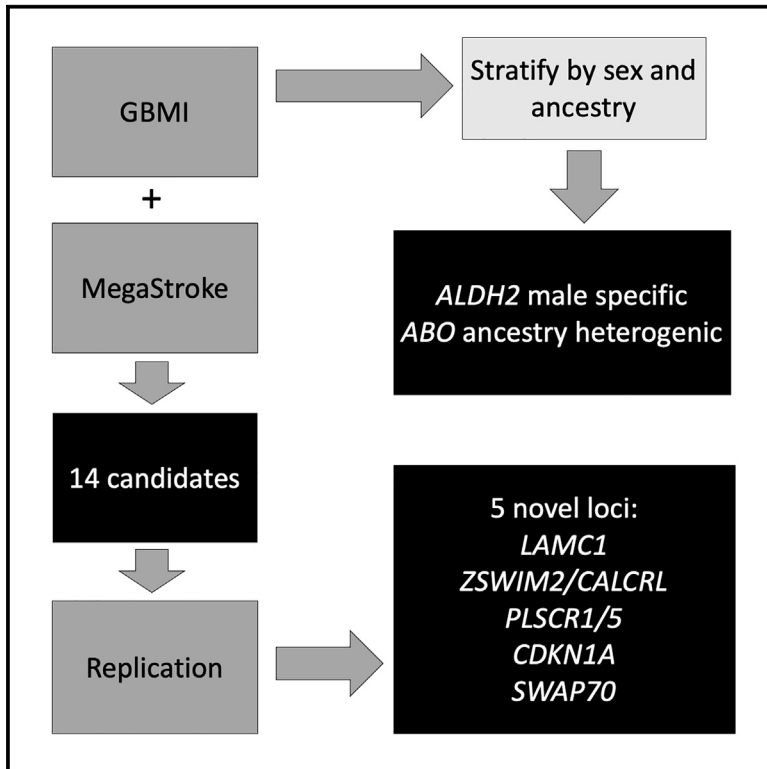# Multi-ancestry meta-analysis identifies 5 novel loci for ischemic stroke and reveals heterogeneity of effects between sexes and ancestries

Graphical abstract

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.

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

Article

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,[4,5,15,16] Benjamin M. Neale,[4,5,16] Cristen J. Willer[1,2,9,17] and on behalf of the Global Biobank Meta-analysis Initiative (GBMI)

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

Seventy-one loci have been identified using genome-wide association study (GWAS) methods[6–28] 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
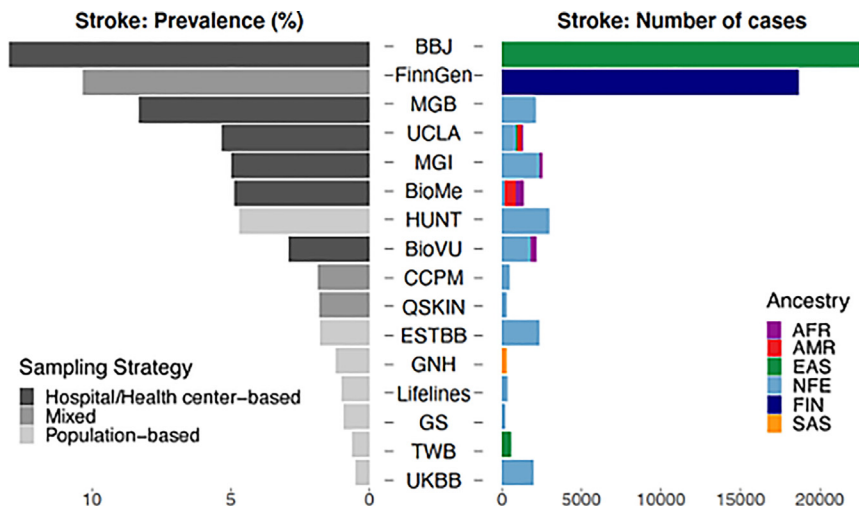
Figure 1. Breakdown of stroke meta-analysis ancestries

This figure presents the prevalence and number of cases by cohort participating in the GBMI stroke meta-analysis. BBJ, Biobank Japan; MGB, Mass General Brigham Biobank; UCLA, UCLA Precision Health Biobank; MGI, Michigan Genomics Initiative; BioMe, Mount Sinai BioMe Biobank; HUNT, Trøndelag Health Study; BioVU, Biorepository at Vanderbilt University; CCPM, Colorado Center for Personalized Medicine; QSKIN, Queensland Skin Study; ESTBB, Estonian Biobank; GNH, East London Genes & Health; GS, Generation Scotland; TWB, Taiwan Biobank; UKBB, UK Biobank; AFR, African ancestry; AMR, admixed American ancestry; EAS, East Asian ancestry; NFE, non-Finnish European ancestry; FIN, Finnish ancestry; SAS, South Asian ancestry.

to higher prevalence of that subtype (~80% of the cases) and thus more power to detect an association.[29] Additionally, stroke prevalence has been shown to differ between populations of different ancestry and sexes,[30] 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[6] 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[31] (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[6] (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,[32] 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[33,34] and related traits[2,35] (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 R^2 = 0.62). This suggests that the population-based sampling and the ICD-based phenotyping in the GBMI, in contrast to the majorly 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

Locus	rsID	Chromosome: position hg38	EA/ NEA	EAF	Discovery OR (95% CI)	Discovery p	Replication OR (95% CI)	Replication p
LAMC1[a]	rs4652772	1:183,054,561	C/G	0.560	1.03 (1.02; 1.05)	1.76e−8	1.03 (1.01; 1.05)	3.77e−4
ZSWIM2/ CALCRL	rs72896798	2:187,282,743	A/T	0.102	1.05 (1.03; 0.107)	3.22e−8	1.04 (1.02; 1.06)	8.49e−4
PLSCR1/PLSCR5	rs6762490	3:146,658,664	A/T	0.363	1.04 (1.02; 1.05)	1.48e−8	1.03 (1.02; 1.05)	1.51e−4
CDKN1A	rs762624	6:36,677,811	A/C	0.663	1.04 (1.03; 1.06)	3.65e−10	1.03 (1.02; 1.05)	1.2e−4
SWAP70[a]	rs360156	11:9,735,177	T/C	0.448	0.96 (0.95; 0.97)	2.45e−10	0.97 (0.97; 0.99)	5.3e−4

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.
[a]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 3/12) 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 (ALHD2, p < 1.6e−3, Bonferroni correction for 31 tests; Table S6). The ALDH2 locus has been associated with stroke in previous GWASs,[6] and the sex-heterogenic effects on ischemic stroke have been indirectly implied.[37] 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 HMGB1 expression,[38] which promotes atherosclerosis and has been implicated in multiple cardiovascular diseases.[39] CALCRL is known to regulate blood pressure, vascular biology, and inflammation.[40] 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[41] and has additionally been associated with

blood coagulation.[42] *CDKN1A* knockout mice have been shown to develop age-dependent cardiac hypertrophy.[43] Three of the putative novel loci, *PRDM16* (PR/SET domain 16), *GRK5* (G protein-coupled receptor kinase 5), and *PLEKHA1*, showed only nominal association in the replication. Of these, *PRDM16* and *GRK5* have been previously linked to cardiac hypertrophy.[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.[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.[47] The *ALDH2* exonic lead variant, rs671, is a well-known polymorphism linked to alcohol consumption and hypertension in EAS individuals.[37] 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,[13] this locus is associated with blood pressure traits in Chinese individuals with a higher body mass index,[48] and the expression of the *FGF5* gene has been shown to be different in patients with hypertension.[49] Additionally, recent studies have shown that the fibroblast growth factors (FGFs) could potentially be used to treat stroke in animal models.[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 despite the large overall size of the meta-analysis. First, the ischemic stroke phenotype was defined in the GBMI meta-analyses using an electronic health record (EHR)-derived phenotype definition only. This approach does not allow for dissection of results by subphenotypes, which has previously been done in large stroke meta-analysis efforts.[6] Our comparison analysis with the MegaStroke results showed that this ICD-code-based phenotyping together with the population-based sampling reduced the observed effect sizes, albeit with the caveat that MegaStroke results near genome-wide significance may be biased upward due to the winner's curse. Furthermore, with our investigation comparing the GBMI results in all the previously published stroke loci, we observed that only loci with an original association to more general phenotypes (stroke or ischemic stroke) were replicated in the GBMI meta-analysis. Secondly, for some of the ancestry groups, the number of participating biobanks and total

number of individuals are low (Finnish ancestry [FIN], one biobank n = 180,062; SAS, one biobank n = 21,940; AMR, two biobanks n = 15,064), resulting in limited power to fully test for ancestry-specific effects. Finally, as the two large efforts (GBMI and MegaStroke) combined in the discovery stage had majority of datasets imputed 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
- QUANTIFICATION AND STATISTICAL ANALYSIS

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, Huiling Zhao, Shinichi Namba, Ida Surakka, Brooke N. Wolford, Valeria Lo Faro, Esteban A. Lopera-Maya, Kristi Läll, Marie-Julie Favé, Juulia J. Partanen, Sinéad B. Chapman, Juha Karjalainen, Mitja Kurki, Mutaamba Maasha, Ben M. Brumpton, Sameer Chavan, Tzu-

Cell Genomics
Article

Ting Chen, Michelle Daya, Yi Ding, Yen-Chen A. Feng, Lindsay A. Guare, Christopher R. Gignoux, Sarah E. Graham, Whitney E. Hornsby, Nathan Ingold, Said I. Ismail, Ruth Johnson, Triin Laisk, Kuang Lin, Jun Lv, Iona Y. Millwood, Sonia Moreno-Grau, Kisung Nam, Priit Palta, Anita Pandit, Michael H. Preuss, Chadi Saad, Shefali S. Setia, Unnur Thorsteinsdottir, Jasmina Uzunovic, Anurag Verma, Matthew Zawistowski, Xue Zhong, Nahla Afifi, Kawthar M. Al-Dabhani, Asma Al Thani, Yuki Bradford, Archie Campbell, Kristy Crooks, Geertruida H. de Bock, Scott M. Damrauer, Nicholas J. Douville, Sarah Finer, Lars G. Fritsche, Eleni Fthenou, Gilberto Gonzalez-Arroyo, Christopher J. Griffiths, Yu Guo, Karen A. Hunt, Alexander Ioannidis, Nomdo M. Jansonius, Takahiro Konuma, Ming Ta Michael Lee, Arturo Lopez-Pineda, Yuta Matsuda, Riccardo E. Marioni, Babak Moatamed, Marco A. Nava-Aguilar, Kensuke Numakura, Snehal Patil, Nicholas Rafaels, Anne Richmond, Agustin Rojas-Muñoz, Jonathan A. Shortt, Peter Straub, Ran Tao, Brett Vanderwerff, Manvi Vernekar, Yogasudha Veturi, Kathleen C. Barnes, Marike Boezen, Zhengming Chen, Chia-Yen Chen, Judy Cho, George Davey Smith, Hilary K. Finucane, Lude Franke, Eric R. Gamazon, Andrea Ganna, Tom R. Gaunt, Tian Ge, Hailiang Huang, Jennifer Huffman, Nicholas Katsanis, Jukka T. Koskela, Clara Lajonchere, Matthew H. Law, Liming Li, Cecilia M. Lindgren, Ruth J.F. Loos, Stuart MacGregor, Koichi Matsuda, Catherine M. Olsen, David J. Porteous, Jordan A. Shavit, Harold Snieder, Tomohiro Takano, Richard C. Trembath, Judith M. Vonk, David C. Whiteman, Stephen J. Wicks, Cisca Wijmenga, John Wright, Jie Zheng, Xiang Zhou, Philip Awadalla, Michael Boehnke, Carlos D. Bustamante, Nancy J. Cox, Segun Fatumo, Daniel H. Geschwind, Caroline Hayward, Kristian Hveem, Eimear E. Kenny, Seunggeun Lee, Yen-Feng Lin, Hamdi Mbarek, Reedik Mägi, Hilary C. Martin, Sarah E. Medland, Yukinori Okada, Aarno V. Palotie, Bogdan Pasaniuc, Daniel J. Rader, Marylyn D. Ritchie, Serena Sanna, Jordan W. Smoller, Kari Stefansson, David A. van Heel, Robin G. Walters, Sebastian Zöllner, Biobank of the Americas, Biobank Japan Project, BioMe, BioVU, CanPath - Ontario Health Study, China Kadoorie Biobank Collaborative Group, Colorado Center for Personalized Medicine, deCODE Genetics, Estonian Biobank, FinnGen, Generation Scotland, Genes & Health Research Team, LifeLines, Mass General Brigham Biobank, Michigan Genomics Initiative, National Biobank of Korea, Penn Medicine BioBank, Qatar Biobank, The Qskin Sun and Health Study, Taiwan Biobank, The HUNT Study, UCLA ATLAS Community Health Initiative, Uganda Genome Resource, UK Biobank, Alicia R. Martin, Cristen J. Willer, Mark J. Daly, and Benjamin M. Neale.

The member of the NHLBI Trans-Omics for Precision Medicine (TOPmed) Stroke Working Group are Christopher D. Anderson, Tim Assimes, Paul Auer, Alexa Beiser, Joshua Bis, Cara Carty, Kei Hang Katie Chan, Jaeyoon Chung, John Cole, Adolfo Correa, Nicole Davis Armstrong, Ron Do, Zhaohui Du, Nauder Faraday, Myriam Fornage, Jeff Haessler, Scott Heemann, Yao Hu, Rebecca Jackson, Thomas Jaworek, Michelle Kim, Charles Kooperberg, Christina Kourkoulis, Leslie Lange, Simin Liu, Will Longstreth, Jessica Lundin, Tracy Madsen, Sandro Marini, Julie Mikulla, Braxton D. Mitchell, Paul Nyquist, James Perry, Ulrike Peters, Gloria Quach, Vasan S. Ramachandran, Alex Reiner, Stephen Rich, Jonathan Rosand, Jerome Rotter, Muralidharan Sargurupremrajm Chloé Sarnowski, Jenny Schoenberg, Sudha Seshadri, Brian Silver, Sylvia Smoller, David Tirschwell, Dhananjay Vaidya, Joanna von Berg, Huichun Xu, and Ying Zhou.

The members of the Million Veteran Program (MVP) are Samuel Aguayo, Sunil Ahuja, Kathrina Alexander, Dean P. Argyres, Prakash Balasubramanian, Zuhair Ballas, Jean Beckham, Sujata Bhushan, Edward Boyko, James Breeling, Jessica V. Brewer, Mary T. Brophy, Juan P. Casas, Kyong-Mi Chang, Kelly Cho, Lori Churby, David Cohen, Todd Connor, Joseph Constans, Louis Dellitalia, Nhan Do, Scott L. DuVall, Joseph Fayad, Daryl Fujii, Melinda Gaddy, Saib Gappy, J. Michael Gaziano, Michael Godschalk, Jennifer Greco, Samir Gupta, Salvador Gutierrez, Kimberly Hammer, Mark Hamner, John Harley, Elizabeth Hauser, Grant Huang, Donald E. Humphries, Adriana Hung, Mostaqul Huq, Robin Hurley, Pran Iruvanti, Frank Jacono, Darshana Jhala, Scott Kinlay, Jon Klein, Peter Liang, Suthat Liangpunsakul, Jack Lichy, Stephen Mastorides, Roy Mathew, Kristin Mattocks, Rachel McArdle, Paul Meyer, Jonathan Moorman, Timothy Morgan, Jennifer Moser, Sumitra Muralidhar, Maureen Murdoch, James Norton, Olaoluwa Okusaga, Kris Ann Oursler, Ana Palacio, Saiju Pyarajan, Rachel Ramoni, Nora Ratcliffe, Michael Rauchman, Brooks Robey, Luis E. Selva, Richard Servatius, Satish Sharma, Shahpoor Shayan, River Smith, Peruvemba Sriram, Todd Stapley, Brady Stephens, Patrick Strollo, Yan Sun, Neeraj Tandon, Philip S. Tsao, Gerardo Villareal, Agnes Wallbom, Jessica Walsh, John Wells, Stacey B. Whitbourne, Jeffrey Whittle, Mary Whooley, Peter Wilson, Junzhe Xu, Shing Shing Yeh, and Hongyu Zhao.

AUTHOR CONTRIBUTIONS

Study design, I.S., W.Z., M.J.D., B.M.N., and C.J.W.; bioinformatic analysis, I.S., K.-H.W., B.N.W., W.Z., A.P., Y.H., B.B.; HUNT, I.S., B.B., A.H.S., M.E.G., and K.H.; TOPMed replication, Y.H.; CKMVP replication, J.E.H. and P.W.F.W.; writing, I.S., K.-H.W., F.S., A.P., W.H., N.R.S., and C.J.W.; revision, I.S., K.-H.W., B.N.W., F.S., W.H., R.G.W., M.J.D., S.Z., N.R.S., B.M.N., and C.J.W.

DECLARATION OF INTERESTS

N.R.S. is an advisor for Abbott, Philips, and Shockwave and has received honoraria for speaking from Zoll, Cordis. C.J.W.'s spouse works at Regeneron Pharmaceuticals.

REFERENCES

1. Katan, M., and Luft, A. (2018). Global burden of stroke (2018). Semin. Neurol. *38*, 208–211.

2. Nelson, C.P., Goel, A., Butterworth, A.S., Kanoni, S., Webb, T.R., Marouli, E., Zeng, L., Ntalla, I., Lai, F.Y., Hopewell, J.C., et al. (2017). Association analyses based on false discovery rate implicate new loci for coronary artery disease. Nat. Genet. *49*, 1385–1391.

3. Mahajan, A., Taliun, D., Thurner, M., Robertson, N.R., Torres, J.M., Rayner, N.W., Payne, A.J., Steinthorsdottir, V., Scott, R.A., Grarup, N., et al. (2018). Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. Nat. Genet. *50*, 1505–1513.

4. Hankey, G.J. (2017). Stroke. Lancet *389*, 641–654.

5. Bevan, S., Traylor, M., Adib-Samii, P., Malik, R., Paul, N.L.M., Jackson, C., Farrall, M., Rothwell, P.M., Sudlow, C., Dichgans, M., and Markus, H.S. (2012). Genetic heritability of ischemic stroke and the contribution of previously reported candidate gene and genomewide associations. Stroke *43*, 3161–3167.

6. Malik, R., Chauhan, G., Traylor, M., Sargurupremraj, M., Okada, Y., Mishra, A., Rutten-Jacobs, L., Giese, A.K., van der Laan, S.W., Gretarsdottir, S., et al. (2018). Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. Nat. Genet. *50*, 524–537.

7. Malik, R., Rannikmäe, K., Traylor, M., Georgakis, M.K., Sargurupremraj, M., Markus, H.S., Hopewell, J.C., Debette, S., Sudlow, C.L.M., and Dichgans, M.; MEGASTROKE consortium and the International Stroke Genetics Consortium (2018). Genome-wide meta-analysis identifies 3 novel loci associated with stroke. Ann. Neurol. *84*, 934–939.

8. Woo, D., Falcone, G.J., Devan, W.J., Brown, W.M., Biffi, A., Howard, T.D., Anderson, C.D., Brouwers, H.B., Valant, V., Battey, T.W.K., et al. (2014). Meta-analysis of genome-wide association studies identifies 1q22 as a susceptibility locus for intracerebral hemorrhage. Am. J. Hum. Genet. *94*, 511–521.

9. NINDS Stroke Genetics Network SiGN; International Stroke Genetics Consortium ISGC (2016). Loci associated with ischaemic stroke and its subtypes (SiGN): a genome-wide association study. Lancet Neurol. *15*, 174–184.

10. von Berg, J., van der Laan, S.W., McArdle, P.F., Malik, R., Kittner, S.J., Mitchell, B.D., Worrall, B.B., de Ridder, J., and Pulit, S.L. (2020). Alternate approach to stroke phenotyping identifies a genetic risk locus for small vessel stroke. Eur. J. Hum. Genet. *28*, 963–972.

11. Yamada, Y., Kato, K., Oguri, M., Horibe, H., Fujimaki, T., Yasukochi, Y., Takeuchi, I., and Sakuma, J. (2018). Identification of nine genes as novel susceptibility loci for early-onset ischemic stroke, intracerebral hemorrhage, or subarachnoid hemorrhage. Biomed. Rep. *9*, 8–20.

12. Cheng, Y.-C., Stanne, T.M., Giese, A.K., Ho, W.K., Traylor, M., Amouyel, P., Holliday, E.G., Malik, R., Xu, H., Kittner, S.J., et al. (2016). Genome-wide association analysis of young-onset stroke identifies a locus on chromosome 10q25 near HABP2. Stroke *47*, 307–316.

13. Sakaue, S., Kanai, M., Tanigawa, Y., Karjalainen, J., Kurki, M., Koshiba, S., Narita, A., Konuma, T., Yamamoto, K., Akiyama, M., et al. (2021). A cross-population atlas of genetic associations for 220 human phenotypes. Nat. Genet. *53*, 1415–1424.

14. Kumar, A., Chauhan, G., Sharma, S., Dabla, S., Sylaja, P.N., Chaudhary, N., Gupta, S., Agrawal, C.S., Anand, K.S., Srivastava, A.K., et al. (2021). Association of SUMOylation pathway genes with stroke in a genome-wide association study in India. Neurology *97*, e345–e356.

15. Ishigaki, K., Akiyama, M., Kanai, M., Takahashi, A., Kawakami, E., Sugishita, H., Sakaue, S., Matoba, N., Low, S.K., Okada, Y., et al. (2020). Large-scale genome-wide association study in a Japanese population identifies novel susceptibility loci across different diseases. Nat. Genet. *52*, 669–679.

16. Traylor, M., Persyn, E., Tomppo, L., Klasson, S., Abedi, V., Bakker, M.K., Torres, N., Li, L., Bell, S., Rutten-Jacobs, L., et al. (2021). Genetic basis of lacunar stroke: a pooled analysis of individual patient data and genome-wide association studies. Lancet Neurol. *20*, 351–361.

17. Traylor, M., Malik, R., Nalls, M.A., Cotlarciuc, I., Radmanesh, F., Thorleifsson, G., Hanscombe, K.B., Langefeld, C., Saleheen, D., Rost, N.S., et al. (2017). Genetic variation at 16q24.2 is associated with small vessel stroke. Ann. Neurol. *81*, 383–394.

18. Keene, K.L., Hyacinth, H.I., Bis, J.C., Kittner, S.J., Mitchell, B.D., Cheng, Y.C., Pare, G., Chong, M., O'Donnell, M., Meschia, J.F., et al. (2020). Genome-wide association study meta-analysis of stroke in 22 000 individuals of African descent identifies novel associations with stroke. Stroke *51*, 2454–2463.

19. Ikram, M.A., Seshadri, S., Bis, J.C., Fornage, M., DeStefano, A.L., Aulchenko, Y.S., Debette, S., Lumley, T., Folsom, A.R., van den Herik, E.G., et al. (2009). Genomewide association studies of stroke. N. Engl. J. Med. *360*, 1718–1728.

20. Malik, R., Dau, T., Gonik, M., Sivakumar, A., Deredge, D.J., Edeleva, E.V., Götzfried, J., van der Laan, S.W., Pasterkamp, G., Beaufort, N., et al. (2017). Common coding variant in SERPINA1 increases the risk for large artery stroke. Proc. Natl. Acad. Sci. USA *114*, 3613–3618.

21. Carty, C.L., Keene, K.L., Cheng, Y.C., Meschia, J.F., Chen, W.M., Nalls, M., Bis, J.C., Kittner, S.J., Rich, S.S., Tajuddin, S., et al. (2015). Meta-analysis of genome-wide association studies identifies genetic risk factors for stroke in African Americans. Stroke *46*, 2063–2068.

22. Traylor, M., Farrall, M., Holliday, E.G., Sudlow, C., Hopewell, J.C., Cheng, Y.C., Fornage, M., Ikram, M.A., Malik, R., Bevan, S., et al. (2012). Genetic risk factors for ischaemic stroke and its subtypes (the METASTROKE collaboration): a meta-analysis of genome-wide association studies. Lancet Neurol. *11*, 951–962.

23. Song, Y., Choi, J.E., Kwon, Y.J., Chang, H.J., Kim, J.O., Park, D.H., Park, J.M., Kim, S.J., Lee, J.W., and Hong, K.W. (2021). Identification of susceptibility loci for cardiovascular disease in adults with hypertension, diabetes, and dyslipidemia. J. Transl. Med. *19*, 85.

24. Neurology Working Group of the Cohorts for Heart and Aging Research in Genomic Epidemiology CHARGE Consortium, the Stroke Genetics Network SiGN, and the International Stroke Genetics Consortium ISGC (2016). Identification of additional risk loci for stroke and small vessel disease: a meta-analysis of genome-wide association studies. Lancet Neurol. *15*, 695–707.

25. Dichgans, M., Malik, R., König, I.R., Rosand, J., Clarke, R., Gretarsdottir, S., Thorleifsson, G., Mitchell, B.D., Assimes, T.L., Levi, C., et al. (2014). Shared genetic susceptibility to ischemic stroke and coronary artery disease: a genome-wide analysis of common variants. Stroke *45*, 24–36.

26. Gretarsdottir, S., Thorleifsson, G., Manolescu, A., Styrkarsdottir, U., Helgadottir, A., Gschwendtner, A., Kostulas, K., Kuhlenbäumer, G., Bevan, S., Jonsdottir, T., et al. (2008). Risk variants for atrial fibrillation on chromosome 4q25 associate with ischemic stroke. Ann. Neurol. *64*, 402–409.

27. Holliday, E.G., Maguire, J.M., Evans, T.J., Koblar, S.A., Jannes, J., Sturm, J.W., Hankey, G.J., Baker, R., Golledge, J., Parsons, M.W., et al. (2012). Common variants at 6p21.1 are associated with large artery atherosclerotic stroke. Nat. Genet. *44*, 1147–1151.

28. International Stroke Genetics Consortium ISGC; Wellcome Trust Case Control Consortium 2 WTCCC2; Bellenguez, C., Bevan, S., Gschwendtner, A., Spencer, C.C.A., Burgess, A.I., Pirinen, M., Jackson, C.A., Traylor, M., et al. (2012). Genome-wide association study identifies a variant in HDAC9 associated with large vessel ischemic stroke. Nat. Genet. *44*, 328–333.

29. Donkor, E.S. (2018). Stroke in the 21st century: a snapshot of the burden, epidemiology and quality of life. Stroke Res. Treat. *2018*, 3238165.

30. Guzik, A., and Bushnell, C. (2017). Stroke epidemiology and risk factor management. Continuum *23*, 15–39.

31. Zhou, W., Kanai, M., Wu, K.H.H., Rasheed, H., Tsuo, K., Hirbo, J.B., Wang, Y., Bhattacharya, A., Zhao, H., Namba, S., and Surakka, I. (2022). Global Biobank Meta-analysis Initiative: powering genetic discovery across human diseases. Cell Genom. *2*, 100192.

32. Burgess, D.J. (2020). Reaching completion for GTEx. Nat. Rev. Genet. *21*, 717.

33. Graham, S.E., Clarke, S.L., Wu, K.H.H., Kanoni, S., Zajac, G.J.M., Ramdas, S., Surakka, I., Ntalla, I., Vedantam, S., Winkler, T.W., et al. (2021). The power of genetic diversity in genome-wide association studies of lipids. Nature *600*, 675–679.

34. Evangelou, E., Warren, H.R., Mosen-Ansorena, D., Mifsud, B., Pazoki, R., Gao, H., Ntritsos, G., Dimou, N., Cabrera, C.P., Karaman, I., et al. (2018). Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. Nat. Genet. *50*, 1412–1425.

35. Mahajan, A., Spracklen, C.N., Zhang, W., Ng, M.C.Y., Petty, L.E., Kitajima, H., Yu, G.Z., Rüeger, S., Speidel, L., Kim, Y.J., et al. (2022). Multi-ancestry genetic study of type 2 diabetes highlights the power of diverse populations for discovery and translation. Nat. Genet. *54*, 560–572.

36. Mishra, A., Malik, R., Hachiya, T., Jürgenson, T., Namba, S., Posner, D.C., Kamanu, F.K., Koido, M., Le Grand, Q., Shi, M., et al. (2022). Stroke genetics informs drug discovery and risk prediction across ancestries. Nature *611*, 115–123.

37. Millwood, I.Y., Walters, R.G., Mei, X.W., Guo, Y., Yang, L., Bian, Z., Bennett, D.A., Chen, Y., Dong, C., Hu, R., et al. (2019). Conventional and genetic evidence on alcohol and vascular disease aetiology: a prospective study of 500 000 men and women in China. Lancet *393*, 1831–1842.

38. Hebbel, R.P., Wei, P., Milbauer, L., Corban, M.T., Solovey, A., Kiley, J., Pattee, J., Lerman, L.O., Pan, W., and Lerman, A. (2020). Abnormal endothelial gene expression associated with early coronary atherosclerosis. J. Am. Heart Assoc. *9*, e016134.

39. Cai, J., Wen, J., Bauer, E., Zhong, H., Yuan, H., and Chen, A.F. (2015). The role of HMGB1 in cardiovascular biology: danger signals. Antioxidants Redox Signal. *23*, 1351–1369.

40. Angenendt, L., Bormann, E., Pabst, C., Alla, V., Görlich, D., Braun, L., Dohlich, K., Schwöppe, C., Bohlander, S.K., Arteaga, M.F., et al. (2019). The neuropeptide receptor calcitonin receptor-like (CALCRL) is a potential therapeutic target in acute myeloid leukemia. Leukemia *33*, 2830–2841.

41. Bassé, F., Stout, J.G., Sims, P.J., and Wiedmer, T. (1996). Isolation of an erythrocyte membrane protein that mediates Ca2+-dependent transbilayer movement of phospholipid. J. Biol. Chem. *271*, 17205–17210.

42. Bing, P.-F., Xia, W., Wang, L., Zhang, Y.H., Lei, S.F., and Deng, F.Y. (2016). Common marker genes identified from various sample types for systemic lupus erythematosus. PLoS One *11*, e0156234.

43. Hauck, L., Grothe, D., and Billia, F. (2016). p21(CIP1/WAF1)-dependent inhibition of cardiac hypertrophy in response to Angiotensin II involves Akt/Myc and pRb signaling. Peptides *83*, 38–48.

44. Cibi, D.M., Bi-Lin, K.W., Shekeran, S.G., Sandireddy, R., Tee, N., Singh, A., Wu, Y., Srinivasan, D.K., Kovalik, J.P., Ghosh, S., et al. (2020). Prdm16 deficiency leads to age-dependent cardiac hypertrophy, adverse remodeling, mitochondrial dysfunction, and heart failure. Cell Rep. *33*, 108288.

45. Hullmann, J.E., Grisanti, L.A., Makarewich, C.A., Gao, E., Gold, J.I., Chuprun, J.K., Tilley, D.G., Houser, S.R., and Koch, W.J. (2014). GRK5-mediated exacerbation of pathological cardiac hypertrophy involves facilitation of nuclear NFAT activity. Circ. Res. *115*, 976–985.

46. Paré, G., Chasman, D.I., Parker, A.N., Zee, R.R.Y., Mälarstig, A., Seedorf, U., Collins, R., Watkins, H., Hamsten, A., Miletich, J.P., and Ridker, P.M. (2009). Novel associations of CPS1, MUT, NOX4, and DPEP1 with plasma homocysteine in a healthy population: a genome-wide evaluation of 13 974 participants in the Women's Genome Health Study. Circ. Cardiovasc. Genet. *2*, 142–150.

47. Larsson, S.C., Traylor, M., and Markus, H.S. (2019). Homocysteine and small vessel stroke: a mendelian randomization analysis. Ann. Neurol. *85*, 495–501.

48. Li, J., Shi, J., Huang, W., Sun, J., Wu, Y., Duan, Q., Luo, J., Lange, L.A., Gordon-Larsen, P., Zheng, S.L., and Yuan, W. (2015). Variant near FGF5 has stronger effects on blood pressure in Chinese with a higher a body mass index. Am. J. Hypertens. *28*, 1031–1037.

49. Ren, Y., Jiao, X., and Zhang, L. (2018). Expression level of fibroblast growth factor 5 (FGF5) in the peripheral blood of primary hypertension and its clinical significance. Saudi J. Biol. Sci. *25*, 469–473.

50. Dordoe, C., Chen, K., Huang, W., Chen, J., Hu, J., Wang, X., and Lin, L. (2021). Roles of fibroblast growth factors and their therapeutic potential in treatment of ischemic stroke. Front. Pharmacol. *12*, 671131.

51. Willer, C.J., Li, Y., and Abecasis, G.R. (2010). METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics *26*, 2190–2191.

52. Hunter-Zinck, H., Shi, Y., Li, M., Gorman, B.R., Ji, S.G., Sun, N., Webster, T., Liem, A., Hsieh, P., Devineni, P., et al. (2020). Genotyping array design and data quality control in the Million Veteran Program. Am. J. Hum. Genet. *106*, 535–548.

53. Klarin, D., Damrauer, S.M., Cho, K., Sun, Y.V., Teslovich, T.M., Honerlaw, J., Gagnon, D.R., DuVall, S.L., Li, J., Peloso, G.M., et al. (2018). Genetics of blood lipids among ~300,000 multi-ethnic participants of the Million Veteran Program. Nat. Genet. *50*, 1514–1523.

54. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience *4*, 7.

55. Fang, H., Hui, Q., Lynch, J., Honerlaw, J., Assimes, T.L., Huang, J., Vujkovic, M., Damrauer, S.M., Pyarajan, S., Gaziano, J.M., and DuVall, S.L. (2019). Harmonizing genetic ancestry and self-identified race/ethnicity in genome-wide association studies. Am. J. Hum. Genet. *105*, 763–772.

56. Hu, Y., Haessler, J.W., Manansala, R., Wiggins, K.L., Moscati, A., Beiser, A., Heard-Costa, N.L., Sarnowski, C., Raffield, L.M., Chung, J., et al. (2022). Whole-genome sequencing association analyses of stroke and its subtypes in ancestrally diverse populations from Trans-Omics for Precision Medicine project. Stroke *53*, 875–885.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
GBMI summary statistics	Zhou et al.[31]	https://www.globalbiobankmeta.org/resources and http://results.globalbiobankmeta.org
MegaStroke Ischemic stroke summary statistics	Malik et al.[6]	GWAS catalog: GCST006908
Replication data	This paper	Table S2
GWAs catalog	https://www.ebi.ac.uk/gwas/	N/A
GTeX	Burgess[32]	https://gtexportal.org/home/
Software and algorithms		
METAL	Willer et al.[51]	https://csg.sph.umich.edu/abecasis/metal/
R package coloc	https://cran.r-project.org/web/packages/coloc/index.html	N/A
R	https://www.r-project.org	N/A

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[52,53] analysis was performed using plink2a[54] and with the EHR based stroke phenotype and covariates defined in Klarin et al.[53] Analysis was completed within each HARE-defined ancestry group[55] 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.[56] 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[32] (https://gtexportal.org/home/). For all significant SNP × 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.[31] The joint meta-analysis of GBMI and MegaStroke was conducted using METAL[51] 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 = 1×10^{-4}, p2 = 1×10^{-4}, p12 = 1×10^{-5}).