Slowik, MD, PhD,[9] Tatjana Rundek, MD, PhD,[10] Rainer Malik, PhD,[3] Martin Dichgans, MD,[3,11,12] Jonathan Rosand, MD,[1,2,13,14] Christopher D. Anderson, MD,[1,2,14,15] and NINDS Stroke Genetics Network (SiGN)

Objective: Ischemic stroke etiology remains undetermined in 30% of cases. We explored the genetic architecture of stroke classified as undetermined to test if mechanisms and risk factors underlying large-artery atherosclerotic (LAAS), cardioembolic (CES), and small-vessel stroke (SVS) contribute to its pathogenesis.
Methods: We analyzed genome-wide data from 16,851 ischemic stroke cases and 32,473 controls. Using polygenic risk scores for LAAS, CES, and SVS, we assessed the genetic overlap with stroke of undetermined source and used pairwise genomewide association study (GWAS-PW) to search for shared loci. We then applied Mendelian randomization (MR) to identify potentially causal risk factors of stroke of undetermined source.
Results: Genetic risk for LAS, CES, and SVS was associated with stroke of undetermined source pointing to overlap in their genetic architecture. Pairwise analyses revealed 19 shared loci with LAAS, 2 with CES, and 5 with SVS that have been implicated in atherosclerosis-related phenotypes. Genetic liability to both carotid atherosclerosis and atrial fibrillation was associated with stroke of undetermined source, but the association with atrial fibrillation was attenuated after excluding cases with incomplete diagnostic workup. MR analyses showed effects of genetically determinants of blood pressure, diabetes, waist-to-hip ratio, inflammatory pathways (IL-6 signaling, MCP-1/CCL2 levels), and factor XI levels on stroke of undetermined source.

View this article online at wileyonlinelibrary.com. DOI: 10.1002/ana.26332

Received Oct 5, 2021, and in revised form Feb 16, 2022. Accepted for publication Feb 16, 2022.

Address correspondence to Dr Georgakis, Center for Genomic Medicine, Massachusetts General Hospital, 185 Cambridge Street, Boston, MA 02114, USA. E-mail: mgeorgakis@mgh.harvard.edu, marios.georgakis@med.uni-muenchen.de

From the [1]Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA; [2]Program in Medical and Population Genetics, Broad Institute of Harvard and the Massachusetts Institute of Technology, Boston, MA, USA; [3]Institute for Stroke and Dementia Research (ISD), University Hospital, Ludwig-Maximilians-University (LMU) Munich, Munich, Germany; [4]Second Department of Neurology, Attikon University Hospital, School of Medicine, National and Kapodistrian University of Athens, Athens, Greece; [5]Department of Neurology, University of Tennessee Health Sciences Center, Memphis, TN, USA; [6]Cardiology Division, Massachusetts General Hospital, Boston, MA, USA; [7]Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA, USA; [8]Cardiovascular Disease Initiative, Broad institute of Harvard and the Massachusetts Institute of Technology, Cambridge, MA, USA; [9]Department of Neurology, Jagiellonian University, Krakow, Poland; [10]Department of Neurology, Evelyn F. McKnight Brain Institute, Miller School of Medicine, University of Miami, Miami, FL, USA; [11]Munich Cluster for Systems Neurology (SyNergy), Munich, Germany; [12]German Centre for Neurodegenerative Diseases (DZNE), Munich, Germany; [13]Department of Neurology, Massachusetts General Hospital, Boston, MA, USA; [14]McCance Center for Brain Health, Massachusetts General Hospital, Boston, MA, USA; and [15]Department of Neurology, Brigham and Women's Hospital, Boston, MA, USA

Additional supporting information can be found in the online version of this article.

Interpretation: Stroke of undetermined source shares genetic and vascular risk factors with other stroke subtypes, especially LAAS, thus highlighting the diagnostic limitations of current subtyping approaches. The potentially causal associations with carotid atherosclerosis and atherosclerotic risk factors might have implications for prevention strategies targeting these mechanisms.

ANN NEUROL 2022;91:640–651

Stroke is the second leading cause of death and the third leading cause of disability worldwide.[1] Diagnostic workup following a stroke is focused on identifying the underlying etiology to pursue focused secondary prevention. Still, in one third of all cases of ischemic stroke, the underlying etiology remains undetermined.[2] These patients do not receive targeted secondary preventive treatments, although stroke recurrence rates are similar to those of other stroke subtypes.[2,3] The current prevailing hypothesis is that most ischemic strokes of undetermined source result from emboli of arterial origin from nonstenosis atherosclerotic lesions, emboli of cardiac origin due to paroxysmal atrial fibrillation, or venous thromboembolism leading to paradoxical emboli due to patent foramen ovale.[4] Still, even after extensive diagnostic workup, no source can be found for the large proportion of these events.

Previous studies have explored potentially underlying causes of stroke of undetermined source. For example, it has been suggested that carotid stenoses less advanced than the cutoff of 50% required to define a stroke as of atherosclerotic origin, might represent unrecognized sources of embolism. Unstable plaques with specific morphological characteristics, even when not causing a stenosis, are more common in patients with stroke of undetermined source than in patients with cardioembolic sources.[5,6] Paroxysmal atrial fibrillation that remains undetected in the acute phase of stroke might represent another stroke source.[7] Implantation of a cardiac event-triggered monitor can detect more paroxysmal atrial fibrillation episodes in patients with stroke of undetermined source,[8] but it remains uncertain if these episodes have a causal relationship with the original event.[9] Randomized controlled trials have shown that oral anticoagulants are not superior to antiplatelet agents for secondary prevention after stroke of undetermined source.[10,11]

The limited existing data regarding the etiology of stroke of undetermined source arise from observational studies, which are prone to confounding. Examining human genetics may be a valuable resource in elucidating the mechanisms of stroke of undetermined source. Using causal inference methods, such as Mendelian randomization (MR), conclusions can be drawn about the potential causes of stroke of undetermined source, as well as the effects of modifiable risk factors and potentially druggable mechanisms that could guide the design of future trials.[12,13] Here, we analyzed genetic data from 16,851 cases of ischemic stroke including up to 4,755 strokes of undetermined source, and 32,473 controls with the following aims: (1) to explore if there is a genetic overlap between stroke of undetermined source and the most common subtypes of defined etiology (large-artery atherosclerotic stroke [LAAS], cardioembolic stroke [CES], small-vessel stroke [SVS]); (2) to identify genetic loci that jointly influence risk of stroke of undetermined source and defined stroke subtypes that could point to specific pathophysiological mechanisms; (3) to use MR to test whether the most common stroke etiologies (carotid atherosclerosis, atrial fibrillation, and cerebral small vessel disease) also causally influence risk of stroke of undetermined source; and (4) to explore the risk factor profile of stroke of undetermined source.

Methods

Study Population: The Stroke Genetics Network

For this study, we used data from the European dataset of the National Institute of Neurological Disorders and Stroke (NINDS)-Stroke Genetics Network (SiGN), a collaborative genomewide association study (GWAS) meta-analysis of ischemic stroke, initiated within the International Stroke Genetics Consortium (ISGC). SiGN includes a network of 31 cohorts involving 16,851 cases of ischemic stroke and 32,473 stroke-free controls. A detailed description of the project is available elsewhere.[14] In SiGN, ischemic stroke cases have been classified according to the Trial of org 10,172 in Acute Stroke Treatment (TOAST) system[15] and the Causative Classification of Stroke (CCS) system.[16] The TOAST classification is available for 74% of the cases, whereas CCS is available for all cases. In both TOAST and CCS, ischemic stroke cases are classified as LAAS, CES, SVS, stroke of undetermined source, and stroke of other determined causes. Our analysis focused on stroke of undetermined source. Stroke of undetermined source according to TOAST was defined as brain infarction that is not attributable to a source of definite cardioembolism, large artery atherosclerosis, or small artery disease despite standard vascular, cardiac, and serologic evaluation.[15] The category of stroke of undetermined etiology in the TOAST classification includes cases with less well-established potential causes of cardiac embolism, such as patent foramen ovale,

aortic arch atheroma, and mitral valve strands, as well as potential prothrombotic disorders, as well as stroke of undetermined cases with two or more equally plausible identified causes of stroke.[15] We used both classification systems, because TOAST is widely used in the clinic and CCS prioritizes the most probable cause when multiple causes are present, thus including fewer cases with competing etiologies in the category of undetermined source. CCS-defined cases of undetermined source were further subclassified into (i) cases with incomplete diagnostic workup or unclassified cases due to presence of more than one competing etiologies (incomplete diagnostics/unclassified) and (ii) cases with no identifiable cause after complete diagnostic workup or with identified minor cardioembolic sources, such as patent foramen ovale, mitral annular calcification, atrial septal aneurysm, and complex aortic atheroma (cryptogenic/minor cardioembolic). These definitions were used for sensitivity analyses.

Standard Protocol Approvals and Patient Consent

All participating cohorts in SiGN received ethical approval from the appropriate local committees. All patients included in the studies have provided informed consent for their participation.

Genetic Overlap Between Defined Stroke Subtypes and Stroke of Undetermined Source

We explored the genetic overlap throughout the genome between defined stroke subtypes and stroke of undetermined source using 2 methods. First, using summary statistics from the MEGASTROKE dataset of individuals of European ancestry (4,373 LAAS cases, 7,193 CES cases, 5,386 SVS cases, and 406,111 controls),[17] we calculated a *z*-score for each single-nucleotide polymorphism (SNP) included in the GWASs for LAAS, CES, and SVS, as well as for the strokes of undetermined source in the SiGN datasets.[18] Following clumping for linkage disequilibrium (LD) at $r^2 < 0.1$ (according to the European 1,000 Genomes phase III panel), we calculated the Pearson's r between z-scores of the respective datasets at different p value thresholds ($p < 5 \times 10^{-8}$, $p < 10^{-6}$, $p < 10^{-4}$, and $p < 0.05$). Second, we constructed polygenic risk scores (PRS) for LAAS, CES, and SVS using log-transformed odds ratios (ORs) derived from the more recent MEGASTROKE dataset.[17] We selected genetic variants to be included in the PRSs on the basis of 4 different p value thresholds ($p < 5\text{x}10^{-8}$, $p < 10^{-6}$, $p < 10^{-4}$, and $p < 0.05$) after clumping LD at $r^2 < 0.1$. In order to explore whether genetic predisposition to defined stroke subtypes also influences the risk of stroke of undetermined source, we then explored the effects of the PRSs of the defined stroke subtypes on risk of ischemic stroke of undetermined source in SiGN summary statistics, as previously described.[19] For both methods, we corrected for multiple testing with the false discovery rate (FDR) approach and set statistical significance at an FDR-adjusted p value < 0.05.

Pairwise Genomewide Association Analyses

Because we found significant correlations between defined stroke subtypes and stroke of undetermined source at a genomewide level, as a next step, we aimed to identify specific genomic loci that jointly contribute to risk of stroke of undetermined source and risk of defined stroke subtypes that could point to hints about the shared biology that underlies these associations. We used the pairwise GWAS (GWAS-PW) software tool, which explored in an unbiased and explorative way the colocalization between individual loci across the genome.[20] Specifically, this Bayesian method uses summary statistics for pairs of traits and calculates the posterior probabilities of association (PPA) that LD-independent genomic regions are specifically influencing only 1 of the 2 traits (models 1 and 2), jointly both traits (model 3), or both traits but due to different variants (model 4). Genomic regions with PPA ≥0.9 in model 3 were considered to jointly influence both traits under study (pleiotropy effect), whereas regions with a PPA ≥0.6 were considered suggestive of a joint influence. We did not perform alternative methods, such as MTAG[21] or genomic SEM,[22] because they are based on the assumption that the examined pairs of traits represent different manifestations of a common underlying mechanism, which in our analysis remains unclear. For the defined stroke subtypes, the MEGASTROKE data from Europeans were used as an input,[17] whereas the SiGN data were used for stroke of undetermined source. Because the controls of the analysis performed in SiGN were all also included in the MEGASTROKE data, there was a substantial control overlap between the compared phenotypes. According to the description of the original method,[20] we calculated between each pair of traits, the expected correlation because of this overlap under the null model, and added this as a correction factor in the respective GWAS-PW analysis.[18]

Gene-Based, Transcription-Based, and Colocalization Analyses

To functionally characterize the regions identified by GWAS-PW, we performed gene-based association analyses for the identified loci using MAGMA (version 1.09a [mac]).[23] Briefly, this method uses p values as input and the test statistic is a linear combination of gene p values.

We used the default gene-based test, which is the mean of the χ^2 for the variants annotated to each gene. In order to account for LD, the 1,000 genomes phase III European panel was used to derive variant-wise LD such that the null distribution can be approximated. In regions reaching a PPA ≥0.6 in model 3 of GWAS-PW, variants were mapped to 436 autosomal protein-coding genes, obtained from the MAGMA website (https://ctg.cncr.nl/software/magma). To account for multiple testing, we considered significant genes reaching a $p < 0.05/436 = 1.15 \times 10^{-4}$ (Bonferroni method).

Next, to explore whether the signals derived from the GWAS-PW analyses could be explained by effects on gene expression, we performed a transcription-based association analysis using S-PrediXcan,[24] as well as an expression enriched expansion of MAGMA, E-MAGMA.[25] S-PrediXcan estimates gene expression by using tissue-dependent prediction models trained in the reference data, and then correlates the estimated gene expression with genetic associations with the trait under study. E-MAGMA uses the interface of MAGMA but assigns risk variants to their putative genes based on tissue-specific expression quantitative trait locus (eQTL) information. We integrated transcriptome data from expression models based on the GTEx project version 7p (PredictDB; http://predictdb.org) from arterial tissues, heart tissues, and blood (aorta, tibial artery, coronary artery, left ventricle, left atrial appendage, and whole blood). We tested associations with all 436 genes located in the regions identified by GWAS-PW. Statistical significance was set at $p < 0.05/436 = 1.15 \times 10^{-4}$ (Bonferroni method).

Finally, a colocalization analysis[26] was carried out at each locus identified by GWAS-PW to estimate the PPA of a shared causal variant between gene expression at the abovementioned tissues and association with stroke of undetermined source. PPA ≥0.6 was considered significant.

Mendelian Randomization

To explore whether the risk for stroke of undetermined source can be partially explained by established stroke etiologies, we explored in MR analyses whether genetic predisposition to carotid atherosclerosis (proxied by common carotid artery intima media thickness [cIMT] in 71,128 individuals of European ancestry from the CHARGE Consortium[27]), a main cause of LAAS, to atrial fibrillation (65,446 cases, 522,744 controls from a trans-ethnic meta-analysis of the AFGen Consortium, the UK Biobank, the Broad AF Study, and the Biobank Japan[28]), the primary source of CES, and to cerebral small vessel disease (proxied by white matter hyperintensities [WMHs] volume in 42,310 individuals in the UK Biobank[29]), the main cause of SVS, is associated with stroke of undetermined source in the SiGN data. As a less established cause of cardioembolism, we also explored associations with genetically predicted left atrial structural traits,[30] as derived from deep learning algorithms in cardiac magnetic resonance imaging (MRI) data (40,558 individuals from the UK Biobank[31]).

To explore the risk factor profile of stroke of undetermined source, we performed MR analyses between a set of vascular risk factors known to influence stroke risk and stroke of undetermined source. The vascular risk factors were selected on the basis of known associations with defined stroke subtypes both in observational studies and MR analyses[13] with the aim to explore whether the risk factor profile of stroke of undetermined source resembles that of other defined subtypes. On the basis of this rationale, the risk factors we explored included: systolic and diastolic blood pressure (757,601 individuals of European ancestry from the meta-analysis of ICBP Consortium and the UK Biobank),[32] circulating LDL- and HDL-cholesterol levels (617,303 individuals from the trans-ethnic meta-analysis of the MVP data and the GLGC Consortium),[33] type 2 diabetes mellitus (74,124 cases, 824,006 controls of European ancestry from the DIAGRAM Consortium),[34] glycated hemoglobin A1C (based on 421,923 individuals from the UK Biobank data),[35] a lifetime smoking index (462,690 individuals from the UK Biobank),[36] body mass index (681,275 individuals of European ancestry from a meta-analysis of the GIANT Consortium and the UK Biobank),[37] waist-to-hip ratio (694,649 individuals of European ancestry from a meta-analysis of the GIANT Consortium and the UK Biobank), interleukin-6 (IL-6) signaling (522,681 individuals from a meta-analysis of the CHARGE Consortium and the UK Biobank),[38,39] circulating monocyte chemoattractant protein-1/CC-chemokine ligand-2 (MCP-1/CCL2) levels (8,293 individuals of Finnish ancestry),[40] and factor XI levels (16,169 individuals of European ancestry).[41] For lipoprotein-(a) levels, IL-6 signaling activity, MCP-1 levels, and factor XI levels, we used previously described genetic instruments.[38–42] Genetic instruments for all other exposures were selected based on genomewide significant associations ($p < 5 \times 10^{-8}$) and clumped for LD at $r^2 < 0.001$. The variants used as instruments are available in Supplementary Table S1.

We applied 2-sample MR analyses based on variant associations derived from the abovementioned sources. Following extraction of the association estimates and harmonization of their direction-of-effect alleles, we computed MR estimates with fixed-effects inverse-variance weighted (IVW) analyses. For all analyses, we set statistical significance at an FDR-adjusted p value < 0.05. This was preferred over a more stringent Bonferroni-based

threshold because of the high overlap between both the examined exposures (eg, systolic and diastolic blood pressure) and the examined outcomes (eg, 4 different versions of stroke of undetermined source). Furthermore, although we also present associations with defined stroke subtypes for comparison, these were not the focus of the current work, and as such a Bonferroni-based approach taking all these comparisons into account would lead to overcorrection and underestimation of potentially important relationships. Associations not reaching this threshold, but showing a $p < 0.05$ were considered suggestive of an association. The IVW method was our primary MR analysis approach,[43] but the derived estimates could be biased in the case of directional pleiotropy. As a measure of pleiotropy, we assessed heterogeneity across the MR estimates for each instrument in the IVW MR analyses with the Cochran's Q statistic ($p < 0.05$ was considered significant).[44] To address the presence of potential directional pleiotropy in our results, we further applied alternative MR methods, which are more robust to the use of pleiotropic instruments: the weighted median estimator allows the use of invalid instruments under the assumption that at least half of the instruments used in the MR analysis are valid[45]; MR-Egger regression allows for the estimation of an intercept term, provides less precise estimates, and relies on the assumption that the strengths of potential pleiotropic instruments are independent of their direct associations with the outcome.[46] The intercept obtained from MR-Egger regression was used as a measure of directional pleiotropy ($p < 0.05$ indicated significance).[46] MR analyses were performed in R (version 3.5.0; The R Foundation for Statistical Computing) using the MendelianRandomization package.

Data Availability

The data analyzed in the current study are publicly available. Genetic instruments used for the MR analyses are provided as Supplementary Material.

Results

The study hypotheses, the examined research questions, and the study design are graphically presented in Figure 1. A total of 3,593 and 4,755 patients with ischemic stroke of undetermined source defined by TOAST and CCS, respectively, were included in SiGN. The CCS-defined cases were further subclassified into those with either "incomplete" diagnostic workup or of "unclassified" etiology due to presence of multiple competing causes (n = 2,310) and to "cryptogenic/minor cardioembolic" strokes without any evidence of other probable etiology after complete diagnostic workup (n = 2,392). A total of 32,473 stroke-free individuals served as controls.

Genetic Overlap Between Stroke of Undetermined Source and Other Stroke Subtypes

First, we explored the genetic correlation between defined stroke subtypes in MEGASTROKE (LAAS, CES, and SVS) and stroke of undetermined source in SiGN. There were high correlations between z-scores of the SNPs associated with determined stroke subtypes and the respective z-scores for stroke of undetermined source, as defined according to both the TOAST and the CCS definitions. Correlations were more prominent for variants selected to be associated with defined stroke subtypes at lower p value thresholds (for $p < 5 \times 10^{-8}$ Pearson's r: 0.97 between TOAST-defined stroke of undetermined source and LAAS

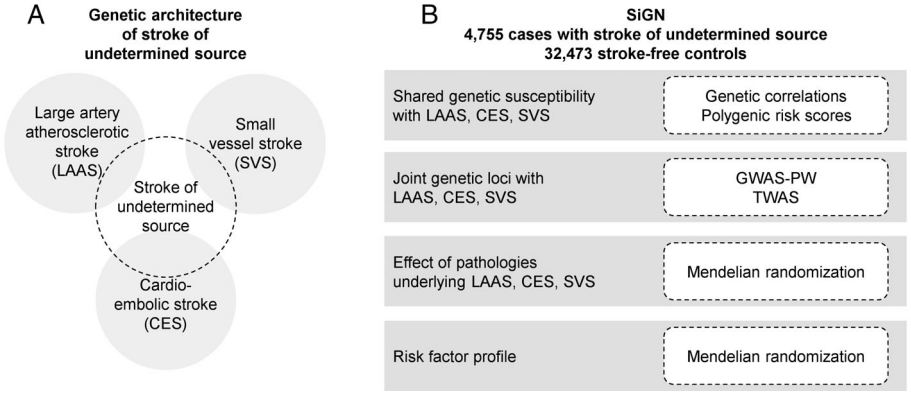

FIGURE 1: Schematic representation of the study hypothesis, research questions, and applied methods. (A) Our hypothesis is that the mechanisms underlying the most common defined stroke subtypes (LAAS, CES, SVS) also contribute to the risk of stroke of undetermined source and as such their genetic architecture would be overlapping. (B) Using data from the NINDS Stroke Genetics Network (SiGN) we explored (i) the genetic overlap between stroke of undetermined source and LAAS, CES, and SVS, (ii) if there are shared genetic loci between stroke of undetermined source and LAAS, CES, and SVS, (iii) the causal effect of pathologies underlying LAAS, CES, and SVS on risk of stroke of undetermined source, and (iv) the causal effects of conventional modifiable vascular risk factors on stroke of undetermined source. CES = cardioembolic stroke; GWAS-PW = pairwise genomewide association study; LAAS = large-artery atherosclerotic stroke; SVS = small-vessel stroke.

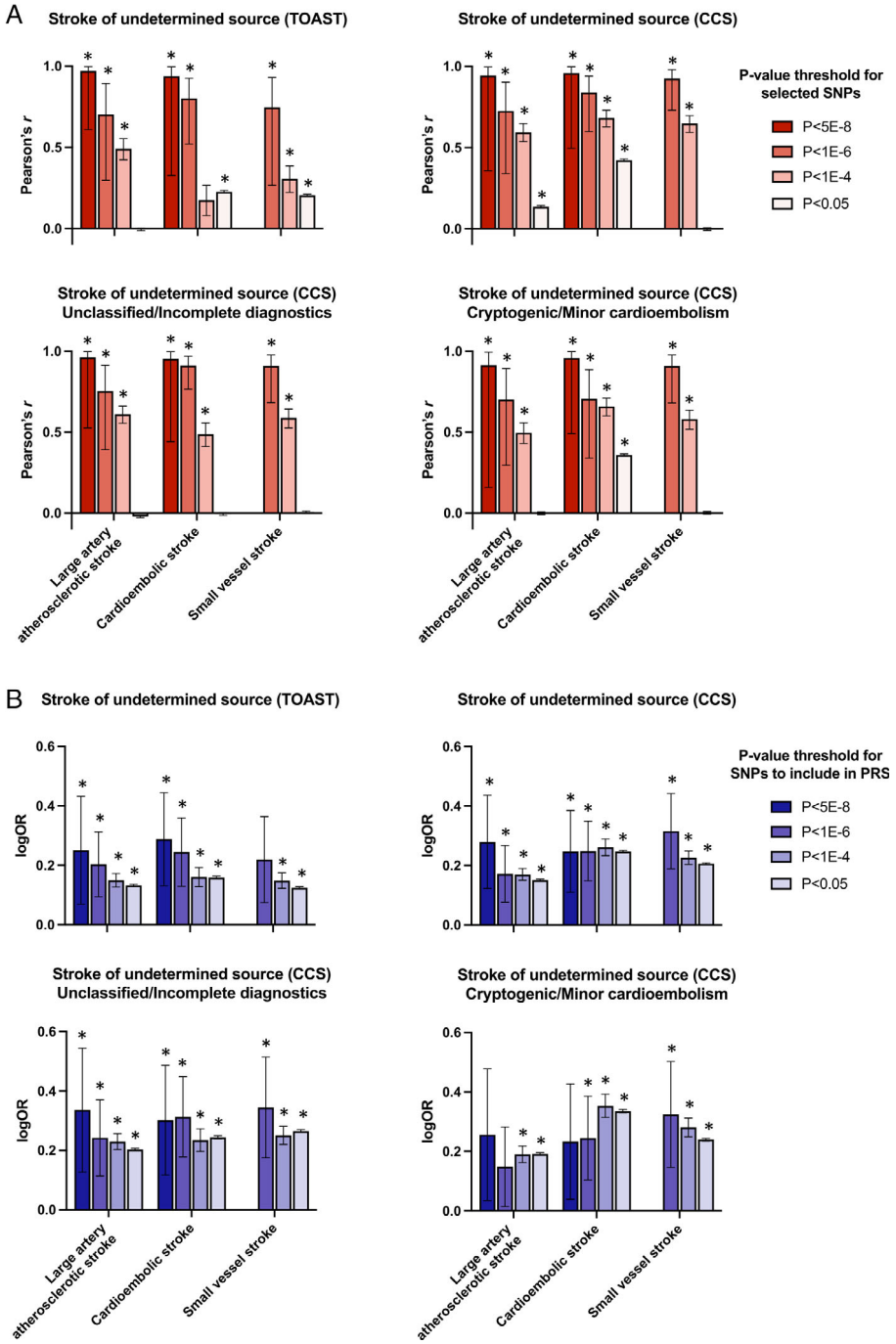

FIGURE 2: Shared genetic susceptibility between defined stroke subtypes and stroke of undetermined source. (A) Shown are Pearson's *r* correlation coefficients between the z-score of the associations of individual variants with defined stroke subtypes and those with ischemic stroke of undetermined source according to the TOAST and CCS criteria, as well as the subgroups of the CCS-defined cases: ischemic stroke of undetermined source due to incomplete diagnostic workup or competing etiologies, and cryptogenic or minor cardioembolic stroke. (B) Polygenic risk scores (PRS) for defined stroke subtypes in association with ischemic stroke of undetermined source. Different sets of genetic variants associated with the determined stroke subtypes at different *p* value levels (different red shades) are included in the analyses. The variants are selected from the MEGASTROKE data from individuals of European ancestry following clumping for linkage disequilibrium at $r^2 < 0.1$. *FDR-adjusted *p* value < 0.05. CCS = Causative Classification of Stroke; FDR = false discovery rate; SNP = single-nucleotide polymorphism; TOAST = Trial of org 10,172 in Acute Stroke Treatment. [Color figure can be viewed at www.annalsofneurology.org]

and 0.93 with CES, for $p < 1 \times 10^{-6}$: Pearson's *r*: 0.75 for SVS; Fig 2A). Similarly, higher PRSs for stroke of determined etiology were strongly associated with higher odds for stroke of undetermined source indicating overlapping genetic susceptibility (the R^2 in genetic susceptibility of TOAST-defined stroke of undetermined source

by PRSs constructed by variants associated with LAAS, CES, and SVS at $p < 1 \times 10^{-6}$ were: 0.31, 0.24, and 0.26, respectively; Fig 2B). Similar results were obtained when examining the CCS-defined subgroups of incomplete diagnostics/unclassified and cryptogenic/minor cardioembolic stroke of stroke of undetermined source (see Fig 2).

Shared Genetic Loci Between Stroke of Undetermined Source and Other Stroke Subtypes

We next explored whether there were genetic loci jointly influencing the risk of stroke of undetermined source and defined stroke subtypes using GWAS-PW. We found 2 loci reaching a PPA ≥0.9 and 17 a PPA ≥0.6 for a shared association between LAAS and any of the 4 phenotypes of strokes of undetermined source (Fig 3A, and Supplementary Table S2). Additionally, 2 and 5 loci reached a PPA ≥0.6 for a shared association of CES and SVS with stroke of undetermined stroke, respectively (see Fig 3A, and Supplementary Table S2).

To functionally characterize these shared loci, we performed gene-based and transcriptome-based analyses using eQTL data from 6 vasculature-relevant tissues (aorta, tibial artery, coronary artery, left ventricle, left atrial appendage, and whole blood; Fig 3B, and Supplementary Tables S3 and S4). We found 7 significant gene expression/trait associations for ischemic stroke of undetermined source in 4 genes: *ALDH2* in whole blood (joint signals between stroke of undetermined source and both LAAS and SVS), *PDE8B* in aorta (joint signal with LAAS), *NBEAL* 1 in aorta, tibial artery, and coronary artery, and *CARF* in whole blood (joint signals with SVS). For all genes, there was some evidence of significant colocalization between expression levels in the relevant tissues and associations with strokes of undetermined source (see Fig 3B).

We explored locus-specific pleiotropic associations with other phenotypes in the PhenoScanner database (Supplementary Table S5). We found that lead SNPs in the shared loci were associated (at $p < 1 \times 10^{-5}$) with atherosclerosis-related outcomes (mainly coronary artery disease manifestations), thrombotic complications (such as venous thromboembolism and pulmonary embolism), vascular risk factors (hypertension, cholesterol levels, and diabetes), obesity-related traits, as well as platelet, red blood cell, and leukocyte counts in peripheral blood (Fig 3C).

Genetic Predisposition to Established Causes of Stroke and Stroke of Undetermined Source

Using MR, we next tested whether genetic predisposition to carotid atherosclerosis, atrial fibrillation, and cerebral small vessel disease, three main causes of LAAS, CES, and SVS, respectively, are associated with stroke of

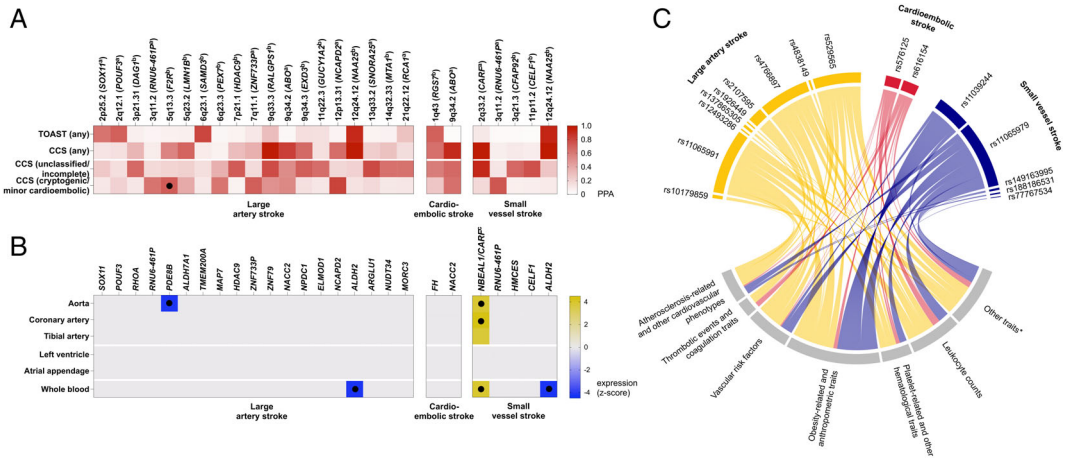

FIGURE 3: Locus-level genomic overlap between defined stroke subtypes and ischemic stroke of undetermined source and functional relevance of the shared loci. (A) Loci reaching a posterior probability of association (PPA) >0.6 in pairwise GWAS analyses (GWAS-PW) between determined stroke subtypes and at least one of the categories of stroke of undetermined source. The lead genetic variants (with the lowest *p* value at each locus for association with stroke of undetermined source) were annotated to the closest gene (A) if in non-coding regions or to their respective genes using MAGMA (B). Dots represent a significant association in the gene-based MAGMA analysis. (B) Transcription-based analysis for the identified loci using summary statistics from the strongest-associated subtype of stroke of undetermined source and expression data from GTEx for 6 relevant tissues, as performed in S-PrediXcan. Only statistically significant results are depicted. Dots represent statistically significant results (PPA ≥0.6) from colocalization analysis. The cNBEAL1 showed statistically significant association for aorta, coronary artery, and tibial artery, whereas CARF was significant in whole blood. (C) Associations of the lead SNPs in the identified loci with other phenotypes in the Phenoscanner database. Associations of $p < 1 \times 10\text{-}5$ or lower are depicted. The individual SNPs are clustered by associations with the determined stroke subtypes. CCS = Causative Classification of Stroke; CI = confidence interval; TOAST = Trial of org 10,172 in Acute Stroke Treatment. [Color figure can be viewed at www.annalsofneurology.org]

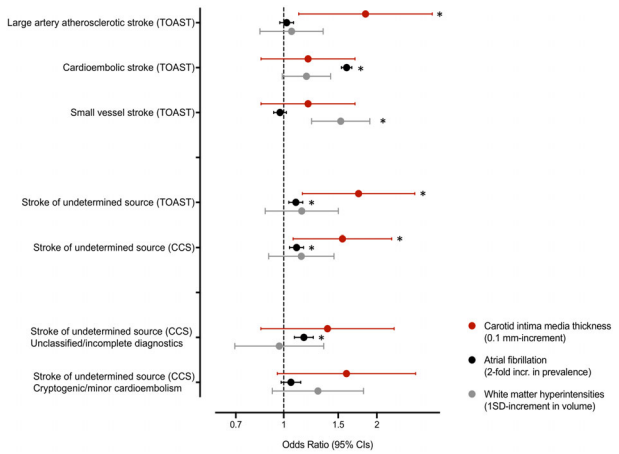

FIGURE 4: Mendelian randomization associations between genetic predisposition to established risk factors for ischemic stroke with defined stroke subtypes and stroke of undetermined source. Associations with determined ischemic stroke subtypes (TOAST-defined), ischemic stroke of undetermined sources according to the TOAST and CCS criteria, as well as the subgroups of the CCS-defined cases (ischemic stroke of undetermined source due to incomplete diagnostic workup or competing etiologies, and cryptogenic or minor cardioembolic stroke) are shown, according to the inverse-variance weighted model. * FDR-adjusted *p* value <0.05. CCS = Causative Classification of Stroke; FDR = false discovery rate; TOAST = Trial of org 10,172 in Acute Stroke Treatment. [Color figure can be viewed at www.annalsofneurology.org]

undetermined source. We found higher genetically predicted cIMT, as well as genetic liability to atrial fibrillation to be associated with stroke of undetermined source. These effects held for subtypes defined according to TOAST (OR per 0.1-mm-increment in cIMT: 1.75, [1.18–2.58], OR per 2-fold-increase in prevalence of atrial fibrillation: 1.09 [1.04–1.15]) and CCS criteria (OR per 0.1-mm-increment in cIMT: 1.55, [1.09–2.20], OR per 2-fold-increase in prevalence of atrial fibrillation: 1.10 [1.04–1.15]; Fig 4, and Supplementary Table S6). There was no evidence of an effect of genetically predicted WMH volume on stroke of undetermined source. Of note, across the subgroups of CCS-defined stroke of undetermined source, there was significantly greater effect of genetic liability to atrial fibrillation on stroke of undetermined source due to incomplete diagnostic workup or competing etiologies, but no effect on cryptogenic/minor cardioembolic stroke. The effects of genetically predicted cIMT were not significant for these 2 subgroups, but consistent in terms of directionality and magnitude with those of the larger categories (see Fig 4). There was no evidence of an effect of genetically predicted MRI-derived left atrial structural traits, and any of the undetermined stroke subtypes (Supplementary Table S7). The effects were consistent when using the median approach and MR Egger that are more robust to pleiotropy and there was no significant heterogeneity in the performed analyses (Supplementary Table S6).

Genetic Predisposition to Modifiable Risk Factors and Stroke of Undetermined Source

As a last step, we used MR to determine the risk factor profile for stroke of undetermined source, and explore causal links with potentially druggable mechanisms (Fig 5, and Supplementary Table S8). Similarly to defined stroke subtypes, higher genetically predicted systolic and diastolic blood pressures were associated with higher odds of stroke

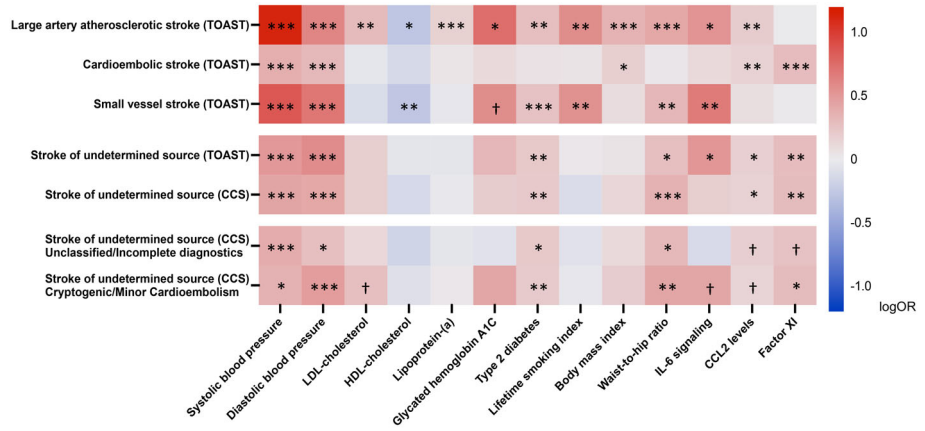

FIGURE 5: Mendelian randomization associations between genetically predicted vascular risk factors both defined stroke subtypes and stroke of undetermined source. Associations between determined ischemic stroke subtypes (TOAST-defined), ischemic stroke of undetermined sources according to the TOAST and CCS criteria, as well as the subgroups of the CCS-defined cases (ischemic stroke of undetermined source due to incomplete diagnostic workup or competing etiologies, and cryptogenic or minor cardioembolic stroke) are shown, according to the inverse-variance weighted model. Results are scaled to 1-SD increment of the exposure for continuous traits and to a 2-fold increase in prevalence for binary traits. † *p* < 0.05; * FDR-adjusted *p* value <0.05; ** FDR-adjusted *p* value <0.01; * FDR-adjusted *p* value <0.001. CCS = Causative Classification of Stroke; FDR = false discovery rate; TOAST = Trial of org 10,172 in Acute Stroke Treatment. [Color figure can be viewed at www.annalsofneurology.org]**

of undetermined source, when defined by either TOAST (OR per 1-SD-increment in systolic blood pressure: 1.65, [95% confidence interval [CI] = 1.34–2.04], OR per 1-SD-increment in diastolic blood pressure: 1.76 [95% CI = 1.46–2.13]), or CCS criteria (OR per SD-increment in systolic blood pressure: 1.53, [95% CI = 1.26–1.82], OR per 1-SD-increment in diastolic blood pressure: 1.51 [95% CI = 1.27–1.78]). Similarly, genetic liability to type 2 diabetes (OR per 2-fold increase in prevalence: 1.10 [95% CI = 1.03–1.16] for TOAST-defined, and 1.10 [95% CI = 1.04–1.15] for CCS-defined) and genetically predicted waist-to-hip ratio (OR per 1-SD-increment: 1.42 [95% CI = 1.19–1.70] for TOAST-defined and 1.30 [95% CI = 1.06–1.60] for CCS-defined) were associated with stroke of undetermined source (see Fig 5). There was also suggestive evidence ($p < 0.05$) that genetic predisposition to higher LDL-cholesterol levels was associated with higher odds of stroke of undetermined source.

Interestingly, we found genetically upregulated IL-6 signaling, a pathway associated with LAAS and SVS and a potential therapeutic target for atherosclerosis,[38] to be associated with TOAST-defined stroke of undetermined source. Similarly, higher genetically predicted MCP-1/CCL2 levels, previously associated with other atherosclerotic phenotypes,[40] also showed suggestive effects on stroke of undetermined source defined according to either of the 2 classification systems (see Fig 5). Finally, genetic predisposition to higher factor XI levels, a member of the intrinsic coagulation pathway that is a target for next-generation anticoagulants,[47] was associated with a higher risk of CCS-defined stroke of undetermined source, similarly to its effects on CES (see Fig 5). There was only moderate evidence of heterogeneity across MR analyses, and for significant effects with evidence of heterogeneity the results remained consistent when using the median approach and MR Egger, thus suggesting only minor influence of directional pleiotropy on our results (Supplementary Table S8).

Discussion

Leveraging human genetic data, we found considerable genetic overlap between stroke of undetermined source and the most common ischemic stroke subtypes. We identified 19 shared genomic loci with LAAS, 2 with CES, and 5 with SVS, pointing to vascular risk factors, atherosclerosis- and thrombosis-related mechanisms, as well as altered transcription profiles in arterial tissues and blood. Using MR, we found a significant association between genetic predisposition to carotid atherosclerosis and stroke of undetermined source. MR analyses further provided evidence for causal links with traditional vascular risk factors, such as blood pressure, diabetes, and obesity, but also potentially druggable pro-inflammatory and coagulation mechanisms (IL-6 signaling, MCP-1, and factor XI).

The substantial genetic overlap between strokes of defined etiology and stroke of undetermined source likely indicates that a proportion of strokes of undetermined source is caused by undetected well-established causes of stroke. The fact that genetic determinants of the different stroke subtypes do not respect diagnostic boundaries highlights the limits of the existing classification systems and the most widely used diagnostic procedures to differentiate the most common stroke etiologies. Indeed, both TOAST and CCS require specific criteria for characterizing an ischemic stroke as LAAS, CES, or SVS that in the clinical setting might lead to several "grey zones," such as large artery stenoses below 50%, indirect signs, and high risk of atrial fibrillation in patients with sinus rhythm, or presumably lacunar strokes that are slightly larger than 15 mm on MRI.[48,49] The similarity in the genetic architecture of the diagnostic entity of stroke of undetermined source with the ones of LAAS, CES, and SVS implies that at least some of these strokes might be the result of such "grey zone" pathologies, rather than rarer causes of stroke missed during diagnostic workup. Given the importance of determining stroke etiology for focused secondary prevention and for selecting patient subgroups for clinical trials, our results highlight the need for more pragmatic clinical stroke classifications and also lend support to the heterogeneity of underlying mechanisms in patients with stroke of undetermined source.[4]

Exploring this genetic overlap at the locus-level, we found a higher number of shared loci with LAAS, as compared to CES and SVS. Many of these loci have been previously associated with coronary artery disease or traditional cardiometabolic risk factors, such as hypertension, diabetes, and obesity, thus further highlighting the overlap between stroke of undetermined source and traits traditionally associated with atherosclerosis. Some of these loci were further associated with altered transcription of genes (*PDE8B*, *ALDH2*, *NBEAL1*, and *CARF*) in arterial tissues and in blood. Although some of these genes have been previously involved in other forms of cardiovascular disease, their potential relevance specifically for stroke of undetermined source would need to be tested in a larger sample that would offer more power.

Our MR analyses further highlighted that higher genetically predicted cIMT associates with stroke of undetermined source, thus pointing to a role of carotid atherosclerosis beyond that which is needed to define a stroke as of large artery origin. Observational studies generally agree with this finding: complex non-stenosing plaques with specific morphological features, such as

intraplaque hemorrhages or a thin fibrous cap, are more common in patients with stroke of undetermined source.[5,6] Our MR results thus triangulate these findings supporting a causal role of carotid atherosclerosis even in strokes that are currently characterized as of undetermined source.

An important role of atherosclerosis in the development of stroke of undetermined source is further highlighted by the risk factor profile of this subtype in our MR analyses. Beyond the effects of blood pressure, which are significant for all stroke subtypes, we found evidence supporting associations with genetic liability to type 2 diabetes, as well as higher genetically predicted waist-to-hip ratio. Furthermore, there were suggestive effects of genetically predicted lipid (LDL and HDL cholesterol) and inflammatory traits (IL-6 signaling and MCP-1 levels) that share links with atherosclerotic disease.[38–40,50] These results have potentially important implications. First, they highlight the need of managing traditional vascular risk factors in patients with strokes of undetermined source. Second, they provide support that patients with stroke of undetermined source could be candidates for trials testing specific under development secondary preventive strategies, such as anti-inflammatory or lipid-lowering treatments.

An important finding of our study is the potentially causal effect of factor XI levels on risk of stroke of undetermined source, similar to CES. Whereas 2 randomized trials found novel oral anticoagulants (dabigatran and rivaroxaban, respectively) not superior to antiplatelets for reducing recurrent stroke risk in patients with stroke of undetermined source,[10,11] next-generation anticoagulants targeting factor XI appear to have a similar efficacy for preventing venous thromboembolism and a more favorable bleeding risk profile in phase II trials.[47,51] Our MR results cannot answer whether factor XI inhibitors would be superior to antiplatelets for prevention of stroke of undetermined source, but provide genetic support for their potential efficacy. Superiority to antiplatelets in prevention of ischemic endpoints must still be tested in a clinical trial directly comparing these agents.

A key component of the diagnostic workup for patients with stroke of undetermined source is focused on detection of atrial fibrillation.[8] Although our MR results suggest a potentially causal role of atrial fibrillation in stroke of undetermined source, the association was significantly attenuated after excluding cases of undetermined etiology due to presence of multiple competing etiologies or due to incomplete diagnostic workup. However, extant GWASs for atrial fibrillation include only cases with detected arrhythmia detection was possible, and might therefore under-represent cases with undetected paroxysmal atrial fibrillation, such as in stroke of undetermined source, thus underestimating their genetic effect. We should note although our results agree with a recent trial that found no benefit for stroke prevention among high-risk patients who received an implantable loop recorder, despite a 3-fold increase in paroxysmal atrial fibrillation detection.[52]

Our study has limitations. First, there could be substantial misclassification between stroke subtypes in the included causes, particularly the inclusion of stroke cases without adequate diagnostic workup, to the category of stroke of undetermined source. Similarly, cases with competing etiologies are traditionally assigned to this category using TOAST criteria, although CCS partially controls for this issue. To overcome this limitation, we performed analyses for both TOAST- and CCS-defined stroke subtypes, as well as for CCS-defined subcategories that separate cryptogenic or minor cardioembolic strokes from those that did not receive complete diagnostic workup or those that remained unclassified due to presence of competing etiologies. Still, the inclusion of minor cardioembolic cases in this category could bias the results. Moreover, we were unable to perform a sensitivity analysis on the subgroup of patients with embolic stroke of undetermined source that represent a potentially more therapeutically relevant entity.[4] Our findings supporting the overlap between stroke cases defined as undetermined and defined stroke subtypes could be explained by classification error, misclassification due to incomplete diagnostic workup, or shared etiology. We believe that the classification error is minimal, because stroke classification for the purposes of the current study was centrally performed for all patients by trained investigators based on standard criteria. This is supported by the robust genetic associations of the defined stroke subtypes with their established etiologies (eg, the very strong effect of genetic liability to atrial fibrillation only on cardioembolic stroke). As a second explanation, mixing patients with stroke classified as undetermined following complete diagnostic workup with patients who did not have a complete diagnostic workup could partly explain the findings. However, our sensitivity analyses restricted to cases defined as undetermined after complete diagnostic workup according to CCS still support a strong genetic overlap with defined stroke subtypes. Therefore, we believe that the main genetic findings are not driven by classification errors, but rather by shared etiology between stroke classified as of undetermined source and the defined stroke subtypes. This is, we assume, the result of inherent limitations of the established classification systems leading to the diagnostic separation of biologically common entities. Future studies with better phenotyping could explore in larger datasets the genetics of smaller groups of strokes of undetermined source defined according to stricter phenotypic criteria. Alternatively, larger studies could apply principal component analysis in the genetic architecture of stroke of undetermined source to explore whether these patients cluster in different groups representing different biological entities.

Second, our results were based on European populations and may thus not be generalizable to other ancestries. Third, the GWAS-PW and the polygenic risk score analyses could be biased due to an overlap in the control samples between GWASs for the strokes of determined etiology from MEGASTROKE and the GWASs for strokes of undetermined source in SiGN. However, in our GWAS-PW analyses, the underlying correlation between samples has been taken into account. Fourth, it was not possible to explore genetic correlations between defined stroke subtypes and stroke of undetermined source with more dedicated tools, such as LD score regression, due to volatility in heritability estimates in the stroke subtypes GWASs caused by low statistical power.

In conclusion, we found that the stroke of undetermined source shares genetic and modifiable risk factors with the most common stroke subtypes, especially atherosclerotic stroke, thus highlighting specific limitations of current subtyping approaches. Our analyses further support potentially causal links with conventional vascular risk factors, as well as inflammatory and thrombotic mechanisms, which could inform the design of future secondary prevention trials in this large proportion of patients who have a stroke at an ongoing risk for secondary ischemic events.

Acknowledgments

M.G. is supported by a Walter-Benjamin fellowship from the German Research Foundation (Deutsche Forschungsgemeinschaft [DFG], GZ: GE 3461/1-1) and the FöFoLe program of LMU Munich (FöFoLe-Forschungsprojekt Reg. Nr. 1120). M.D. is also supported by DFG under Germany's Excellence Strategy within the framework of the Munich Cluster for Systems Neurology (EXC 2145 SyNergy – ID 390857198) and as part of CRC 1123 (B3) and DI 722/16-1. J.R. is supported by the NIH (R01NS036695, R01NS093870, R01NS100178, R01NS103924, R01AG067019, U01NS102289, and UO1NS069763) and the American Heart Association (812095). C.D.A. is supported by the NIH (R01NS103924 and U01NS069763), the American Heart Association (812095 and 18SFRN34150007), and Massachusetts General Hospital. Open Access funding enabled and organized by Projekt DEAL.

Author Contributions

M.G. and C.D.A. contributed to the conception and design of the study. All authors contributed to the acquisition and analysis of data. All authors contributed to drafting the text or preparing the figures.

Potential Conflicts of Interest

J.R. serves as a consultant to Takeda and Boehringer Ingelheim outside the current work. C.D.A. receives sponsored research support from Bayer AG, and has served as a consultant for ApoPharma outside the current work. The other authors have nothing to disclose.

References

1. GBD 2016 Stroke Collaborators. 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 (10):795–820.
2. Li L, Yiin GS, Geraghty OC, et al. Incidence, outcome, risk factors, and long-term prognosis of cryptogenic transient ischaemic attack and ischaemic stroke: a population-based study. Lancet Neurol 2015;14:903–913.
3. Bang OY, Lee PH, Joo SY, et al. Frequency and mechanisms of stroke recurrence after cryptogenic stroke. Ann Neurol 2003;54: 227–234.
4. Hart RG, Diener HC, Coutts SB, et al. Embolic strokes of undetermined source: the case for a new clinical construct. Lancet Neurol 2014;13:429–438.
5. Kopczak A, Schindler A, Bayer-Karpinska A, et al. Complicated carotid artery plaques as a cause of cryptogenic stroke. J Am Coll Cardiol 2020;76:2212–2222.
6. Brunner G, Virani SS, Sun W, et al. Associations between carotid artery plaque burden, plaque characteristics, and cardiovascular events: the ARIC carotid magnetic resonance imaging study. JAMA Cardiol 2021;6(1):79–86. doi: 10.1001/jamacardio.2020.5573
7. Saver JL. CLINICAL PRACTICE. Cryptogenic Stroke. N Engl J Med 2016;374:2065–2074.
8. Gladstone DJ, Spring M, Dorian P, et al. Atrial fibrillation in patients with cryptogenic stroke. N Engl J Med 2014;370:2467–2477.
9. Tsivgoulis G, Katsanos AH, Grory BM, et al. Prolonged cardiac rhythm monitoring and secondary stroke prevention in patients with cryptogenic cerebral ischemia. Stroke 2019;50:2175–2180.
10. Diener HC, Sacco RL, Easton JD, et al. Dabigatran for prevention of stroke after embolic stroke of undetermined source. N Engl J Med 2019;380:1906–1917.
11. Hart RG, Sharma M, Mundl H, et al. Rivaroxaban for stroke prevention after embolic stroke of undetermined source. N Engl J Med 2018;378:2191–2201.
12. Holmes MV, Ala-Korpela M, Smith GD. Mendelian randomization in cardiometabolic disease: challenges in evaluating causality. Nat Rev Cardiol 2017;14:577–590.
13. Georgakis MK, Gill D. Mendelian randomization studies in stroke: exploration of risk factors and drug targets with human genetic data. Stroke 2021;52:2992–3003.
14. Network NSG. International stroke genetics C. loci associated with ischaemic stroke and its subtypes (SiGN): a genome-wide association study. Lancet Neurol 2016;15:174–184.
15. Adams HP Jr, Bendixen BH, Kappelle LJ, et al. 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.
16. Ay H, Benner T, Arsava EM, et al. A computerized algorithm for etiologic classification of ischemic stroke: the causative classification of stroke system. Stroke 2007;38:2979–2984.

17. Malik R, Chauhan G, Traylor M, 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.

18. Pulit SL, Weng LC, McArdle PF, et al. Atrial fibrillation genetic risk differentiates cardioembolic stroke from other stroke subtypes. Neurol Genet 2018;4:e293.

19. International Consortium for Blood Pressure Genome-Wide Association S, Ehret GB, Munroe PB, et al. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. Nature 2011;478:103–109.

20. Pickrell JK, Berisa T, Liu JZ, et al. Detection and interpretation of shared genetic influences on 42 human traits. Nat Genet 2016;48: 709–717.

21. Turley P, Walters RK, Maghzian O, et al. Multi-trait analysis of genome-wide association summary statistics using MTAG. Nat Genet 2018;50:229–237.

22. Grotzinger AD, Rhemtulla M, de Vlaming R, et al. Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits. Nat Hum Behav 2019;3:513–525.

23. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. PLoS Comput Biol 2015;11: e1004219.

24. Barbeira AN, Dickinson SP, Bonazzola R, et al. Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. Nat Commun 2018;9:1825.

25. Gerring ZF, Mina-Vargas A, Gamazon ER, Derks EM. E-MAGMA: an eQTL-informed method to identify risk genes using genome-wide association study summary statistics. Bioinformatics 2021;37:2245–2249.

26. Giambartolomei C, Vukcevic D, Schadt EE, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet 2014;10:e1004383.

27. Franceschini N, Giambartolomei C, de Vries PS, et al. GWAS and colocalization analyses implicate carotid intima-media thickness and carotid plaque loci in cardiovascular outcomes. Nat Commun 2018; 9:5141.

28. Roselli C, Chaffin MD, Weng LC, et al. Multi-ethnic genome-wide association study for atrial fibrillation. Nat Genet 2018;50: 1225–1233.

29. Persyn E, Hanscombe KB, Howson JMM, et al. Genome-wide association study of MRI markers of cerebral small vessel disease in 42,310 participants. Nat Commun 2020;11:2175.

30. Yaghi S, Kamel H, Elkind MSV. Atrial cardiopathy: a mechanism of cryptogenic stroke. Expert Rev Cardiovasc Ther 2017;15:591–599.

31. Pirruccello JP, Di Achille P, Choi SH, et al. Deep learning of left atrial structure and function provides link to atrial fibrillation risk. MedRxiv 2021. doi: 10.1101/2021.08.02.21261481.

32. Evangelou E, Warren HR, Mosen-Ansorena D, et al. Publisher correction: genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. Nat Genet 2018;50:1755.

33. Klarin D, Damrauer SM, Cho K, et al. Genetics of blood lipids among ~300,000 multi-ethnic participants of the million veteran program. Nat Genet 2018;50:1514–1523.

34. Mahajan A, Taliun D, Thurner M, et al. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. Nat Genet 2018;50:1505–1513.

35. Georgakis MK, Harshfield EL, Malik R, et al. Diabetes mellitus, glycemic traits, and cerebrovascular disease: a Mendelian randomization study. Neurology 2021;96:e1732–e1742.

36. Wootton RE, Richmond RC, Stuijfzand BG, et al. Evidence for causal effects of lifetime smoking on risk for depression and schizophrenia: a Mendelian randomisation study. Psychol Med 2020;50:2435–2443.

37. Yengo L, Sidorenko J, Kemper KE, et al. Meta-analysis of genome-wide association studies for height and body mass index in approximately 700000 individuals of European ancestry. Hum Mol Genet 2018;27:3641–3649.

38. Georgakis MK, Malik R, Gill D, et al. Interleukin-6 signaling effects on ischemic stroke and other cardiovascular outcomes: a Mendelian randomization study. Circ Genom Precis Med 2020;13:e002872.

39. Georgakis MK, Malik R, Li X, et al. Genetically downregulated Interleukin-6 signaling is associated with a favorable cardiometabolic profile: a phenome-wide association study. Circulation 2021;143: 1177–1180.

40. Georgakis MK, Gill D, Rannikmae K, et al. Genetically determined levels of circulating cytokines and risk of stroke. Circulation 2019; 139:256–268.

41. Gill D, Georgakis MK, Laffan M, et al. Genetically determined FXI (factor XI) levels and risk of stroke. Stroke 2018;49:2761–2763.

42. Burgess S, Ference BA, Staley JR, et al. Association of LPA variants with risk of coronary disease and the implications for lipoprotein(a)-lowering therapies: a Mendelian randomization analysis. JAMA Cardiol 2018;3:619–627.

43. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. Genet Epidemiol 2013;37:658–665.

44. Bowden J, Hemani G, Davey SG. Invited commentary: detecting individual and global horizontal pleiotropy in Mendelian randomization-a job for the humble heterogeneity statistic? Am J Epidemiol 2018;187:2681–2685.

45. Hartwig FP, Davey Smith G, Bowden J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. Int J Epidemiol 2017;46:1985–1998.

46. 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.

47. Weitz JI, Bauersachs R, Becker B, et al. Effect of Osocimab in preventing venous thromboembolism among patients undergoing knee arthroplasty: the FOXTROT randomized clinical trial. JAMA 2020;323:130–139.

48. Radu RA, Terecoasa EO, Bajenaru OA, Tiu C. Etiologic classification of ischemic stroke: where do we stand? Clin Neurol Neurosurg 2017; 159:93–106.

49. Meschia JF, Barrett KM, Chukwudelunzu F, et al. Interobserver agreement in the trial of org 10172 in acute stroke treatment classification of stroke based on retrospective medical record review. J Stroke Cerebrovasc Dis 2006;15:266–272.

50. Georgakis MK, van der Laan SW, Asare Y, et al. Monocyte-chemoattractant Protein-1 levels in human atherosclerotic lesions associate with plaque vulnerability. Arterioscler Thromb Vasc Biol 2021;41:2038–2048.

51. Verhamme P, Yi BA, Segers A, et al. Abelacimab for prevention of venous thromboembolism. N Engl J Med 2021;385:609–617.

52. Svendsen JH, Diederichsen SZ, Hojberg S, et al. Implantable loop recorder detection of atrial fibrillation to prevent stroke (the LOOP study): a randomised controlled trial. Lancet 2021;398:1507–1516.