Multi-ancestry meta-analysis identifies 5 novel loci for ischemic stroke and reveals heterogeneity of effects between sexes and ancestries

Ida Surakka, University of Michigan Medical School
Kuan-Han Wu, University of Michigan Medical School
Whitney Hornsby, University of Michigan Medical School
Brooke N. Wolford, University of Michigan Medical School
Fred Shen, University of Michigan Medical School
Wei Zhou, Emory University
Jennifer E. Huffman, VA Boston Healthcare System
Anita Pandit, University of Michigan School of Public Health
Yao Hu, Fred Hutchinson Cancer Research Center
Ben Brumpton, Norges Teknisk-Naturvitenskapelige Universitet

Only first 10 authors above; see publication for full author list.

Journal Title: Cell Genomics
Volume: Volume 3, Number 8
Publisher: Elsevier | 2023-08-09, Pages 100345-100345
Type of Work: Article | Final Publisher PDF
Publisher DOI: 10.1016/j.xgen.2023.100345
Permanent URL: https://pid.emory.edu/ark:/25593/w8671

Final published version: http://dx.doi.org/10.1016/j.xgen.2023.100345

Copyright information:
© 2023 The Authors

This is an Open Access work distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (https://creativecommons.org/licenses/by-nc-nd/4.0/).

Accessed July 30, 2024 7:02 PM EDT
Multi-ancestry meta-analysis identifies 5 novel loci for ischemic stroke and reveals heterogeneity of effects between sexes and ancestries

**Graphical abstract**

- GBMI + MegaStroke → 14 candidates → Replication → Stratify by sex and ancestry
- ALDH2 male specific ABO ancestry heterogeneous
- 5 novel loci: LAMC1, ZSWIM2/CALCRL, PLSCR1/5, CDKN1A, SWAP70

**Highlights**
- Five novel loci associated with ischemic stroke were discovered
- Previously reported ABO locus shows significant ancestry heterogeneity
- Known association in ALDH2 locus is sex and ancestry specific

**Authors**
Ida Surakka, Kuan-Han Wu, Whitney Hornsby, ..., Benjamin M. Neale, Cristen J. Willer, on behalf of the Global Biobank Meta-analysis Initiative (GBMI)

**Correspondence**
isurakka@umich.edu

**In brief**
By combining data from diverse Global Biobank Meta-analysis Initiative (GBMI) and the MegaStroke consortium, Surakka et al. identified five novel loci associated with ischemic stroke and showed that some genetic associations differ between sexes and ancestries. These results highlight the importance of sex- and ancestry-informed investigations in complex disease genetics.

Surakka et al., 2023, Cell Genomics 3, 100345
August 9, 2023 © 2023 The Authors.
https://doi.org/10.1016/j.xgen.2023.100345
Multi-ancestry meta-analysis identifies 5 novel loci for ischemic stroke and reveals heterogeneity of effects between sexes and ancestries

Ida Surakka,1,18,* Kuan-Han Wu,2 Whitney Hornsby,1 Brooke N. Wolford,2,3 Fred Shen,1 Wei Zhou,4,5 Jennifer E. Huffman,6 Anita Pandit, 3 Yao Hu, 7 Ben Brumpton, 8,9,10 Anne Heidi Skogholt, 8 Maiken E. Gabrielsen, 8 Robin G. Walters, 11,12 The TOPMed Stroke Working Group, Million Veteran Program (MVP), Kristian Hveem,8,10 Charles Kooperberg,7 Sebastian Zöllner,3 Peter W.F. Wilson,13,14 Nadia R. Sutton,1 Mark J. Daly,4,5,16 Benjamin M. Neale,4,5,16 Cristen J. Willer,1,2,9,17 and on behalf of the Global Biobank Meta-analysis Initiative (GBMI)

1Department of Internal Medicine, University of Michigan, Ann Arbor, MI, USA
2Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, USA
3Center for Statistical Genetics, University of Michigan School of Public Health, Ann Arbor, MI, USA
4Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital, Boston, MA, USA
5Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA
6Massachusetts Physicians Epidemiology Research and Information Center (MAVERIC), VA Boston Healthcare System, Boston, MA, USA
7Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, USA
8K.G. Jebsen Center for Genetic Epidemiology, Department of Public Health and Nursing, NTNU, Norwegian University of Science and Technology, Trondheim, Norway
9Clinic of Medicine, St. Olav’s Hospital, Trondheim University Hospital, Trondheim, Norway
10HUNT Research Centre, Department of Public Health and Nursing, Norwegian University of Science and Technology, Levanger, Norway
11Medical Research Council Population Health Research Unit, Nuffield Department of Population Health, University of Oxford, Oxford, UK
12Clinical Trial Service Unit and Epidemiological Studies Unit, Nuffield Department of Population Health, University of Oxford, Oxford, UK
13Atlanta VA Health Care System, Decatur, GA, USA
14Department of Epidemiology, Emory University Rollins School of Public Health, Atlanta, GA, USA
15Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland
16Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA
17Department of Human Genetics, University of Michigan, Ann Arbor, MI, USA
18Lead contact
*Correspondence: isurakka@umich.edu

https://doi.org/10.1016/j.xgen.2023.100345

SUMMARY

Stroke is the second leading cause of death and disability worldwide. Stroke prevalence varies by sex and ancestry, possibly due to genetic heterogeneity between subgroups. We performed a genome-wide meta-analysis of 16 biobanks across multiple ancestries to study the genetics of ischemic stroke (60,176 cases, 1,310,725 controls) as part of the Global Biobank Meta-analysis Initiative (GBMI) and further combined the results with previously published MegaStroke. Five novel loci for ischemic stroke (LAMC1, CALCRL, PLSCR1, CDKN1A, and SWAP70) were identified after replication in four additional datasets. One previously reported locus showed significant ancestry heterogeneity (ABO), and one showed significant sex heterogeneity (ALDH2). The ALDH2 association was male specific (males p = 1.67e−24, females p = 0.126) and was additionally observed only in the East Asian ancestry (male) samples. These findings emphasize the need for more diverse datasets with large sample sizes to further understand the genetic predisposition of stroke in different ancestry and sex groups.

INTRODUCTION

Cerebrovascular accidents (stroke) are the second leading cause of death and disability worldwide due to brain infarction (ischemic stroke) or intracerebral hemorrhage.1 The former can be further divided into different subgroups, including cardioembolic, large vessel, and small vessel stroke, and the latter into lobar and non-lobar hemorrhagic stroke. Mapping genetic variants associated with stroke has been more challenging than for other homogeneous complex diseases, such as coronary artery disease2 or type 2 diabetes,3 given that stroke subgroups have different etiologies4 and heritability.5 Seventy-one loci have been identified using genome-wide association study (GWAS) methods6–28 despite the complex phenotypic heterogeneity of stroke. These studies consist of sample sizes up to 900,000 (72,000 cases with all-cause stroke) and show that genetic predisposition varies between subgroups. Most known loci are associated with ischemic stroke, likely due
to higher prevalence of that subtype (~80% of the cases) and thus more power to detect an association. Additionally, stroke prevalence has been shown to differ between populations of different ancestry and sexes, suggesting possible heterogeneity of environmental and/or genetic factors contributing to the risk for stroke. Here, we perform a new GWAS as part of the Global Biobank Meta-analysis Initiative (GBMI) to evaluate the potential role that biobanks could play in genetic discovery of complex phenotypes such as stroke. We further combine the summary statistics with the previously published MegaStroke consortium to further examine the genetic variants and to test whether the observed associations for ischemic stroke show either ancestry- or sex-specific effects.

RESULTS

Ischemic stroke locus discovery

We initially assessed association summary statistics from 16 biobanks with participants from various ancestries (Figure 1) and identified 2 potentially novel and 10 previously published loci (Table S1) with genome-wide significant association (p < 5e-8). To further increase our discovery sample size, we combined the new GBMI summary statistics with the previously published summary statistics from the MegaStroke consortium (all ischemic stroke summary statistics). As the MegaStroke consortium included the stroke cases from the BioBank Japan (BBJ), which also contributed to the GBMI meta-analysis, for the joint meta-analysis, we used GBMI summary statistics that excluded BBJ. By combining all available overlapping variants in these two datasets, we identified a total of 37 genome-wide significant loci (Table S1), of which 12 had not been previously published with genome-wide significance for any stroke subtype.

To further validate the 14 potentially novel ischemic stroke-associated loci emerging from GBMI (2 loci) and the joint meta-analysis with MegaStroke (12 loci), we performed replication in four additional biobanks (STAR Methods; Table S2). We were able to replicate 5 loci with a p < 3.6e-3 (Bonferroni correction for 14 tests; Table 1), with 3 additional loci showing nominal significance (Table S2). Given that the 5 replicated variants were either intergenic or intronic, we looked for the possible evidence of variant-gene expression association for the variants in the GTEx database, followed by a colocalization analysis of the tissue-specific gene expression association and the stroke meta-analysis association for each of the loci (Table S3). Findings show PLSCR5 and SWAP70 as likely candidate genes by having posterior probability for a shared underlying causal variant >80% with the ischemic stroke association results. Additionally, we searched for any previous publications linked to the variants and the region around them in a GWAS catalog, together with summary statistics for some of the well-known heart disease risk-related biomarkers and related traits (Table S4), to better understand the mechanisms underlying the novel ischemic stroke association. The results suggest that the first novel locus, LAMC1, might increase the stroke risk through an effect on lipid levels. This effect does not show in the association of the variant with type 2 diabetes or coronary artery disease. Furthermore, we observe significant associations (p < 0.01, Bonferroni for 5 tests) between the CDKN1A lead variant and triglyceride levels, systolic blood pressure, pulse pressure, and type 2 diabetes and between the SWAP70 lead variant and high-density lipoprotein cholesterol, total cholesterol, all three blood pressure traits, and coronary artery disease.

In 6 of the previously known loci, the GBMI trans-ancestry meta-analysis is pointing toward a different lead variant compared with the joint meta-analysis with MegaStroke (Table S1), partly due to the denser imputation reference panels used by the GBMI. For two of these loci, CENPQ and ALDH2, the new GBMI lead variant falls into the protein-coding region of the gene. When looking at the GBMI association results across all the previously published MegaStroke loci (Table S5A; Figure S1), we can see that the association results from the GBMI show deflated effect sizes compared with the MegaStroke transethnic AIS (all ischemic stroke) results (correlation of the effect sizes R² = 0.62). This suggests that the population-based sampling and the ICD-based phenotyping in the GBMI, in contrast to the majority case-control sampling and clinically defined phenotype in MegaStroke, reduces the observed
Table 1. The 5 newly identified and replicated loci associated with ischemic stroke

<table>
<thead>
<tr>
<th>Locus</th>
<th>rsID</th>
<th>Chromosome:</th>
<th>EA/NEA</th>
<th>EAF</th>
<th>Discovery OR (95% CI)</th>
<th>Discovery p</th>
<th>Replication OR (95% CI)</th>
<th>Replication p</th>
</tr>
</thead>
<tbody>
<tr>
<td>LAMC1*</td>
<td>rs4652772</td>
<td>1:183,054,561</td>
<td>C/G</td>
<td>0.560</td>
<td>1.03 (1.01; 1.05)</td>
<td>1.76e-8</td>
<td>1.03 (1.01; 1.05)</td>
<td>3.77e-4</td>
</tr>
<tr>
<td>ZSWIM2/ CALCRL</td>
<td>rs72896798</td>
<td>2:187,282,743</td>
<td>A/T</td>
<td>0.102</td>
<td>1.05 (1.03; 1.07)</td>
<td>3.22e-8</td>
<td>1.04 (1.02; 1.06)</td>
<td>8.49e-4</td>
</tr>
<tr>
<td>PLSCR1/ PLSCR5</td>
<td>rs6762490</td>
<td>3:146,658,664</td>
<td>A/T</td>
<td>0.363</td>
<td>1.04 (1.02; 1.05)</td>
<td>1.48e-8</td>
<td>1.03 (1.02; 1.05)</td>
<td>1.51e-4</td>
</tr>
<tr>
<td>CDKN1A</td>
<td>rs762624</td>
<td>6:36,677,811</td>
<td>A/C</td>
<td>0.663</td>
<td>1.04 (1.03; 1.06)</td>
<td>3.65e-10</td>
<td>1.03 (1.02; 1.05)</td>
<td>1.2e-4</td>
</tr>
<tr>
<td>SWAP70*</td>
<td>rs301556</td>
<td>11:9,735,177</td>
<td>T/C</td>
<td>0.448</td>
<td>0.96 (0.95; 0.97)</td>
<td>2.45e-10</td>
<td>0.97 (0.97; 0.99)</td>
<td>5.3e-4</td>
</tr>
</tbody>
</table>

The discovery results are presented from the joint meta-analysis of GBMI and MegaStroke summary statistics. EA, effect allele; NEA, non-effect allele; OR, odds ratio; CI, confidence interval.

*Also genome-wide significant in the recently published GIGAStroke meta-analysis.36

Effect sizes notably (mean deflation = 52%). If we look at the GBMI results across all previously published stroke loci (Table S5B), we observe only one of the variants with a mean allelic frequency (MAF) <0.05 with a significant p value in the GBMI lookup (p < 4.8e-4, Bonferroni for 104 tests). The original association was observed in diabetic African American samples for ischemic stroke with an odds ratio (OR) of 3.51 (GBMI OR = 1.17). Another trend included a higher change of replicated association in the GBMI for variants reported between 2018 and 2021 (33 out of 79) compared with older studies (7/25), suggesting that the increased sample size in the original study increases the probability of replicating the association (lower type I error) (Table S5B). If we dissect the data using the reported associated trait of all IS associations (includes large and small vessel stroke, cardioembolic stroke, and ischemic stroke), GBMI shows significant association for 29 out of 74 variants. There was no distinct trend for a specific subtype of ischemic stroke (small vessel stroke 2/7, large artery stroke 2/12, and cardioembolic stroke 2/12) to be more or less likely to be replicated in GBMI. However, for the subtype-specific loci without significant association to any other stroke type, for example cardioembolic stroke loci with no previously reported association to more general ischemic stroke, there was no significant association in GBMI in any of the 6 loci. The same trend is observed for the 7 large and 4 small vessel stroke-specific loci. These results suggest that only loci that additionally show association to ischemic stroke in the previous studies are replicable in the GBMI meta-analysis.

Ancestry and sex heterogeneity of the ischemic stroke-associated loci

To further study the possible ancestry and sex differences in the ischemic stroke-associated loci, we used the GBMI summary statistics, as MegaStroke results were not available for the subgroups. First, the ancestry-specific results of the discovered ischemic stroke associations, both previously known (26 loci) and the replicated novel loci (5 loci), were examined. For loci where the lead variants were different between the two discovery approaches, we used the GBMI lead variant for these analyses. We observed one locus showing significant ancestry heterogeneity (p < 1.6e-3, Bonferroni correction for 31 tests; Table S6). The lead variant, rs635634, is close to the ABO gene and had consistent effects in the European (non-Finnish European [NFE]) and South Asian (SAS) ancestry populations, about a two times larger effect in the African (AFR) ancestry population, and an opposite effect in the admixed American (AMR) individuals. In addition to ABO, we observed nominally significant ancestry heterogeneity for three loci: LOC100505841, CDKN2B-AS1, and COL4A1.

Next, we looked for evidence of sex-specific heterogeneity in the associated loci. One of the 31 associated loci showed significant sex heterogeneity (ALDH2, p < 1.6e-3, Bonferroni correction for 31 tests; Table S6). The ALDH2 locus has been associated with stroke in previous GWASs, and the sex-heterogenic effects on ischemic stroke have been indirectly implied. In GBMI, the stroke association was stronger in males than for females (effect sizes for males –0.160 and for females –0.031) and was significant only in males (males p = 1.67e-24 and females p = 0.126). Additionally, the association was East Asian (EAS) specific, reflecting the absence or very low frequency of this variant in other ancestries in the GBMI meta-analysis. Furthermore, the allele frequency of the lead variant, rs671, was 25.5% for the EAS population and ≤0.05% for other ancestries in the publicly available gnomAD database. In addition to the one significant sex-heterogenic locus, we observed two lead variants that showed nominally significant sex heterogeneity (HDAC9 and COL4A1).

DISCUSSION

By performing a genome-wide association analysis of ischemic stroke within GBMI followed by a joint analysis combining GBMI and previously published MegaStroke ischemic stroke summary statistics, we identified 5 novel genetic loci associated with ischemic stroke, confirmed through replication in four independent datasets. Moreover, we observed significant ancestry heterogeneity in the ABO locus and significant sex heterogeneity in the ALDH2 locus.

Many of the novel associations fall into/near candidate genes with a plausible biological link with stroke risk. LAMC1 gene expression has been shown to be inversely linked with the HMGBl expression, which promotes atherosclerosis and has been implicated in multiple cardiovascular diseases. CALCR is known to regulate blood pressure, vascular biology, and inflammation. PLSCR1, the most studied of the of the scramblase family members, is known to redistribute phospholipids across the plasma membrane in response to an increase of intracellular calcium and has additionally been associated with...
blood coagulation.\textsuperscript{42} CDKN1A knockout mice have been shown to develop age-dependent cardiac hypertrophy.\textsuperscript{43} Three of the putative novel loci, PRDM16 (PR/SET domain 16), GRK5 (G protein-coupled receptor kinase 5), and PLEKH1, showed only nominal association in the replication. Of these, PRDM16 and GRK5 have been previously linked to cardiac hypertrophy.\textsuperscript{44,45} Functional follow-up studies are needed to examine the possible connection between these candidate genes and the observed stroke association.

We observed two protein-altering lead variants that resided in CENPQ and ALDH2 genes. CENPQ has been previously reported to be associated with blood homocysteine levels, a known risk factor for stroke.\textsuperscript{46} CENPQ has also been suggested to regulate MUT expression, one of the driver genes in the causal relationship between blood homocysteine levels and small vessel stroke.\textsuperscript{47} The ALDH2 exonic lead variant, rs671, is a well-known polymorphism linked to alcohol consumption and hypertension in EAS individuals.\textsuperscript{48} This variant also showed significant sex heterogeneity where the effect of this variant is observed more strongly in men, likely due to cultural and societal differences in patterns of alcohol consumption between sexes. We observed an association with ischemic stroke for this variant only in males and, more specifically, in those with EAS ancestry.

Both exonic variants were lead variants only in the GBMI meta-analysis, highlighting the importance of the dense imputation panels allowing for better coverage of the locus in the association testing. One additional locus that was not significantly associated with ischemic stroke in the joint meta-analysis was observed in the GBMI meta-analysis near the FGF5 gene. In addition to the previous association with ischemic stroke,\textsuperscript{13} this locus is associated with blood pressure traits in Chinese individuals with a higher body mass index,\textsuperscript{49} and the expression of the FGF5 gene has been shown to differ in patients with hypertension.\textsuperscript{49} Additionally, recent studies have shown that the fibroblast growth factors (FGFs) could potentially be used to treat stroke in animal models.\textsuperscript{50}

Limitations of the study
The high representation of EAS individuals was a notable strength of this investigation, likely leading to the discovery of EAS ancestry-driven sex-specific effects in the previously published ALDH2 stroke locus. Our study has several limitations due to the large overall size of the meta-analysis. First, the ischemic stroke phenotype was defined in the GBMI meta-analyses using different imputation panels, the total number of variants in the joint meta-analysis was restricted by the less dense dataset, allowing for a future study combining the datasets with state-of-the-art imputation reference panels.

OUTLOOK/CONCLUSION
We present 5 novel loci associated with ischemic stroke and show that some stroke loci have sex- and/or ancestry-specific patterns. Utilization of the EHR-based biobanks allowed investigation of large sample sizes with higher diversity. However, this comes with the cost of reduced effect sizes, likely due to imprecision of the phenotypes and more heterogeneity of case definitions. These findings emphasize the need for more diverse datasets, especially with clinically defined phenotypes and large case representation, with larger sample sizes to further understand the genetic predisposition of stroke in different ancestry and sex groups. Further effort invested in artificial intelligence (AI) or clinically informed phenotyping using full EHR information may further improve power of genetic discovery from EHR-based datasets, especially in complement to traditional case-ascertainment discovery to ensure that estimated effect sizes are appropriate.

STAR METHODS
Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **RESOURCE AVAILABILITY**
  - Lead contact
  - Materials availability
  - Data and code availability
- **METHOD DETAILS**
  - GBMI multi-ancestry meta-analysis
  - Joint meta-analysis of GBMI and MegaStroke
  - Replication cohorts
  - eQTL look-ups and colocalization analysis
  - GWAs catalog data
- **SUPPLEMENTAL INFORMATION**
  Supplemental information can be found online at https://doi.org/10.1016/j.xgen.2023.100345.

CONSORTIA
The members of the Global Biobank Meta-analysis Initiative are Wei Zhou, Masahiro Kanai, Kuan-Han H. Wu, Humaira Rasheed, Kristin Tsuo, Jibril B. Hirbo, Ying Wang, Arjun Bhattacharya, Huling Zhao, Shinichi Namba, Ida Surakka, Brooke N. Wolford, Valeria Lo Faro, Esteban A. Lopera-Mayá, Kristi Läli, Marie-Julie Favé, Juulia J. Partanen, Sinéad B. Chapman, Juha Karjalainen, Mitja Kurki, Mutaamba Maasha, Ben M. Brumpton, Sameer Chavan, Tzu-
ACKNOWLEDGMENTS

We thank Dr. Robin Walters for the critical review of our manuscript. We would like to express our gratitude to all contributors to GBMI and the biobank participants who provided their data for biomedical research. Particularly, we are grateful to GBMI study cohorts, MGI and BioMe, who assisted with replication and interpretation of our findings. The authors acknowledge the participants, recruitment teams, and project managers of the GBMI for providing data aggregation, management, and distribution services in support of the research reported in this publication. We acknowledge BBJ (Yukiniro Okada, Koichi Matsua, and Masahiro Kanai); BioMe (Ruth Loos, Judy Cho, Eimear Kenny, Michael Preuss, and Simon Lee); BioVU (Nancy Cox and Jibril Hirbo); Canadian Partnership for Tomorrow (Philip Awadalla and Marie-Julie Fave); China Kadoorie (Robin Walters, Kuang Lin, and Iona Millwood); Colorado Center for Personalized Medicine (Kathleen Barnes, Michelle Daya, and Chris Gignou); deCODE Genetics (Kari Stefansson and Unnur þorsteinsdóttir); East London A. van Heel Health (David A. van Heel, Sarah Finer, and Richard Trembath); Estonian Biobank (Andres Metspalu, Reelik Mägi, Tõnu Esko, and Halli Huang); UCLA Precision Health Biobank (Ruth Johnson, Yi Ding, Alex Chiu, Bogdan Pasaniuc, and Daniel Geschwind); and UK Biobank (Konrad Karczewski and Alicia Martin). The Trandelag Health Study (HUNT) is a collaboration between the HUNT Research Center (Faculty of Medicine and Health Sciences, NTNU, Norwegian University of Science and Technology), the Trandelag County Council, the Central Norway Regional Health Authority, and the Norwegian Institute of Public Health. The genotyping in HUNT was financed by the National Institutes of Health, the Central Norway County Council, the Central Norway Regional Health Authority, and the Norwegian Institute of Public Health. The genotyping in HUNT was supported by the Norwegian University of Science and Technology, the Trandelag County Council, the Central Norway Regional Health Authority, and the Norwegian Institute of Public Health. The genotyping in HUNT was supported by the Norwegian University of Science and Technology, the Trandelag County Council, the Central Norway Regional Health Authority, and the Norwegian Institute of Public Health. The genotyping in HUNT was supported by the Norwegian University of Science and Technology, the Trandelag County Council, the Central Norway Regional Health Authority, and the Norwegian Institute of Public Health.
REFERENCES


STAR METHODS

KEY RESOURCES TABLE

<table>
<thead>
<tr>
<th>REAGENT or RESOURCE</th>
<th>SOURCE</th>
<th>IDENTIFIER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deposited data</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GBMI summary statistics</td>
<td>Zhou et al. 31</td>
<td><a href="https://www.globalbiobankmeta.org/resources">https://www.globalbiobankmeta.org/resources</a> and <a href="http://results.globalbiobankmeta.org">http://results.globalbiobankmeta.org</a></td>
</tr>
<tr>
<td>MegaStroke Ischemic stroke summary statistics</td>
<td>Malik et al. 6</td>
<td>GWAS catalog: GCST006908</td>
</tr>
<tr>
<td>Replication data</td>
<td>This paper</td>
<td>Table S2</td>
</tr>
<tr>
<td>GWAs catalog</td>
<td><a href="https://www.ebi.ac.uk/gwas/">https://www.ebi.ac.uk/gwas/</a></td>
<td>N/A</td>
</tr>
<tr>
<td>GTex</td>
<td>Burgess 32</td>
<td><a href="https://gtexportal.org/home/">https://gtexportal.org/home/</a></td>
</tr>
<tr>
<td>METAL</td>
<td>Willer et al. 51</td>
<td><a href="https://csg.sph.umich.edu/abecasis/metal/">https://csg.sph.umich.edu/abecasis/metal/</a></td>
</tr>
<tr>
<td>R package coloc</td>
<td><a href="https://cran.r-project.org/web/packages/coloc/index.html">https://cran.r-project.org/web/packages/coloc/index.html</a></td>
<td>N/A</td>
</tr>
<tr>
<td>R</td>
<td><a href="https://www.r-project.org">https://www.r-project.org</a></td>
<td>N/A</td>
</tr>
</tbody>
</table>

RESOURCE AVAILABILITY

Lead contact
Ida Surakka, University of Michigan, isurakka@umich.edu.

Materials availability
No materials were produced in this project.

Data and code availability
The GBMI summary statistics are available at https://www.globalbiobankmeta.org/resources and browsing at the browser http://results.globalbiobankmeta.org, and the MegaStroke summary statistics through GWAS Catalog (https://www.ebi.ac.uk/gwas/). No significant code was produced.

METHOD DETAILS

GBMI multi-ancestry meta-analysis
The GBMI ischemic stroke meta-analysis was conducted from genome-wide association results of 16 biobanks using inverse variance weighted meta-analysis. The overall stroke dataset has 1.9% of African/African American (AFR), 19.6% of East Asian (EAS), 75.8% of European (non-Finnish European: NFE and Finns: FIN), 1.1% of Latino or admixed American (AMR), and 1.6% of South Asian (SAS) ancestry (Figure 1). A detailed description of the meta-analysis methods can be found here. 31 The GBMI stroke phenotype was defined using PheCode 433.21 (Cerebral artery occlusion, with cerebral infarction).

Joint meta-analysis of GBMI and MegaStroke
Publicly available summary statistics for the MegaStroke 6 “All Ischemic Stroke” were downloaded from the GWAs-catalog (https://www.ebi.ac.uk/gwas/). The positions in the MegaStroke summary statistics were mapped to hg38 using liftover command line tool (https://genome.ucsc.edu/cgi-bin/hgLiftOver) and the overlapping SNPs between GBMI (34,124,639 variants) and MegaStroke (7,458,253 variants) were combined using METAL 51 resulting in meta-analysis of 5,633,313 SNPs (indels were excluded).

Replication cohorts
Replication of all 14 potentially novel lead variants were requested from two additional GBMI cohorts, Penn Medicine Biobank (PMBB, NFE: 1,226 cases, 25,228 controls, AFR: 787 cases, 9,319 controls) and Canadian Partnership for Tomorrow’s Health (CanPath, NFE: 70 cases, 7,190 controls). Analyses in these two cohorts were performed using the standard GBMI analysis pipeline. 31 Furthermore, we received replication results from Million Veteran Program (MVP) and Trans-Omics for Precision Medicine
(TOPMed) Program. The MVP analysis was performed using plink2a and with the EHR based stroke phenotype and covariates defined in Klarin et al. Analysis was completed within each HARE-defined ancestry group with 7,581/112,230 AFR (African), 21199/429518 EUR (European), 1983/49053 HIS (Hispanic) and 217/7979 ASN (Asian) ancestry cases/controls. The TOPMed results have been previously published and the analysis details can be found from the original publication. All TOPMed cohorts are included in the MegaStroke effort so none of the variants emerging from the joint meta-analysis were tested in the TOPMed in the replication stage. Additionally, TOPMed includes one overlapping biobank with GBMI (BioMe), which was not accounted for as neither of the two variants where TOPMed replication was used replicated.

**eQTL look-ups and colocalization analysis**

For each of the replicated novel variants, significant eQTLs were looked up from the GTEx database (https://gtexportal.org/home/). For all significant SNP x tissue specific gene expression pairs colocalization analysis was run using R package coloc (https://cran.r-project.org/web/packages/coloc/index.html) function coloc.abf() using the full GTEx cis-eQTL results (for each gene, 1Mb upstream and downstream included). We restricted the analysis to a ±500Kb window around the stroke associated lead variant.

**GWAs catalog data**

We downloaded reported significant (P-value < 5e-8) associations from the GWAs Catalog (https://www.ebi.ac.uk/gwas/) to define the novelty of our associated variants (GWAs catalog traits: Stroke, Ischemic Stroke) and to find any other phenotypic associations for our novel replicated variants (look-up by ±100Kb and ±250Kb window around the lead variant). Both data were downloaded on June 15th, 2022.

**QUANTIFICATION AND STATISTICAL ANALYSIS**

A detailed description of the GBMI meta-analysis methods can be found here. The joint meta-analysis of GBMI and MegaStroke was conducted using METAL inverse variance weighted meta-analysis. The replication results were combined using inverse variance weighted meta-analysis. The heterogeneity between ancestries and between sexes was tested using Cochran’s Q statistic implemented in METAL software. The colocalization of Stroke association and eQTL association was tested using R package coloc (https://cran.r-project.org/web/packages/coloc/index.html) with default prior probabilities (p1 = 1 x 10^-4, p2 = 1 x 10^-4, p12 = 1 x 10^-5).