Multi-ancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes

Rainer Malik [1*], Ganesh Chauhan [2*], Matthew Traylor [3*], Muralidharan Sargurupremraj [4,5*], Yukinori Okada [6,7,8*], Aniket Mishra [4,5], Loes Rutten-Jacobs [3], Anne-Katrin Giese [9], Sander W van der Laan [10], Solveig Gretarsdottir [11], Christopher D Anderson [12,13,14,14], Michael Chong [15], Hieab HH Adams [16,17], Tetsuro Ago [18], Peter Almgren [19], Philippe Amouyel [20,21], Hakan Ay [22,13], Traci M Bartz [23], Oscar R Benavente [24], Steve Bevan [25], Giorgio B Boncoraglio [26], Robert D Brown, Jr. [27], Adam S Butterworth [28,29], Caty Carrera [30,31], Cara L Carty [32,33], Daniel I Chasman [34,35], Wei-Min Chen [36], John W Cole [37], Adolfo Correa [38], Ioana Cotlarciuc [39], Carlos Cruchaga [40,41], John Danesh [28,42,43,44], Paul IW de Bakker [45,46], Anita L DeStefano [47,48], Marcel den Hoed [49], Qing Duan [50], Stefan T Engelter [51,52], Guido J Falcone [53,54], Rebecca F Gottesman [55], Raji P Grewal [56], Vilmundur Gudnason [57,58], Stefan Gustafsson [59], Jeffrey Haessler [60], Tamara B Harris [61], Ahamad Hassan [62], Aki S Havulinna [63,64], Susan R Heckbert [65], Elizabeth G Holliday [66,67], George Howard [68], Fang-Chi Hsu [69], Hyacinth I Hyacinth [70], M Arfan Ikram [16], Erik Ingelsson [71,72], Marguerite R Irvin [73], Xueqiu Jian [74], Jordi Jiménez-Conde [75], Julie A Johnson [76,77], J Wouter Jukema [78], Masahiro Kanai [6,7,79], Keith L Keene [80,81], Brett M Kissela [82], Dawn O Kleindorfer [82], Charles Kooperberg [60], Michiaki Kubo [83], Leslie A Lange [84], Carl D Langefeld [85], Claudia Langenberg [86], Lenore J Launer [87], Jin-Moo Lee [88], Robin Lemmens [89,90], Didier Leys [91], Cathryn M Lewis [92,93], Wei-Yu Lin [28,94], Arne G Lindgren [95,96], Erik Lorentzen [97], Patrik K Magnusson [98], Jane Maguire [99], Ani Manichaikul [36], Patrick F McArdle [100], James F Meschia [101], Braxton D Mitchell [100,102], Thomas H Mosley [103,104], Michael A Nalls [105,106], Toshiharu Ninomiya [107], Martin J O'Donnell [15,108], Bruce M Psaty [109,110,111,112], Sara L Pulit [113,45], Kristiina Rannikmäe [114,115], Alexander P Reiner [65,116], Kathryn M Rexrode [117], Kenneth Rice [118], Stephen S Rich [36], Paul M Ridker [34,35], Natalia S Rost [9,13], Peter M Rothwell [119], Jerome I Rotter [120,121], Tatjana Rundek [122], Ralph L Sacco [122], Saori Sakaue [7,123], Michele M Sale [124], Veikko Salomaa [63], Bishwa R Sapkota [125], Reinhold Schmidt [126], Carsten O Schmidt [127], Ulf Schminke [128], Pankaj Sharma [39], Agnieszka Slowik [129], Cathie LM Sudlow [114,115], Christian Tanislav [130], Turgut Tatlisumak [131,132], Kent D Taylor [120,121], Vincent NS Thijs [133,134], Gudmar Thorleifsson [11], Unnur Thorsteinsdottir [11], Steffen Tiedt [1], Stella Trompet [135], Christophe Tzourio [5,136,137], Cornelia M van Duijn [138,139], Matthew Walters [140], Nicholas J Wareham [86], Sylvia Wassertheil-Smoller [141], James G Wilson [142], Kerri L Wiggins [109], Qiong Yang [47], Salim Yusuf [15], AFGen consortium [143], Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium [143], International Genomics of Blood Pressure (iGEN-BP) Consortium [143], INVENT consortium [143], STARNET [143], Joshua C Bis [109], Tomi Pastinen [144], Arno Ruusalepp [145,146,147], Eric E Schadt [148], Simon Koplev [148], Johan LM Björkegren [148,149,150,151], Veronica Codoni [152,153], Mete Civelek [124,154], Nicholas L Smith [65,155,156], David A Trégouët [152,153], Ingrid E Christophersen [54,157,158], Carolina Roselli [54], Steven A Lubitz [54,157], Patrick T Ellinor [54,157], E Shyong Tai [159], Jaspal S Kooner [160], Norihiro Kato [161], Jiang He [162], Pim van der Harst [163], Paul Elliott [164], John C Chambers [165,166], Fumihiko Takeuchi [161], Andrew D Johnson [167,48], BioBank Japan Cooperative Hospital Group [143], COMPASS consortium [143], EPIC-CVD consortium [143], EPIC-InterAct consortium [143], International Stroke Genetics Consortium (ISGC) [143], METASTROKE Consortium [143], Neurology Working Group of the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium [143], NINDS Stroke Genetics Network (SiGN) [143], UK Young Lacunar DNA Study [143], MEGASTROKE Consortium [168], Dharambir K Sanghera [125,169,170], Olle Melander [19], Christina Jern [171], Daniel Strbian [172,173], Israel Fernandez-Cadenas [31,30], W T Longstreth, Jr [174,65], Arndt Rolfs [175], Jun Hata [107], Daniel Woo [82], Jonathan Rosand [12,13,14,14], Guillaume Pare [15], Jemma C Hopewell [176], Danish Saleheen [177], Kari Stefansson [11,178*], Bradford B Worrall [179*], Steven J Kittner [37*], Sudha Seshadri [180,48*], Myriam Fornage [74,181*], Hugh S Markus [3*], Joanna MM Howson [28*], Yoichiro Kamatani [6,182*], Stephanie Debette [4,5*§] and Martin Dichgans [1,183,184*§]

1 Institute for Stroke and Dementia Research (ISD), University Hospital, LMU Munich, Munich, Germany
2 Centre for Brain Research, Indian Institute of Science, Bangalore, India
3 Stroke Research Group, Division of Clinical Neurosciences, University of Cambridge, UK
4 INSERM U1219 Bordeaux Population Health Research Center, Bordeaux, France

5 University of Bordeaux, Bordeaux, France
6 Laboratory for Statistical Analysis, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan
7 Department of Statistical Genetics, Osaka University Graduate School of Medicine, Osaka, Japan
8 Laboratory of Statistical Immunology, Immunology Frontier Research Center (WPI-IFReC), Osaka University, Suita, Japan.
9 Department of Neurology, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA
10 Laboratory of Experimental Cardiology, Division of Heart and Lungs, University Medical Center Utrecht, University of Utrecht, Utrecht,Netherlands
11 deCODE genetics/AMGEN inc, Reykjavik, Iceland
12 Center for Genomic Medicine, Massachusetts General Hospital (MGH), Boston, MA, USA
13 J. Philip Kistler Stroke Research Center, Department of Neurology, MGH, Boston, MA, USA
14 Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, USA
15 Population Health Research Institute, McMaster University, Hamilton, Canada
16 Department of Epidemiology, Erasmus University Medical Center, Rotterdam, Netherlands
17 Department of Radiology and Nuclear Medicine, Erasmus University Medical Center, Rotterdam, Netherlands
18 Department of Medicine and Clinical Science, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan
19 Department of Clinical Sciences, Lund University, Malmö, Sweden
20 Univ. Lille, Inserm, Institut Pasteur de Lille, LabEx DISTALZ-UMR1167, Risk factors and molecular determinants of aging-related diseases, F-59000 Lille, France
21 Centre Hosp. Univ Lille, Epidemiology and Public Health Department, F-59000 Lille, France
22 AA Martinos Center for Biomedical Imaging, Department of Radiology, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA
23 Cardiovascular Health Research Unit, Departments of Biostatistics and Medicine, University of Washington, Seattle, WA, USA
24 Division of Neurology, Faculty of Medicine, Brain Research Center, University of British Columbia, Vancouver, Canada
25 School of Life Science, University of Lincoln, Lincoln, UK
26 Department of Cerebrovascular Diseases, Fondazione IRCCS Istituto Neurologico "Carlo Besta", Milano, Italy
27 Department of Neurology, Mayo Clinic Rochester, Rochester, MN, USA
28 MRC/BHF Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK
29 The National Institute for Health Research Blood and Transplant Research Unit in Donor Health and Genomics, University of Cambridge, UK
30 Neurovascular Research Laboratory, Vall d'Hebron Institut of Research, Neurology and Medicine Departments-Universitat Autònoma de Barcelona, Vall d’Hebrón Hospital, Barcelona, Spain
31 Stroke Pharmacogenomics and Genetics, Fundacio Docència i Recerca MutuaTerrassa, Terrassa, Spain
32 Children's Research Institute, Children's National Medical Center, Washington, DC, USA
33 Center for Translational Science, George Washington University, Washington, DC, USA
34 Division of Preventive Medicine, Brigham and Women's Hospital, Boston, MA, USA
35 Harvard Medical School, Boston, MA, USA
36 Center for Public Health Genomics, Department of Public Health Sciences, University of Virginia, Charlottesville, VA, USA
37 Department of Neurology, University of Maryland School of Medicine and Baltimore VAMC, Baltimore, MD, USA
38 Departments of Medicine, Pediatrics and Population Health Science, University of Mississippi Medical Center, Jackson, MS, USA
39 Institute of Cardiovascular Research, Royal Holloway University of London, UK & Ashford and St Peters Hospital, Surrey UK
40 Department of Psychiatry,The Hope Center Program on Protein Aggregation and Neurodegeneration (HPAN),Washington University, School of Medicine, St. Louis, MO, USA

41 Department of Developmental Biology, Washington University School of Medicine, St. Louis, MO, USA
42 NIHR Blood and Transplant Research Unit in Donor Health and Genomics, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK
43 Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK
44 British Heart Foundation, Cambridge Centre of Excellence, Department of Medicine, University of Cambridge, Cambridge, UK
45 Department of Medical Genetics, University Medical Center Utrecht, Utrecht, Netherlands
46 Department of Epidemiology, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, Netherlands
47 Boston University School of Public Health, Boston, MA, USA
48 Framingham Heart Study, Framingham, MA, USA
49 Department of Immunology, Genetics and Pathology and Science for Life Laboratory, Uppsala University, Uppsala, Sweden
50 Department of Genetics, University of North Carolina, Chapel Hill, NC, USA
51 Department of Neurology and Stroke Center, Basel University Hospital, Switzerland
52 Neurorehabilitation Unit, University and University Center for Medicine of Aging and Rehabilitation Basel, Felix Platter Hospital, Basel, Switzerland
53 Department of Neurology, Yale University School of Medicine, New Haven, CT, USA
54 Program in Medical and Population Genetics, The Broad Institute of Harvard and MIT, Cambridge, MA, USA
55 Department of Neurology, Johns Hopkins University School of Medicine, Baltimore, MD, USA
56 Neuroscience Institute, SF Medical Center, Trenton, NJ, USA
57 Icelandic Heart Association Research Institute, Kopavogur, Iceland
58 University of Iceland, Faculty of Medicine, Reykjavik, Iceland
59 Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Uppsala, Sweden
60 Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, USA
61 Laboratory of Epidemiology and Population Science, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA
62 Department of Neurology, Leeds General Infirmary, Leeds Teaching Hospitals NHS Trust, Leeds, UK
63 National Institute for Health and Welfare, Helsinki, Finland
64 FIMM - Institute for Molecular Medicine Finland, Helsinki, Finland
65 Department of Epidemiology, University of Washington, Seattle, WA, USA
66 Public Health Stream, Hunter Medical Research Institute, New Lambton, Australia
67 Faculty of Health and Medicine, University of Newcastle, Newcastle, Australia
68 School of Public Health, University of Alabama at Birmingham, Birmingham, AL, USA
69 Department of Biostatistical Sciences, Wake Forest School of Medicine, Winston-Salem, NC, USA
70 Aflac Cancer and Blood Disorder Center, Department of Pediatrics, Emory University School of Medicine, Atlanta, GA, USA
71 Department of Medicine, Division of Cardiovascular Medicine, Stanford University School of Medicine, CA, USA
72 Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Uppsala, Sweden
73 Epidemiology, School of Public Health, University of Alabama at Birmingham, USA
74 Brown Foundation Institute of Molecular Medicine, University of Texas Health Science Center at Houston, Houston, TX, USA
75 Neurovascular Research Group (NEUVAS), Neurology Department, Institut Hospital del Mar d'Investigació Mèdica, Universitat Autònoma de Barcelona, Barcelona, Spain
76 Department of Pharmacotherapy and Translational Research and Center for Pharmacogenomics, University of Florida, College of Pharmacy, Gainesville, FL, USA
77 Division of Cardiovascular Medicine, College of Medicine, University of Florida, Gainesville, FL, USA
78 Department of Cardiology, Leiden University Medical Center, Leiden, the Netherlands
79 Program in Bioinformatics and Integrative Genomics, Harvard Medical School, Boston, MA, USA

80 Department of Biology, East Carolina University, Greenville, NC, USA
81 Center for Health Disparities, East Carolina University, Greenville, NC, USA
82 University of Cincinnati College of Medicine, Cincinnati, OH, USA
83 RIKEN Center for Integrative Medical Sciences, Yokohama, Japan
84 Department of Medicine, University of Colorado Denver, Anschutz Medical Campus, Aurora, CO, USA
85 Center for Public Health Genomics and Department of Biostatistical Sciences, Wake Forest School of Medicine, Winston-Salem, NC, USA
86 MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine, Institute of Metabolic Science, Cambridge Biomedical Campus, Cambridge, UK
87 Intramural Research Program, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA
88 Department of Neurology, Radiology, and Biomedical Engineering, Washington University School of Medicine, St. Louis, MO, USA
89 KU Leuven – University of Leuven, Department of Neurosciences, Experimental Neurology, Leuven, Belgium
90 VIB Center for Brain & Disease Research, University Hospitals Leuven, Department of Neurology, Leuven, Belgium
91 Univ.-Lille, INSERM U 1171. CHU Lille. Lille, France
92 Department of Medical and Molecular Genetics, King's College London, London, UK
93 SGDP Centre, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK
94 Northern Institute for Cancer Research, Paul O'Gorman Building, Newcastle University, Newcastle, UK
95 Department of Clinical Sciences Lund, Neurology, Lund University, Lund, Sweden
96 Department of Neurology and Rehabilitation Medicine, Skåne University Hospital, Lund, Sweden
97 Bioinformatics Core Facility, University of Gothenburg, Gothenburg, Sweden
98 Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden
99 University of Technology Sydney, Faculty of Health, Ultimo, Australia
100 Department of Medicine, University of Maryland School of Medicine, MD, USA
101 Department of Neurology, Mayo Clinic, Jacksonville, FL, USA
102 Geriatrics Research and Education Clinical Center, Baltimore Veterans Administration Medical Center, Baltimore, MD, USA
103 Division of Geriatrics, School of Medicine, University of Mississippi Medical Center, Jackson, MS, USA
104 Memory Impairment and Neurodegenerative Dementia Center, University of Mississippi Medical Center, Jackson, MS, USA
105 Laboratory of Neurogenetics, National Institute on Aging, National institutes of Health, Bethesda, MD, USA
106 Data Tecnica International, Glen Echo MD, USA
107 Department of Epidemiology and Public Health, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan
108 Clinical Research Facility, Department of Medicine, NUI Galway, Galway, Ireland
109 Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA, USA
110 Department of Epidemiology, University of Washington, Seattle, WA
111 Department of Health Services, University of Washington, Seattle, WA, USA
112 Kaiser Permanente Washington Health Research Institute, Seattle, WA, USA
113 Brain Center Rudolf Magnus, Department of Neurology, University Medical Center Utrecht, Utrecht, The Netherlands
114 Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK
115 Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh, UK
116 Fred Hutchinson Cancer Research Center, University of Washington, Seattle, WA, USA
117 Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA
118 Department of Biostatistics, University of Washington, Seattle, WA, USA
119 Nuffield Department of Clinical Neurosciences, University of Oxford, UK

120 Institute for Translational Genomics and Population Sciences, Los Angeles Biomedical Research Institute at Harbor-UCLA Medical Center, Torrance, CA, USA
121 Division of Genomic Outcomes, Department of Pediatrics, Harbor-UCLA Medical Center, Torrance, CA, USA
122 Department of Neurology, Miller School of Medicine, University of Miami, Miami, FL, USA
123 Department of Allergy and Rheumatology, Graduate School of Medicine, the University of Tokyo, Tokyo, Japan
124 Center for Public Health Genomics, University of Virginia, Charlottesville, VA, USA
125 Department of Pediatrics, College of Medicine, University of Oklahoma Health Sciences Center, Oklahoma City, OK, USA
126 Department of Neurology, Medical University of Graz, Graz, Austria
127 University Medicine Greifswald, Institute for Community Medicine, SHIP-KEF, Greifswald, Germany
128 University Medicine Greifswald, Department of Neurology, Greifswald, Germany
129 Department of Neurology, Jagiellonian University, Krakow, Poland
130 Department of Neurology, Justus Liebig University, Giessen, Germany
131 Department of Clinical Neurosciences/Neurology, Institute of Neuroscience and Physiology, Sahlgrenska Academy at University of Gothenburg, Gothenburg, Sweden
132 Sahlgrenska University Hospital, Gothenburg, Sweden
133 Stroke Division, Florey Institute of Neuroscience and Mental Health, University of Melbourne, Heidelberg, Australia
134 Austin Health, Department of Neurology, Heidelberg, Australia
135 Department of Internal Medicine, Section Gerontology and Geriatrics, Leiden University Medical Center, Leiden, the Netherlands
136 INSERM U1219, Bordeaux, France
137 Department of Public Health, Bordeaux University Hospital, Bordeaux, France
138 Genetic Epidemiology Unit, Department of Epidemiology, Erasmus University Medical Center Rotterdam, Netherlands
139 Center for Medical Systems Biology, Leiden, Netherlands
140 School of Medicine, Dentistry and Nursing at the University of Glasgow, Glasgow, UK
141 Department of Epidemiology and Population Health, Albert Einstein College of Medicine, NY, USA
142 Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, MS, USA
143 A full list of members and affiliations appears in the Supplementary Note
144 Department of Human Genetics, McGill University, Montreal, Canada
145 Department of Pathophysiology, Institute of Biomedicine and Translation Medicine, University of Tartu, Tartu, Estonia
146 Department of Cardiac Surgery, Tartu University Hospital, Tartu, Estonia
147 Clinical Gene Networks AB,Stockholm, Sweden
148 Department of Genetics and Genomic Sciences, The Icahn Institute for Genomics and Multiscale Biology Icahn School of Medicine at Mount Sinai, New York, NY , USA
149 Department of Pathophysiology, Institute of Biomedicine and Translation Medicine, University of Tartu, Biomeedikum, Tartu, Estonia
150 Integrated Cardio Metabolic Centre, Department of Medicine, Karolinska Institutet, Karolinska Universitetssjukhuset, Huddinge, Sweden.
151 Clinical Gene Networks AB, Stockholm, Sweden
152 Sorbonne Universités, UPMC Univ. Paris 06, INSERM, UMR_S 1166, Team Genomics & Pathophysiology of Cardiovascular Diseases, Paris, France
153 ICAN Institute for Cardiometabolism and Nutrition, Paris, France
154 Department of Biomedical Engineering, University of Virginia, Charlottesville, VA, USA
155 Group Health Research Institute, Group Health Cooperative, Seattle, WA, USA
156 Seattle Epidemiologic Research and Information Center, VA Office of Research and Development, Seattle, WA, USA
157 Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA, USA
158 Department of Medical Research, Bærum Hospital, Vestre Viken Hospital Trust, Gjettum, Norway
159 Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore

160 National Heart and Lung Institute, Imperial College London, London, UK
161 Department of Gene Diagnostics and Therapeutics, Research Institute, National Center for Global Health and Medicine, Tokyo, Japan
162 Department of Epidemiology, Tulane University School of Public Health and Tropical Medicine, New Orleans, LA, USA
163 Department of Cardiology,University Medical Center Groningen, University of Groningen, Netherlands
164 MRC-PHE Centre for Environment and Health, School of Public Health, Department of Epidemiology and Biostatistics, Imperial College London, London, UK
165 Department of Epidemiology and Biostatistics, Imperial College London, London, UK
166 Department of Cardiology, Ealing Hospital NHS Trust, Southall, UK
167 National Heart, Lung and Blood Research Institute, Division of Intramural Research, Population Sciences Branch, Framingham, MA, USA
168 A full list of members and affiliations appears at the end of the manuscript
169 Department of Phamaceutical Sciences, Collge of Pharmacy, University of Oklahoma Health Sciences Center, Oklahoma City, OK, USA
170 Oklahoma Center for Neuroscience, Oklahoma City, OK, USA
171 Department of Pathology and Genetics, Institute of Biomedicine, The Sahlgrenska Academy at University of Gothenburg, Gothenburg, Sweden
172 Department of Neurology, Helsinki University Hospital, Helsinki, Finland
173 Clinical Neurosciences, Neurology, University of Helsinki, Helsinki, Finland
174 Department of Neurology, University of Washington, Seattle, WA, USA
175 Albrecht Kossel Institute, University Clinic of Rostock, Rostock, Germany
176 Clinical Trial Service Unit and Epidemiological Studies Unit, Nuffield Department of Population Health, University of Oxford, Oxford, UK
177 Department of Genetics, Perelman School of Medicine, University of Pennsylvania, PA, USA
178 Faculty of Medicine, University of Iceland, Reykjavik, Iceland
179 Departments of Neurology and Public Health Sciences, University of Virginia School of Medicine, Charlottesville, VA, USA
180 Department of Neurology, Boston University School of Medicine, Boston, MA, USA
181 Human Genetics Center, University of Texas Health Science Center at Houston, Houston, TX, USA
182 Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan
183 Munich Cluster for Systems Neurology (SyNergy), Munich, Germany
184 German Center for Neurodegenerative Diseases (DZNE), Munich, Germany

Short title: The MEGASTROKE study

These authors contributed equally to this work: Rainer Malik, Ganesh Chauhan, Matthew Traylor, Muralidharan Sargurupremraj and Yukinori Okada

These authors jointly supervised this work: Kari Stefansson, Bradford B Worrall, Steven J Kittner, Sudha Seshadri, Myriam Fornage, Hugh S Markus, Joanna MM Howson, Yoichiro Kamatani, Stephanie Debette and Martin Dichgans

Corresponding authors:

Martin Dichgans, MD
Institute for Stroke and Dementia Research (ISD), University Hospital, LMU Munich, Munich, Germany
Munich Cluster for Systems Neurology (SyNergy), Munich, Germany
Feodor-Lynen-Str. 17, 81377 Munich, Germany
T: +49-89-4400-46018
E: martin.dichgans@med.uni-muenchen.de
ORCID ID: orcid.org/0000-0002-0654-387X

Stephanie Debette, MD, PhD
INSERM U1219 Bordeaux Population Health Research Center, University of Bordeaux, Bordeaux, France
Department of Neurology, Institute for Neurodegenerative Disease, Bordeaux University Hospital, Bordeaux, France.

T: +33-5-57-57-16-59
E: stephanie.debette@u-bordeaux.fr
ORCID ID: orcid.org/0000-0001-8675-7968

Journal subject codes: Stroke, ischemic stroke, population genetics, genome-wide association studies, gene expression

Stroke has multiple etiologies but the underlying genes and pathways are largely unknown. We conducted a multi-ancestry genome-wide association meta-analysis in 521,612 individuals (67,162 cases and 454,450 controls) and discovered 22 novel stroke risk loci bringing the total to 32. We further found shared genetic variation with related vascular traits including blood pressure, cardiac traits, and venous thromboembolism at individual loci (N=18), and using genetic risk scores and LD score regression. Several loci exhibited distinct association and pleiotropy patterns for etiological stroke subtypes. Eleven novel loci point to mechanisms not previously implicated in stroke pathophysiology, with prioritization of risk variants and genes accomplished through bioinformatics analyses using extensive functional datasets. Stroke risk loci were significantly enriched in drug targets for antithrombotic therapy.

Stroke is the second leading cause of death and disability-adjusted life-years worldwide.[1,2] Characterized by a neurological deficit of sudden onset, stroke is mostly caused by brain infarction (ischemic stroke) and, less often, intracerebral hemorrhage (ICH). Common etiological subtypes of ischemic stroke include large artery atherosclerotic stroke (LAS), cardioembolic stroke (CES), and stroke caused by small vessel disease (small vessel stroke, SVS), the latter being also the leading cause of ICH. Previous genome-wide association studies (GWAS) in predominantly European ancestry groups have identified 10 loci robustly associated with stroke.[3-12] In most instances, the association with stroke could be attributed to individual subtypes of ischemic stroke, such as LAS[5,8,9], CES[3,4], and SVS[10,12] or of ICH[6] although some loci were associated with two or more stroke subtypes[7,9,11,13] or with any stroke.[10] We hypothesized that combining a substantially enlarged sample size with a transethnic analytic approach would identify additional risk loci and improve fine mapping of causal variants. Hence, we combined all available stroke samples with published or unpublished GWAS data including samples of non-European ancestry that were underrepresented in previous GWAS. We further hypothesized that stroke shares genetic influences with vascular risk factors, intermediate phenotypes for stroke (e.g., carotid artery plaque, cPL), and related phenotypes (e.g., coronary artery disease, CAD) and that a systematic approach to identify genetic influences shared among these traits would provide insights into stroke pathophysiology.

RESULTS

We tested ~8 million single nucleotide polymorphisms (SNPs) and InDels with minor allele frequency (MAF) $\geq$ 0.01 in up to 67,162 stroke cases and 454,450 controls for association with stroke. One analysis was of European participants only (40,585 cases; 406,111 controls) and a second involved participants of European, East-Asian (17,369; 28,195), African (5,541; 15,154), South-Asian (2,437; 6,707), mixed Asian (365; 333), and Latin-American (865; 692) ancestry (**Fig. 1).** Participants were drawn from 29 studies with genome-wide genotypes imputed to 1000 Genomes phase 1v3 or similar[14] (The MEGASTROKE consortium, **Supplementary Note**, **Supplementary Tables 1-2**). Ancestry-specific meta-analyses were conducted followed by fixed-effects transethnic meta-analyses and MANTRA transethnic meta-analyses.[15] Analyses were performed for any stroke, comprising ischemic stroke, ICH, and stroke of unknown or undetermined type (any stroke, AS, N=67,162), any ischemic stroke regardless of subtype (AIS, N=60,341) and ischemic stroke subtypes (LAS, N=6,688; CES, N=9,006; SVS, N=11,710).

Genome-wide association results

New genome-wide significant stroke loci

We identified 32 genome-wide significant loci, 22 of which were novel (**Table 1, Fig. 2**, **Supplementary Tables 3-4**, **Supplementary Fig. 1-7**). Of the 22 novel loci, 18 were identified by transethnic meta-analyses (fixed effects p-value < 5.0×10^{-8} or MANTRA $\log_{10}$(Bayes factor)[BF] > 6)(**Fig. 2** and **Supplementary Fig. 1-5**) and the remaining 4 were identified by the ancestry-specific meta-analysis in European samples (fixed effects p < 5.0×10^{-8}) (**Fig. 2** and **Supplementary Fig. 1-5**). Apart from 2 novel loci with a MAF between 0.01 and 0.05 and large effect size estimates (odds ratios [ORs] of 2.33 and 1.95), the remaining 20 novel loci harbored common variants (MAF 0.16-0.48) with observed ORs between 1.05 and 1.20 (**Table 1**). Comparison of the 32 loci across Europeans and East-Asians, the two largest ethnic subgroups, demonstrated significant correlations of risk allele frequencies and ORs between populations (**Supplementary Fig. 8**), although 6 loci exhibited population-specific association (defined as p < 5.0×10^{-8} in Europeans and p > 0.05 in East-Asians or MAF in East-Asians < 0.01)(**Supplementary Table 5**). Estimates for the phenotypic variance explained by the 32 lead variants ranged between 0.6% and 1.8% (**Supplementary Table 6**).
Gene-based tests using VEGAS2[16] (**Supplementary Fig. 9**) confirmed the loci identified by the GWAS analyses above, and yielded a novel significant (p < 2.02×10^{-6}, Bonferroni corrected for the number of genes) association for the neighbouring genes *ICA1L* and *WDR12* with SVS (**Supplementary Table 7**, **Supplementary Fig. 9-10**). Prior studies have demonstrated that variants in this region are associated with white matter hyperintensity (WMH) burden[17] a brain magnetic resonance imaging marker of small vessel disease (SVD). Twenty-one additional loci met a less stringent threshold for suggestive evidence of association ($\log_{10}$[BF] > 5.0 or p < 1.0×10^{-6} in the transethnic fixed effects analysis)(**Supplementary Table 8**), among them three loci previously implicated in Mendelian stroke (*HTRA1*[18,19], *COL4A1*[20], and *COL4A2*[21]).

Associations with etiological stroke subtypes

Eighteen loci (12 novel) reached genome-wide significance for AS, 20 (12 novel) for AIS (20), 6 (3 novel) for LAS, 4 (2 novel) for CES, and 2 (*ICA1L-WDR12* novel, discovered in gene-based tests) for SVS (**Fig. 2**, **Table 1, Supplementary Fig. 1-5 & 10**). Several loci reaching genome-wide significance for one of the ischemic stroke subtypes were also genome-wide significant for AIS or AS, while none reached genome-wide significance for multiple ischemic stroke subtypes (**Fig. 2, Supplementary Table 9**). For some novel loci, the association was strictly confined to a single subtype ($p > 0.5$ for other stroke subtypes): *EDNRA* and *LINC01492* showed association with LAS only, suggesting mechanisms limited to atherosclerosis; *NKX2-5* showed association with CES only, implying that the association may be primarily mediated by cardioembolism. We also found subtype-specificity for previously published loci (*TSPAN2* for LAS and *PITX2* for CES). We further investigated shared genetic influences of individual loci on different stroke subtypes using gwas-pw analyses[22], which estimate the posterior probability that a specified genomic region influences two different traits. Applying a posterior probability cut-off of 90% for shared contribution at a given locus (model 3) we found shared genetic influence between LAS and SVS at *SH2B3*, and between LAS and CES at *ABO* (**Supplementary Table 10 and Supplementary Fig. 11**).

Conditional analysis to identify independent signals within loci

When conditioning all SNPs in a ±0.5 Mb window on the lead SNPs in the Europeans–only analysis, we found two additional independent genome-wide signals at the *PITX2* locus for CES, consistent with known multiple independent loci at *PITX2* for atrial fibrillation (AF),[23] suggesting that a similar genetic architecture at this locus influences both conditions (**Supplementary Fig. 12**). We further found suggestive independent signals at *MMP12, SH2B3,* and *HDAC9-TWIST1* that did not reach genome-wide significance (**Supplementary Table 11**).

Genetic overlap with related vascular traits

Association of individual stroke risk variants with related vascular traits

Several of our loci are in genomic vicinity of established risk loci for vascular risk factors (e.g., blood pressure, BP), and related vascular phenotypes affecting the heart (e.g., CAD), vasculature (e.g., carotid intima media thickness, cIMT), or brain (WMH). To systematically explore genetic overlap between stroke and these traits we surveyed published GWAS for BP, blood lipids, type 2 diabetes (T2D), cIMT, cPL, AF, venous thromboembolism (VTE), CAD, and WMH, assembled through the IGEN-BP[24], ENGAGE[25], DIAGRAM[26], CHARGE[27,28], AFGen[29], INVENT[30], and CARDIoGRAMplusC4D[31] consortia (**Supplementary Table 12**). When constructing sets of index SNPs of the non-stroke phenotypes (Bonferroni adjusted $p < 1.3 \times 10^{-4} = 0.05/32$ loci/12 related vascular traits) and SNPs in high LD ($r^2 > 0.9$ in 1000G EUR) with those index variants, 17 of the 32 stroke lead variants showed overlap with these sets (**Supplementary Table 13, Fig. 3**). Fourteen loci reached genome-wide significance ($p < 5.0 \times 10^{-8}$) for association with one or more of the following phenotypes: BP (5 loci), CAD (5 loci), AF (2 loci), VTE (2 loci), LDL-cholesterol (2 loci), cPL (1 locus), and WMH (1 locus). Among the 21 additional subthreshold loci for stroke (**Supplementary Table 8**) 6 loci have previously been associated with related vascular traits including AF (*PRRX*[32], *CAV1/2*[32]), VTE (*F11*[30]), CAD (*SWAP70, LPA*[31]), blood lipids (*LPA*[31]), and WMH (*ICA1L-WDR12*[28]).

Association of genetic risk scores of related vascular traits

Second, we generated weighted genetic risk scores (wGRS) for VTE, BP-related traits, blood lipids, T2D, and CAD using the lead SNPs from published GWAS and tested these wGRS for association with each stroke phenotype, implementing the inverse-variance weighting approach (**Methods, Supplementary Table 14**). We found significant associations ($p < 5.6 \times 10^{-3}$ correcting for 9 independent phenotypes, see Methods) with wGRS for all traits

examined, except for triglyceride and LDL-cholesterol levels, with clear differences between stroke subtypes (**Fig. 4**). The strongest association was between the wGRS for CAD and LAS consistent with shared pathophysiology through atherosclerosis. We further found associations of all stroke subtypes with wGRS for BP traits. The wGRS for VTE was significantly associated with both LAS and CES (all $p < 1.0x10^{-4}$) but not SVS. The wGRS for HDL-cholesterol showed a significant inverse association with SVS.
In the present setting the wGRS analysis was used primarily to explore the genetic overlap with related vascular traits, not as a tool for establishing causal inference. In sensitivity analyses we conducted an MR-Egger regression to explore whether any of the significant associations between vascular wGRS and stroke may be partly driven by directional pleiotropy. There was no indication of directional pleiotropy except for the association between the SBP wGRS and AS (MR-Egger intercept estimate p=0.015), which was no longer significant after removing 6 of 37 SNPs appearing as outliers from the leave-one-out analysis (**Methods**), leading to causal estimates in broad agreement across regression techniques (**Supplementary Table 15**).

Shared genetic contribution to stroke and related vascular traits at the whole genome level

Third, we applied LD score regression to quantify the extent of shared genetic contributions between traits on a whole genome level.[33,34] Using available GWAS results from individuals of European ancestry, we found significant positive correlations ($r_g > 0$; $p < 5.6x10^{-3}$ correcting for 9 independent phenotypes), mostly corroborating the wGRS results (**Fig. 4** and **Supplementary Table 16**). In addition, we found significant genetic overlap between triglyceride levels and AIS with similar results obtained in available GWAS datasets from East-Asian ancestry (**Supplementary Table 16**). Results did not materially change when removing genome-wide signals for stroke and related vascular traits and their proxies ($r^2 \geq 0.8$ in 1000G EUR).

Global functional interpretation of stroke risk loci

Global epigenetic patterns at the 32 stroke risk loci

To test for cell-specific enrichment in chromatin marks that were previously shown to be phenotypically cell-type specific in ENCODE/RoadMap (H3K4me1, H3K4me3, H3K9ac)[35], we implemented the epigwas tool[35] and the narrow peak information from the latest RoadMap dataset (127 tissues).[36] Epigwas estimates the enrichment score (ratio of the height of the nearest narrow peak over the distance to the peak) for the lead variant and proxies ($r^2 \geq 0.8$ in 1000G cosmopolitan panel) and calculates statistical significance by examining the relative proximity and specificity of the test SNP-set with 10,000 sets of matched background. The analysis showed significant enrichment of enhancer and promoter sites (H3K4me1, H3K4me3) in mesenchymal stem cells, embryonic stem cells, epithelial cells, and blood & T-cells, and of active promoters (H3K9ac) in embryonic stem cells and digestive tissue (**Supplementary Table 17**).

Pathway Analyses

To identify pathways overrepresented in stroke association results we used the DEPICT gene-set enrichment tool[37] using all SNPs with $\log_{10}(BF) > 5$ for the respective stroke subtype. We found three gene-sets to be significantly (FDR < 5%) associated with AS: enlarged heart, decreased cardiac muscle contractility, and oxaloacetate metabolic process (**Supplementary Table 18**). Next, we used Ingenuity Pathway Analysis (https://www.qiagenbioinformatics.com/products/ingenuity-pathway-analysis/) examining genes within the 53 stroke locwith $\log_{10}(BF) > 5$. The extended gene list ($r^2 > 0.5$ in 1000G Europeans or East-Asians, or located within 50kB of the lead SNP) consisted of 214 genes. We found the coagulation system to be the most significant canonical pathway followed by cardiomyocyte differentiation via bone morphogenetic protein receptors (FDR 5%) (**Supplementary Table 19**). Finally, we tested enrichment of VEGAS2 derived gene-based p-values in expert curated and computationally predicted Biosystem gene-sets[38] adapting VEGAS2Pathway,[39] and identified significant association with 18 pathways including various cardiac pathways, muscle cell fate commitment, and nitric oxide metabolic process with CES (FDR 5%) (**Supplementary Table 20**).

Prioritizing potential causal variants

Fine-mapping derived from credible SNP set analyses

To reduce the number of candidate variants per locus to the most noteworthy associations we constructed 95% credible SNP sets for each of the 32 loci (lead SNP and proxy SNPs $r^2 > 0.1$ in 1000G panels) assuming one causal SNP per locus and uniform priors.[40] Credible SNP sets were generated in all stroke phenotypes and for European, East-Asian, and African ancestries separately. We found a marked reduction of credible SNP sets for most loci, expectedly most pronounced for the phenotype showing the strongest association signal (**Supplementary Table 21**). The greatest refinement was observed at *RGS7*, *HDAC9-TWIST1*, and *SH2B3*, where the lead SNP was the only SNP contained in the 95% credible set for the stroke phenotype showing the strongest association.

Stroke loci with nonsynonymous or predicted deleterious variants

To determine SNPs that have protein-altering effects, we annotated all SNPs using ANNOVAR.[41] Of the 32 lead SNPs three were exonic, of which two were non-synonymous (rs3184504 [p.Arg262Trp] in *SH2B3* and rs1052053 [p.Gln75Arg] in *PMF1*). p.Arg262Trp is a loss-of function variant that leads to expansion of hematopoetic stem cells and enhanced megakaryopoiesis in humans.[42] Both variants are predicted to be benign or tolerated by PolyPhen[43] and SIFT.[44] In addition, we identified a proxy SNP (r^2=0.99 in 1000G EUR) for another lead SNP, that was non-synonymous (rs6050 [p.Thr331Ala] in *FGA*), also predicted as benign or tolerated.

Investigation of eQTLs, meQTLs, and pQTLs in different tissues

We interrogated genome-wide gene expression (expression quantitative trait loci, eQTLs), methylation (meQTLs), and protein expression (pQTLs) in extensive publicly and non-publicly available datasets to determine whether stroke risk SNPs influenced the *cis* regulation of nearby genes. These datasets encompass numerous tissues and cell types including cardiac, vascular, and brain tissue, circulating cells, and vascular endothelial cells (**Methods**). These comprise: for eQTLs the GTEx V6[45], an expanded version of GRASP2[46,47], HGVD[48], BIOS[49], Blueprint epigenome project (subset)[50], STARNET[51] and the human aortic endothelial cells study[52]; for meQTLs, the Blueprint epigenome project (subset)[50] and the ARIC cohort[53], and for pQTLs the KORA cohort.[54] Only *cis* eQTLs, meQTLs, and pQTLs were considered.

We found that in 18 of the 32 stroke risk loci the lead stroke risk variant either overlapped or was in moderate to high LD (r^2>0.8) with the most significant QTL variant for a nearby gene, in at least one tissue or cell type (**Supplementary Table 22 and 23**). For seven loci, we observed association of the lead SNP and proxies with expression of a single gene (or methylation or protein level), sometimes the nearest gene (*LRCH1*, *CDK6*, *CDKN2B*, *PRPF8*, and *MMP12*), sometimes a more distant nearby gene (*ZCCHC14* for the *ZCCHC14* locus, and

TWIST1 for the *HDAC9-TWIST1* locus), within the datasets we explored. Associations were mostly found in stroke-relevant tissues and cell types, including vascular tissues, aortic endothelial cells, brain, blood, and immune cells. In most instances (11 loci, 61.1%), the risk SNP affected expression of multiple genes suggesting that at individual loci pleiotropic mechanisms, which might differ according to tissue/cell type, could in some instances influence stroke susceptibility.[55,56] For several of these loci there was a clear predominance of eQTL associations with one gene in stroke-relevant tissues, such as *ZNF318* (chr6p21), *AL049919* (chr12q24), and *FES* (chr15q26) in brain tissues (**Supplementary Table 22-23**). At some loci, meQTLs and eQTLs provided complementary information on the regulatory pattern. For instance, for the *SH3PXD2A* locus, SNPs in high LD with the lead stroke risk variant are eQTLs for multiple genes (*SH3PXD2A*, *SLK*, *GSTO1*, *GSTO2*, *LOC729081*), while several high LD proxies (r^2>0.96) function as the most significant meQTL for CpG probes located in the promoter region of *SH3PXD2A* and not any of the other genes.
For the 149 genes located in the 32 genome-wide significant loci (r^2 > 0.5 in Europeans or East-Asians, or being located ±50kB from the lead SNP, **Methods**), we assigned an empirical functional score based on the presence and number of eQTLs, meQTLs, pQTLs and other biological criteria[57,58] (**Methods** and **Supplementary Table 24**) reasoning that genes with a higher functional score are more likely to be causal, although this score requires validation by experimental data.

Joint modeling of epigenetic marks and association statistics

As an additional approach to identify the most plausible causal variants and genes we used RiVIERA[59], which jointly models summary association statistics and corresponding epigenetic regulatory information in a Bayesian framework to estimate the posterior probability of association (PPA). RiVIERA uses the RoadMap epigenome data of 127 tissue types and information on chromatin (H3K4me1, H3K4me3, H3K36me3, H3K27me3, H3K9me3, H3K27ac, H3K9ac), and DNA accessibility (DNaseI) marks. Three of the stroke risk loci (*PMF1-SEMA4A*, *SH3PXD2A*, and *EDNRA*) displayed a pattern in which the association statistics and epigenetic regulatory information jointly contributed to the modeling of the RiVIERA credible SNP set (the minimum number of SNPs whose PPA, accounting for both association statistics and epigenetic regulatory information, sum up to ≥95%)(**Supplementary Fig. 13**). The variants identified by RiVIERA as having the highest PPA were in moderate to high LD in the 1000G cosmopolitan panel with the respective lead SNP (rs7534434 for *PMF1- SEMA4A* [r^2=0.79 with lead SNP]; rs11191829 for *SH3PXD2A* [r^2=0.99]; rs4835084 for *EDNRA* [r^2=0.35]). Two of these (at *PMF1- SEMA4A* and *SH3PXD2A*) were significantly enriched for RNA Pol II binding in ENCODE cell types[60] including H1-hESC (human embryonic stem cells) (**Supplementary Fig. 13**).

Enrichment in drug target genes

Given previous evidence for utility of GWAS for drug discovery and drug repositioning[57,61,62] we evaluated the overlap between stroke-associated genes and known drug targets. Among the 149 genes located within the 32 stroke risk loci, 16 (11%) were registered as targets of currently approved drugs in the DrugBank database and the Therapeutic Target Database (**Supplementary Table 25**). Of these, two genes (*FGA*, *PDE3A*) were targets of approved drugs for antithrombotic therapy (ATC B01), i.e. alteplase, tenecteplase, reteplase and anistreplase for *FGA*, and cilostazol for *PDE3A* (enrichment OR=5.46, p=0.0369; **Fig. 5**). This enrichment was strengthened after removing the locus with the largest number of genes (*SH2B3*, 73 genes) (OR=8.89, p=0.0166) and after adding 65 genes in 21 suggestive stroke risk loci (OR=7.83, p=0.00606).

DISCUSSION

The current transethnic meta-analysis more than triples the number of stroke risk loci and identifies novel loci for AS, AIS, and all major subtypes of ischemic stroke. Our results highlight several major features of stroke genomics: (i) approximately half of the identified stroke loci show shared genetic association with other vascular traits, the largest genetic correlation being found for BP. We also identified shared genetic association with VTE, with distinct patterns for individual stroke subtypes providing mechanistic insight; (ii) eleven of the novel stroke loci (*ANK2*, *CDK6*, *KCNK3*, *LINC01492*, *LRCH1*, *NKX2-5*, *PDE3A*, *PRPF8*, *RGS7*, *TM4SF4-TM4SF1* and *WNT2B*) point to mechanisms not previously implicated in stroke pathophysiology; some of these suggest a strong link with cardiac mechanisms beyond those expected from established sources of cardioembolism; (iii) the 32 stroke risk loci were significantly enriched in drug targets for antithrombotic therapy, one for an approved thrombolytic drug (alteplase) and the other for an antiplatelet agent (cilostazol) approved for stroke prevention in Asia; (iv) through incorporation of extensive functional datasets and bioinformatics analyses we provide detailed information on prioritization of stroke risk variants and genes as a resource for further experimental follow-up.

The majority of genome-wide associations were identified with both AS and AIS. While this relates in part to a higher power compared to subtypes, we also found shared genetic influences between stroke subtypes, as exemplified by the gwas-pw analyses (*SH2B3* and *ABO*). A notable finding is the identification of *PMF1-SEMA4A* as a risk locus for AIS. *PMF1-SEMA4A* is an established risk locus for non-lobar ICH[6] and thus represents the first locus reaching genome-wide significance for ischemic as well as hemorrhagic stroke. *PMF1-SEMA4A* further reached genome-wide association for WMH burden[28] (**Fig. 3**), an established marker for SVD, and showed a strong signal in the SVS subtype suggesting that the association with stroke is at least in part mediated by SVD. The underlying biological pathways do not seem to involve known vascular risk factors and may thus reveal novel targets for stroke prevention.

Among the novel loci showing associations restricted to specific stroke subtypes, *EDNRA* is consistent with atherosclerotic mechanisms given its association with LAS, cPL[27] and CAD[31] (**Fig. 3**). *LINC01492* and the previously reported *TSPAN2* locus likewise displayed associations restricted to LAS but showed no association with related phenotypes in our look-ups and in prior literature, thus evidencing mechanisms more specific for LAS. *NKX2-5*, showing association restricted to CES, was previously reported as a genome-wide risk locus for heart rate and PR interval[63,64] but not consistently for AF[63,65] thus pointing towards cardiac mechanisms other than AF.

Although the number of loci reaching genome-wide significance for association with SVS remains low, our results suggest an important role for common genetic variation in SVS. First, several of the associations with AS or AIS including at novel loci (*CASZ1*, *LOC100505841*, *SH3PXD2A*, *ICA1L-WDR12*) show predominant association with the SVS subtype (**Supplementary Table 7** and **Supplementary Table 9**). Second, three of the top loci (*PMF1-SEMA4A, LOC100505841, SH3PXD2A)* show genetic overlap with loci for WMH. Third, several suggestive loci ($\log_{10}$[BF] $\geq$5) for AS and SVS harbor genes implicated in monogenic SVD (*HTRA1*, *COL4A1*, *COL4A2*) (**Supplementary Table 8**).

Our extensive exploration of shared genetic variation between stroke and related vascular traits found the most widespread correlations with BP phenotypes consistent with epidemiological data showing high BP to be the leading risk factor for stroke. A quarter of the 32 genome-wide significant stroke loci are BP loci, most of which are novel with respect to stroke risk and show association with risk of AS or AIS. Aside from expected genetic overlap between LAS and CAD, we also identified significant overlap between a wGRS for VTE and both LAS, and CES, but not SVS (**Supplementary Table 14, Fig. 4**) despite a higher power for this subtype, potentially suggesting that thrombotic processes play a less important role in SVS.

Three of our novel loci (*NKX2-5*, *ANK2*, and *LRCH1*) have previously been associated with cardiac pacing.[63,64,66] *NKX2-5* and *ANK2* are further implicated in familial forms of cardiac disease[67-70] but none of the three loci was associated with AF or CAD in the latest published GWAS.[31,65] Apart from *NKX2-5* they were not specifically associated with CES, possibly pointing to an involvement of the underlying genes beyond cardiac development and function. rs9526212, the lead variant in *LRCH1* functions as an eQTL for *LRCH1* in multiple tissues including left ventricle, atherosclerotic aorta, atherosclerotic-lesion free arteries, and blood (**Supplementary Table 22**). Pathway analyses further support a strong link with cardiac mechanisms.

The extensive in silico functional annotation of identified stroke risk loci provides informative elements for future prioritization and follow-up of the most compelling biological candidates. In some instances, the eQTL, meQTL and pQTL information strongly supports involvement of one gene over others in the region, e.g., for *SH3PXD2A*, encoding SH3 and PX domain-containing protein 2A, an adapter protein involved in invadopodia and podosome formation as well as extracellular matrix degradation. For some loci, joint analysis of epigenetic regulatory effects and association statistics enabled prioritization of credible SNPs. When exploring overall epigenetic patterns of identified stroke risk loci, some enrichment of enhancer and promoter sites in developmental tissues was observed, suggesting that some associations may be driven by developmental effects, as recently proposed for the *FOXF2* locus.[10]

RGS7 and *TM4SF4-TM4SF1* showed low minor allele frequencies, high heterogeneity, poor imputation quality in non-Europeans, and large effect size estimates and must therefore be interpreted with caution. Moreover, while our extensive functional exploration provides guidance on gene prioritization for further exploration, additional experiments are required to identify the causal genes and variants. Several studies had limited information on stroke subtypes. Hence sample sizes for ischemic stroke subtypes were still in the lower range. Also, the proportion of the phenotypic variance explained by the 32 lead SNPs was relatively small but comparable to other complex diseases.[71] Collectively, these aspects highlight the potential for gene discovery in the future.

In conclusion, we identify 22 novel stroke risk loci and demonstrate shared genetic variation with multiple related vascular traits. We further identify novel loci offering mechanisms not previously implicated in stroke pathophysiology and provide a framework for prioritization of stroke risk variants and genes for further functional and experimental follow-up. Stroke risk loci are significantly enriched in drug targets for antithrombotic therapy thus highlighting the potential of stroke genetics for drug discovery. Collectively, these findings represent a major advance in understanding the genetic underpinnings of stroke.

MEGASTROKE CONSORTIUM

Rainer Malik [1], Ganesh Chauhan [2], Matthew Traylor [3], Muralidharan Sargurupremraj [4,5], Yukinori Okada [6,7,8], Aniket Mishra [4,5], Loes Rutten-Jacobs [3], Anne-Katrin Giese [9], Sander W van der Laan [10], Solveig Gretarsdottir [11], Christopher D Anderson [12,13,14,14], Michael Chong [15], Hieab HH Adams [16,17], Tetsuro Ago [18], Peter Almgren [19], Philippe Amouyel [20,21], Hakan Ay [22,13], Traci M Bartz [23], Oscar R Benavente [24], Steve Bevan [25], Giorgio B Boncoraglio [26], Robert D Brown, Jr. [27], Adam S Butterworth [28,29], Caty Carrera [30,31], Cara L Carty [32,33], Daniel I Chasman [34,35], Wei-Min Chen [36], John W Cole [37], Adolfo Correa [38], Ioana Cotlarciuc [39], Carlos Cruchaga [40,41], John Danesh [28,42,43,44], Paul IW de Bakker [45,46], Anita L DeStefano [47,48], Marcel den Hoed [49], Qing Duan [50], Stefan T Engelter [51,52], Guido J Falcone [53,54], Rebecca F Gottesman [55], Raji P Grewal [56], Vilmundur Gudnason [57,58], Stefan Gustafsson [59], Jeffrey Haessler [60], Tamara B Harris [61], Ahamad Hassan [62], Aki S Havulinna [63,64], Susan R Heckbert [65], Elizabeth G Holliday [66,67], George Howard [68], Fang-Chi Hsu [69], Hyacinth I Hyacinth [70], M Arfan Ikram [16], Erik Ingelsson [71,72], Marguerite R Irvin [73], Xueqiu Jian [74], Jordi Jiménez-Conde [75], Julie A Johnson [76,77], J Wouter Jukema [78], Masahiro Kanai [6,7,79], Keith L Keene [80,81], Brett M Kissela [82], Dawn O Kleindorfer [82], Charles Kooperberg [60], Michiaki Kubo [83], Leslie A Lange [84], Carl D Langefeld [85], Claudia Langenberg [86], Lenore J Launer [87], Jin-Moo Lee [88], Robin Lemmens [89,90], Didier Leys [91], Cathryn M Lewis [92,93], Wei-Yu Lin [28,94], Arne G Lindgren [95,96], Erik Lorentzen [97], Patrik K Magnusson [98], Jane Maguire [99], Ani Manichaikul [36], Patrick F McArdle [100], James F Meschia [101], Braxton D Mitchell [100,102], Thomas H Mosley [103,104], Michael A Nalls [105,106], Toshiharu Ninomiya [107], Martin J O'Donnell [15,108], Bruce M Psaty [109,110,111,112], Sara L Pulit [113,45], Kristiina Rannikmäe [114,115], Alexander P Reiner [65,116], Kathryn M Rexrode [117], Kenneth Rice [118], Stephen S Rich [36], Paul M Ridker [34,35], Natalia S Rost [9,13], Peter M Rothwell [119], Jerome I Rotter [120,121], Tatjana Rundek [122], Ralph L Sacco [122], Saori Sakaue [7,123], Michele M Sale [124], Veikko Salomaa [63], Bishwa R Sapkota [125], Reinhold Schmidt [126], Carsten O Schmidt [127], Ulf Schminke [128], Pankaj Sharma [39], Agnieszka Slowik [129], Cathie LM Sudlow [114,115], Christian Tanislav [130], Turgut Tatlisumak [131,132], Kent D Taylor [120,121], Vincent NS Thijs [133,134], Gudmar Thorleifsson [11], Unnur Thorsteinsdottir [11], Steffen Tiedt [1], Stella Trompet [135], Christophe Tzourio [5,136,137], Cornelia M van Duijn [138,139], Matthew Walters [140], Nicholas J Wareham [86], Sylvia Wassertheil-Smoller [141], James G Wilson [142], Kerri L Wiggins [109], Qiong Yang [47], Salim Yusuf [15], Najaf Amin [16], Hugo S Aparicio [185,48], Donna K Arnett [186], John Attia [187], Alexa S Beiser [47,48], Claudine Berr [188], Julie E Buring [34,35], Mariana Bustamante [189], Valeria Caso [190], Yu-Ching Cheng [191], Seung Hoan Choi [192,48], Ayesha Chowhan [185,48], Natalia Cullell [31], Jean-François Dartigues [193,194], Hossein Delavaran [95,96], Pilar Delgado [195], Marcus Dörr [196,197], Gunnar Engström [19], Ian Ford [198], Wander S Gurpreet [199], Anders Hamsten [200,201], Laura Heitsch [202], Atsushi Hozawa [203], Laura Ibanez [204], Andreea Ilinca [95,96], Martin Ingelsson [205], Motoki Iwasaki [206], Rebecca D Jackson [207], Katarina Jood [208], Pekka Jousilahti [63], Sara Kaffashian [4,5], Lalit Kalra [209], Masahiro Kamouchi [210], Takanari Kitazono [211], Olafur Kjartansson [212], Manja Kloss [213], Peter J Koudstaal [214], Jerzy Krupinski [215], Daniel L Labovitz [216], Cathy C Laurie [118], Christopher R Levi [217], Linxin Li [218], Lars Lind [219], Cecilia M Lindgren [220,221], Vasileios Lioutas [222,48], Yong Mei Liu [223], Oscar L Lopez [224], Hirata Makoto [225], Nicolas Martinez-Majander [172], Koichi Matsuda [225], Naoko Minegishi [203], Joan Montaner [226], Andrew P Morris [227,228], Elena Muiño [31], Martina Müller-Nurasyid [229,230,231], Bo Norrving [95,96], Soichi Ogishima [203], Eugenio A Parati [232], Leema Reddy Peddareddygari [56], Nancy L Pedersen [98,233], Joanna Pera [129], Markus Perola [63,234], Alessandro Pezzini [235], Silvana Pileggi [236], Raquel Rabionet [237], Iolanda Riba-Llena [30], Marta Ribasés [238], Jose R Romero [185,48], Jaume Roquer [239,240], Anthony G Rudd [241,242], Antti-Pekka Sarin [243,244], Ralhan Sarju [199], Chloe Sarnowski [47,48], Makoto Sasaki [245], Claudia L Satizabal [185,48], Mamoru Satoh [245], Naveed Sattar [246], Norie Sawada [206], Gerli Sibolt [172], Ásgeir Sigurdsson [247], Albert Smith [248], Kenji Sobue [245], Carolina Soriano-Tárraga [240], Tara Stanne [249], O Colin Stine [250], David J Stott [251], Konstantin Strauch [229,252], Takako Takai [203], Hideo Tanaka [253,254], Kozo Tanno [245], Alexander Teumer [255], Liisa Tomppo [172], Nuria P Torres-Aguila [31], Emmanuel Touze [256,257], Shoichiro Tsugane [206], Andre G Uitterlinden [258], Einar M Valdimarsson [259], Sven J van der Lee [16], Henry Völzke [255], Kenji Wakai [253], David Weir [260], Stephen R Williams [261], Charles DA Wolfe [241,242], Quenna Wong [118], Huichun Xu [191], Taiki Yamaji [206], Dharambir K Sanghera [125,169,170], Olle Melander [19], Christina Jern [171], Daniel Strbian [172,173], Israel Fernandez-Cadenas [31,30], W T Longstreth, Jr [174,65], Arndt Rolfs [175], Jun Hata [107], Daniel Woo [82], Jonathan Rosand [12,13,14], Guillaume Pare [15], Jemma C Hopewell [176], Danish Saleheen [177], Kari Stefansson [11,178], Bradford B Worrall [179], Steven J Kittner [37], Sudha Seshadri [180,48], Myriam Fornage [74,181], Hugh S Markus [3], Joanna MM Howson [28], Yoichiro Kamatani [6,182], Stephanie Debette [4,5], Martin Dichgans [1,183,184]

185 Boston University School of Medicine, Boston, MA, USA
186 University of Kentucky College of Public Health, Lexington, KY, USA
187 University of Newcastle and Hunter Medical Research Institute, New Lambton, Australia
188 Univ. Montpellier, Inserm, U1061, Montpellier, France
189 Centre for Research in Environmental Epidemiology, Barcelona, Spain
190 Department of Neurology, Università degli Studi di Perugia, Umbria, Italy
191 Department of Medicine, University of Maryland School of Medicine, Baltimore, MD, USA
192 Broad Institute, Cambridge, MA, USA
193 Univ. Bordeaux, Inserm, Bordeaux Population Health Research Center, UMR 1219, Bordeaux, France
194 Bordeaux University Hospital, Department of Neurology, Memory Clinic, Bordeaux, France
195 Neurovascular Research Laboratory. Vall d'Hebron Institut of Research, Neurology and Medicine Departments-Universitat Autònoma de Barcelona. Vall d'Hebrón Hospital, Barcelona, Spain
196 University Medicine Greifswald, Department of Internal Medicine B, Greifswald, Germany
197 DZHK, Greifswald, Germany
198 Robertson Center for Biostatistics, University of Glasgow, Glasgow, UK
199 Hero DMC Heart Institute, Dayanand Medical College & Hospital, Ludhiana, India
200 Atherosclerosis Research Unit, Department of Medicine Solna, Karolinska Institutet, Stockholm, Sweden
201 Karolinska Institutet, Stockholm, Sweden
202 Division of Emergency Medicine, and Department of Neurology, Washington University School of Medicine, St. Louis, MO, USA
203 Tohoku Medical Megabank Organization, Sendai, Japan
204 Department of Psychiatry, Washington University School of Medicine, St. Louis, MO, USA
205 Department of Public Health and Caring Sciences / Geriatrics, Uppsala University, Uppsala, Sweden
206 Epidemiology and Prevention Group, Center for Public Health Sciences, National Cancer Center, Tokyo, Japan
207 Department of Internal Medicine and the Center for Clinical and Translational Science, The Ohio State University, Columbus, OH, USA
208 Institute of Neuroscience and Physiology, the Sahlgrenska Academy at University of Gothenburg, Goteborg, Sweden
209 Department of Basic and Clinical Neurosciences, King's College London, London, UK
210 Department of Health Care Administration and Management, Graduate School of Medical Sciences, Kyushu University, Japan
211 Department of Medicine and Clinical Science, Graduate School of Medical Sciences, Kyushu University, Japan
212 Landspitali National University Hospital, Departments of Neurology & Radiology, Reykjavik, Iceland
213 Department of Neurology, Heidelberg University Hospital, Germany
214 Department of Neurology, Erasmus University Medical Center
215 Hospital Universitari Mutua Terrassa, Terrassa (Barcelona), Spain
216 Albert Einstein College of Medicine, Montefiore Medical Center, New York, NY, USA
217 John Hunter Hospital, Hunter Medical Research Institute and University of Newcastle, Newcastle, NSW, Australia
218 Centre for Prevention of Stroke and Dementia, Nuffield Department of Clinical Neurosciences, University of Oxford, UK
219 Department of Medical Sciences, Uppsala University, Uppsala, Sweden
220 Genetic and Genomic Epidemiology Unit, Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK
221 The Wellcome Trust Centre for Human Genetics, Oxford, UK
222 Beth Israel Deaconess Medical Center, Boston, MA, USA
223 Wake Forest School of Medicine, Wake Forest, NC, USA
224 Department of Neurology, University of Pittsburgh, Pittsburgh, PA, USA

225 BioBank Japan, Laboratory of Clinical Sequencing, Department of Computational biology and medical Sciences, Graduate school of Frontier Sciences, The University of Tokyo, Tokyo, Japan
226 Neurovascular Research Laboratory, Vall d'Hebron Institut of Research, Neurology and Medicine Departments-Universitat Autònoma de Barcelona. Vall d'Hebrón Hospital, Barcelona, Spain
227 Department of Biostatistics, University of Liverpool, Liverpool, UK
228 Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK
229 Institute of Genetic Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany
230 Department of Medicine I, Ludwig-Maximilians-Universität, Munich, Germany
231 DZHK (German Centre for Cardiovascular Research), partner site Munich Heart Alliance, Munich, Germany
232 Department of Cerebrovascular Diseases, Fondazione IRCCS Istituto Neurologico "Carlo Besta", Milano, Italy
233 Karolinska Institutet, MEB, Stockholm, Sweden
234 University of Tartu, Estonian Genome Center, Tartu, Estonia, Tartu, Estonia
235 Department of Clinical and Experimental Sciences, Neurology Clinic, University of Brescia, Italy
236 Translational Genomics Unit, Department of Oncology, IRCCS Istituto di Ricerche Farmacologiche Mario Negri, Milano, Italy
237 Department of Genetics, Microbiology and Statistics, University of Barcelona, Barcelona, Spain
238 Psychiatric Genetics Unit, Group of Psychiatry, Mental Health and Addictions, Vall d'Hebron Research Institute (VHIR), Universitat Autònoma de Barcelona, Biomedical Network Research Centre on Mental Health (CIBERSAM), Barcelona, Spain
239 Department of Neurology, IMIM-Hospital del Mar, and Universitat Autònoma de Barcelona, Spain
240 IMIM (Hospital del Mar Medical Research Institute), Barcelona, Spain
241 National Institute for Health Research Comprehensive Biomedical Research Centre, Guy's & St. Thomas' NHS Foundation Trust and King's College London, London, UK
242 Division of Health and Social Care Research, King's College London, London, UK
243 FIMM-Institute for Molecular Medicine Finland, Helsinki, Finland
244 THL-National Institute for Health and Welfare, Helsinki, Finland
245 Iwate Tohoku Medical Megabank Organization, Iwate Medical University, Iwate, Japan
246 BHF Glasgow Cardiovascular Research Centre, Faculty of Medicine, Glasgow, UK
247 deCODE Genetics/Amgen, Inc., Reykjavik, Iceland
248 Icelandic Heart Association, Reykjavik, Iceland
249 Institute of Biomedicine, the Sahlgrenska Academy at University of Gothenburg, Goteborg, Sweden
250 Department of Epidemiology, University of Maryland School of Medicine, Baltimore, MD, USA
251 Institute of Cardiovascular and Medical Sciences, Faculty of Medicine, University of Glasgow, Glasgow, UK
252 Chair of Genetic Epidemiology, IBE, Faculty of Medicine, LMU Munich, Germany
253 Division of Epidemiology and Prevention, Aichi Cancer Center Research Institute, Nagoya, Japan
254 Department of Epidemiology, Nagoya University Graduate School of Medicine, Nagoya, Japan
255 University Medicine Greifswald, Institute for Community Medicine, SHIP-KEF, Greifswald, Germany
256 Department of Neurology, Caen University Hospital, Caen, France
257 University of Caen Normandy, Caen, France
258 Department of Internal Medicine, Erasmus University Medical Center, Rotterdam, Netherlands
259 Landspitali University Hospital, Reykjavik, Iceland
260 Survey Research Center, University of Michigan, Ann Arbor, MI, USA
261 University of Virginia Department of Neurology, Charlottesville, VA, USA

Acknowledgments

A full list of acknowledgments appears in the **Supplementary Note.**

Author Contributions

R.M., G.C., M.T., M.S., Y.O., S.D. and M.D. wrote and edited the manuscript. Study design/conception: R.M., M.D., S.D., B.M.P., G.J.F, J.W.J., J.I.R., J.G.W., M.F., H.I.Y., C.J., S.S., W.T.L. O.M..; Statistical analysis: G.J.F.,M.F.,Y.O.,R.M.,M.S.,M.T.,A.M.,E.G.H.,C.D.A.,T.M.B.,C.C.,I.C.,W.Y.L.,S.L.P.,K.Ra.,K.R.,S.T.,J.C.,F.T.;Sample/phenotype contribution: M.D., S.D., C.D.A., C.C., I.C., H.I.H., J.W.J., N.S.R., B.M.P, J.I.R., J.G.W., O.M., C.J., J.C.H., S.S., T.A., G.B.B., R.D.B., A.H., N.L.S., R.L., C.M.L., T.N., P.M.R., V.S., C.O.S., P.S., C.L.M.S., K.D.T., S.T., M.C., D.S., I.F., J.H.; Critical revision of article: R.M., M.D., S.D., B.M.P., C.J., J.I.R., O.M., S.S., G.J.F., J.W.J., W.T.L., C.D.A., D.S., E.G.H., I.F., S.T., C.L.M.S., C.O.S., C.C., G.B.B., I.C., J.C.B., J.H., K.R., S.L.P., N.S.R., S.S.R., T.A., T.N., J.M.M.H, T.M.B., V.S.; Supervision: M.D., S.D., C.D.A., J.M.M.H , J.I.R., S.S., C.M.L., C.L.M.S., J.W.J., V.S., J.C.B; GWAS analyses: R.M., G.C., M.T., S.G., G.T., J.H., A.K.G., M.C., J.B., C.C., A.H., G.J.F., Y.K. Functional annotation: M.S., A.M., R.M., G.C., M.T., L.R.J., A.K.G.; Gene-based analysis: A.M.; Pathway analyses: A.M., R.M., M.C., K.R.; Drug Target Analysis: Y.O.; Scoring method: M.S., R.M., S.D., M.D.; wGRS analysis: M.S., R.M.; LD score regression analysis: R.M., M.S., Y.K.; credible SNP set analysis: R.M., G.C., M.S.

Competing financial interests

S.G., G.T., U.Th. and K.S. are all employees of deCODE Genetics/Amgen, Inc.; M.A.N. is an employee of Data Tecnica International; P.T.E is the PI on a grant from Bayer HealthCare to the Broad Institute focused on the genetics and therapeutics of atrial fibrillation; S.A.L receives sponsored research support from Bayer HealthCare, Biotronik, and Boehringer Ingelheim, and has consulted for St. Jude Medical and Quest Diagnostics; E.I. is a scientific advisor for Precision Wellness, Cellink and Olink Proteomics for work unrelated to the present project; B.M.P. serves on the DSMB of a clinical trial funded by Zoll LifeCor and on the Steering Committee of the Yale Open Data Access Project funded by Johnson & Johnson. The remaining authors have no disclosures.

Disclaimer: The views expressed in this manuscript are those of the authors and do not necessarily represent the views of the National Heart, Lung, and Blood Institute; or the National Institutes of Health.

MAIN TEXT REFERENCES

1. GBD 2015 DALYs and HALE Collaborators. Global, regional, and national disability-adjusted life-years (DALYs) for 315 diseases and injuries and healthy life expectancy (HALE), 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet* **388**, 1603-1658 (2016).
2. GBD 2015 Mortality and Causes of Death Collaborators. Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980-2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet* **388**, 1459-1544 (2016).
3. Gudbjartsson, D.F. *et al.* Variants conferring risk of atrial fibrillation on chromosome 4q25. *Nature* **448**, 353-7 (2007).
4. Gudbjartsson, D.F. *et al.* A sequence variant in ZFHX3 on 16q22 associates with atrial fibrillation and ischemic stroke. *Nat Genet* **41**, 876-8 (2009).
5. International Stroke Genetics Consortium (ISGC) *et al.* Genome-wide association study identifies a variant in HDAC9 associated with large vessel ischemic stroke. *Nat Genet* **44**, 328-33 (2012).
6. Woo, D. *et al.* Meta-analysis of genome-wide association studies identifies 1q22 as a susceptibility locus for intracerebral hemorrhage. *Am J Hum Genet* **94**, 511-21 (2014).
7. Kilarski, L.L. *et al.* Meta-analysis in more than 17,900 cases of ischemic stroke reveals a novel association at 12q24.12. *Neurology* **83**, 678-85 (2014).
8. Traylor, M. *et al.* A novel MMP12 locus is associated with large artery atherosclerotic stroke using a genome-wide age-at-onset informed approach. *PLoS Genet* **10**, e1004469 (2014).
9. NINDS Stroke Genetics Network (SiGN) & International Stroke Genetics Consortium (ISGC). Loci associated with ischaemic stroke and its subtypes (SiGN): a genome-wide association study. *Lancet Neurol* **15**, 174-84 (2015).
10. Neurology Working Group of the Cohorts for Heart Aging Research in Genomic Epidemiology (CHARGE) Consortium, Stroke Genetics Network (SiGN) & International Stroke Genetics Consortium (ISGC). Identification of additional risk loci for stroke and small vessel disease: a meta-analysis of genome-wide association studies. *Lancet Neurol* **15**, 695-707 (2016).
11. Malik, R. *et al.* Low-frequency and common genetic variation in ischemic stroke: The METASTROKE collaboration. *Neurology* **86**, 1217-26 (2016).
12. Traylor, M. *et al.* Genetic Variation at 16q24.2 is associated with small vessel stroke. *Ann Neurol* **81**, 383-94 (2016).
13. Williams, F.M. *et al.* Ischemic stroke is associated with the ABO locus: the EuroCLOT study. *Ann Neurol* **73**, 16-31 (2013).
14. 1000 Genomes Project Consortium *et al.* A global reference for human genetic variation. *Nature* **526**, 68-74 (2015).
15. Morris, A.P. Transethnic meta-analysis of genomewide association studies. *Genet Epidemiol* **35**, 809-22 (2011).
16. Mishra, A. & Macgregor, S. VEGAS2: Software for More Flexible Gene-Based Testing. *Twin Res Hum Genet* **18**, 86-91 (2015).
17. Traylor, M. *et al.* Genome-wide meta-analysis of cerebral white matter hyperintensities in patients with stroke. *Neurology* **86**, 146-53 (2016).
18. Hara, K. *et al.* Association of HTRA1 mutations and familial ischemic cerebral small-vessel disease. *N Engl J Med* **360**, 1729-39 (2009).
19. Verdura, E. *et al.* Heterozygous HTRA1 mutations are associated with autosomal dominant cerebral small vessel disease. *Brain* **138**, 2347-58 (2015).

20. Gould, D.B. *et al.* Role of COL4A1 in small-vessel disease and hemorrhagic stroke. *N Engl J Med* **354**, 1489-96 (2006).
21. Jeanne, M. *et al.* COL4A2 mutations impair COL4A1 and COL4A2 secretion and cause hemorrhagic stroke. *Am J Hum Genet* **90**, 91-101 (2012).
22. Pickrell, J.K. *et al.* Detection and interpretation of shared genetic influences on 42 human traits. *Nat Genet* **48**, 709-17 (2016).
23. Lubitz, S.A. *et al.* Independent susceptibility markers for atrial fibrillation on chromosome 4q25. *Circulation* **122**, 976-84 (2010).
24. Kato, N. *et al.* Trans-ancestry genome-wide association study identifies 12 genetic loci influencing blood pressure and implicates a role for DNA methylation. *Nat Genet* **47**, 1282-93 (2015).
25. Surakka, I. *et al.* The impact of low-frequency and rare variants on lipid levels. *Nat Genet* **47**, 589-97 (2015).
26. Morris, A.P. *et al.* Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat Genet* **44**, 981-90 (2012).
27. Bis, J.C. *et al.* Meta-analysis of genome-wide association studies from the CHARGE consortium identifies common variants associated with carotid intima media thickness and plaque. *Nat Genet* **43**, 940-7 (2011).
28. Verhaaren, B.F. *et al.* Multiethnic genome-wide association study of cerebral white matter hyperintensities on MRI. *Circ Cardiovasc Genet* **8**, 398-409 (2015).
29. Sinner, M.F. *et al.* Integrating genetic, transcriptional, and functional analyses to identify 5 novel genes for atrial fibrillation. *Circulation* **130**, 1225-35 (2014).
30. Germain, M. *et al.* Meta-analysis of 65,734 individuals identifies TSPAN15 and SLC44A2 as two susceptibility loci for venous thromboembolism. *Am J Hum Genet* **96**, 532-42 (2015).
31. Nikpay, M. *et al.* A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat Genet* **47**, 1121-30 (2015).
32. Ellinor, P.T. *et al.* Meta-analysis identifies six new susceptibility loci for atrial fibrillation. *Nat Genet* **44**, 670-5 (2012).
33. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat Genet* **47**, 1236-41 (2015).
34. Bulik-Sullivan, B.K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* **47**, 291-5 (2015).
35. Trynka, G. *et al.* Chromatin marks identify critical cell types for fine mapping complex trait variants. *Nat Genet* **45**, 124-30 (2013).
36. Roadmap Epigenomics Consortium *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317-30 (2015).
37. Pers, T.H. *et al.* Biological interpretation of genome-wide association studies using predicted gene functions. *Nat Commun* **6**, 5890 (2015).
38. Geer, L.Y. *et al.* The NCBI BioSystems database. *Nucleic Acids Res* **38**, D492-6 (2010).
39. Mishra, A. & MacGregor, S. A Novel Approach for Pathway Analysis of GWAS Data Highlights Role of BMP Signaling and Muscle Cell Differentiation in Colorectal Cancer Susceptibility. *Twin Res Hum Genet* **20**, 1-9 (2017).
40. Wakefield, J. A Bayesian measure of the probability of false discovery in genetic epidemiology studies. *Am J Hum Genet* **81**, 208-27 (2007).
41. Yang, H. & Wang, K. Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. *Nat Protoc* **10**, 1556-66 (2015).
42. Wang, W. *et al.* LNK/SH2B3 Loss of Function Promotes Atherosclerosis and Thrombosis. *Circ Res* **119**, e91-e103 (2016).

43. Ramensky, V., Bork, P. & Sunyaev, S. Human non-synonymous SNPs: server and survey. *Nucleic Acids Res* **30**, 3894-900 (2002).
44. Kumar, P., Henikoff, S. & Ng, P.C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* **4**, 1073-81 (2009).
45. GTEx Consortium. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648-60 (2015).
46. Eicher, J.D. *et al.* GRASP v2.0: an update on the Genome-Wide Repository of Associations between SNPs and phenotypes. *Nucleic Acids Res* **43**, D799-804 (2015).
47. Leslie, R., O'Donnell, C.J. & Johnson, A.D. GRASP: analysis of genotype-phenotype results from 1390 genome-wide association studies and corresponding open access database. *Bioinformatics* **30**, i185-94 (2014).
48. Higasa, K. *et al.* Human genetic variation database, a reference database of genetic variations in the Japanese population. *J Hum Genet* **61**, 547-53 (2016).
49. Bonder, M.J. *et al.* Disease variants alter transcription factor levels and methylation of their binding sites. *Nat Genet* **49**, 131-138 (2017).
50. Adams, D. *et al.* BLUEPRINT to decode the epigenetic signature written in blood. *Nat Biotechnol* **30**, 224-6 (2012).
51. Franzen, O. *et al.* Cardiometabolic risk loci share downstream cis- and trans-gene regulation across tissues and diseases. *Science* **353**, 827-30 (2016).
52. Erbilgin, A. *et al.* Identification of CAD candidate genes in GWAS loci and their expression in vascular cells. *J Lipid Res* **54**, 1894-905 (2013).
53. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. *Am J Epidemiol* **129**, 687-702 (1989).
54. Suhre, K. *et al.* Connecting genetic risk to disease end points through the human blood plasma proteome. *Nat Commun* **8**, 14357 (2017).
55. Braenne, I. *et al.* Prediction of Causal Candidate Genes in Coronary Artery Disease Loci. *Arterioscler Thromb Vasc Biol* **35**, 2207-17 (2015).
56. Flister, M.J. *et al.* Identifying multiple causative genes at a single GWAS locus. *Genome Res* **23**, 1996-2002 (2013).
57. Okada, Y. *et al.* Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* **506**, 376-81 (2014).
58. Kemp, J.P. *et al.* Identification of 153 new loci associated with heel bone mineral density and functional involvement of GPC6 in osteoporosis. *Nat Genet* (2017).
59. Li, Y. & Kellis, M. Joint Bayesian inference of risk variants and tissue-specific epigenomic enrichments across multiple complex human diseases. *Nucleic Acids Res* **44**, e144 (2016).
60. Lee, B.K. *et al.* Cell-type specific and combinatorial usage of diverse transcription factors revealed by genome-wide binding studies in multiple human cells. *Genome Res* **22**, 9-24 (2012).
61. Sanseau, P. *et al.* Use of genome-wide association studies for drug repositioning. *Nat Biotechnol* **30**, 317-20 (2012).
62. Nelson, M.R. *et al.* The support of human genetic evidence for approved drug indications. *Nat Genet* **47**, 856-60 (2015).
63. den Hoed, M. *et al.* Identification of heart rate-associated loci and their effects on cardiac conduction and rhythm disorders. *Nat Genet* **45**, 621-31 (2013).
64. Pfeufer, A. *et al.* Genome-wide association study of PR interval. *Nat Genet* **42**, 153-9 (2010).
65. Christophersen, I.E. *et al.* Large-scale analyses of common and rare variants identify 12 new loci associated with atrial fibrillation. *Nat Genet* **49**, 946-952 (2017).
66. Verweij, N. *et al.* Genetic determinants of P wave duration and PR segment. *Circ Cardiovasc Genet* **7**, 475-81 (2014).

67. Le Scouarnec, S. *et al.* Dysfunction in ankyrin-B-dependent ion channel and transporter targeting causes human sinus node disease. *Proc Natl Acad Sci U S A* **105**, 15617-22 (2008).
68. Schott, J.J. *et al.* Congenital heart disease caused by mutations in the transcription factor NKX2-5. *Science* **281**, 108-11 (1998).
69. Ellesoe, S.G. *et al.* Familial Atrial Septal Defect and Sudden Cardiac Death: Identification of a Novel NKX2-5 Mutation and a Review of the Literature. *Congenit Heart Dis* **11**, 283-90 (2016).
70. Mohler, P.J. *et al.* Ankyrin-B mutation causes type 4 long-QT cardiac arrhythmia and sudden cardiac death. *Nature* **421**, 634-9 (2003).
71. Shi, H., Kichaev, G. & Pasaniuc, B. Contrasting the Genetic Architecture of 30 Complex Traits from Summary Association Data. *Am J Hum Genet* **99**, 139-53 (2016).

FIGURE LEGENDS

Figure 1 MEGASTROKE study design. Variants were retained that passed central QC criteria (Methods). Number of cases / number of controls are listed for each ancestry. 1000G, 1000 Genomes; HRC, Haplotype reference consortium; MAF, minor allele frequency; rsq, squared correlation between imputed and true genotypes; imp, measure of imputation quality (Methods); FE, fixed-effects; EUR, European ancestry; AFR African ancestry; EAS, East Asian ancestry; SAS, South Asian ancestry; ASN, mixed Asian ancestries; LAT, Latin American ancestry. P_{het}, heterogeneity p-value; PP_{het}, posterior probability of heterogeneity. * Note the ASN and LAT ancestries were composed of a single study so did not require ancestry specific meta-analysis.

Figure 2 Association results of the transethnic GWAS meta-analysis and the prespecified ancestry-specific meta-analysis in European samples. Shown are novel (red) and replicated (black) genetic loci associated with any stroke or stroke subtypes. The upper panel displays the Manhattan plot from the MANTRA transethnic GWAS meta-analysis for any stroke. The dotted line marks the threshold of statistical significance ($\log_{10}$(Bayes factor) > 6.0).

Figure 3 Genetic overlap between stroke and related vascular traits at the 32 genome-wide significant loci for stroke. (A) Association results from the look-ups in published GWAS data for related vascular traits. Symbol sizes reflect p-values for association with the related trait. (B) Venn diagram. Loci reaching genome-wide significance for association with stroke subtypes are marked by a dagger symbol (for CES), underlined (for LAS), or marked by an asterisk (for SVS). Novel loci are in bold. Note that *SH3PXD2A*, *WNT2B*, *PDE3A* and *OBFC1* have previously been associated with AF (*SH3PXD2A*)[65], DBP (*WNT2B* and *PDE3A*)[24,88] or SBP (*OBFC1*)[89], but the respective lead SNPs were in low LD (r^2<0.1 in 1000G cosmopolitan panel) with variants associated with stroke in the current GWAS. MRI, magnetic resonance imaging; CAD, coronary artery disease; IMT, intima-media thickness. BP, blood pressure; LDL, low density lipoprotein; HDL, high density lipoprotein. Note that the lead variant for TBX3 is not included in the original data sets for BP traits (SBP and DBP). Results are based on a perfect proxy SNP (rs35432, r^2=1 in the European 1000G phase 3 reference).

Figure 4 Shared genetic contribution between stroke and related vascular traits as determined by weighted genetic risk scores (wGRS, upper panel) and LD score regression analysis (lower panel). Effect sizes and significance levels are represented by color and symbol size. β, wGRS effect size; R(g), genetic correlation.Sample sizes for related vascular traits are displayed in Supplementary Table 12.

Figure 5 Connection between stroke risk genes and approved drugs for antithrombotic therapy. Shown are the connections between lead SNPs at stroke risk loci, biological stroke risk genes, and individual targeted drugs. Lead SNPs reaching suggestive evidence for association (MANTRA transethnic meta-analysis $\log_{10}$(Bayes factor) > 5) are shown in grey.

rsID	Chr	Gene(s)	Location relative to gene	Risk allele/ reference allele	Risk allele frequency, %	Phenotype	Analysis	OR	95% CI	P-value	log10 (BF)
Novel associations											
rs880315	1p36	***CASZ1***	Intronic	C/T	40	AS	TRANS	1.05	1.04-1.07	3.62E-10	8.09
rs12037987	1p13	***WNT2B***	Intronic	C/T	16	AS	TRANS	1.07	1.05-1.10	2.73E-08	6.33
rs146390073	1q43	***RGS7***	Intronic	T/C	2	CES	EUR	1.95	1.54-2.47	2.20E-08	NA*
rs12476527	2p23	***KCNK3***	5’-UTR	G/T	48	AS	TRANS	1.05	1.03-1.07	6.44E-08	6.47
rs7610618	3q25	*TM4SF4-TM4SF1*	Intergenic	T/C	1	LAS	EUR	2.33	1.74-3.12	1.44E-08	NA**
rs34311906	4q25	*ANK2*	Intergenic	C/T	41	AIS	EUR	1.07	1.04-1.09	1.07E-08	5.67
rs17612742	4q31	***EDNRA***	Intronic	C/T	21	LAS	TRANS	1.19	1.13-1.26	1.46E-11	9.47
rs6825454	4q31	*FGA*	Intergenic	C/T	31	AIS	TRANS	1.06	1.04-1.08	7.43E-10	7.53
rs11957829	5q23	***LOC100505841***	Intronic	A/G	82	AIS	TRANS	1.07	1.05-1.10	7.51E-09	6.67
rs6891174	5q35	*NKX2-5*	Intergenic	A/G	35	CES	TRANS	1.11	1.07-1.16	5.82E-09	6.96
rs16896398	6p21	*SLC22A7-ZNF318*	Intergenic	T/A	34	AS	TRANS	1.05	1.03-1.07	1.30E-08	6.60
rs42039	7q21	***CDK6***	3’-UTR	C/T	77	AIS	TRANS	1.07	1.04-1.09	6.55E-09	6.84
rs7859727	9p21	**Chr9p21**	ncRNA_intronic	T/C	53	AS	TRANS	1.05	1.03-1.07	4.22E-10	8.01
rs10820405	9q31	***LINC01492***	ncRNA_intronic	G/A	82	LAS	EUR	1.20	1.12-1.28	4.51E-08	4.74
rs2295786	10q24	*SH3PXD2A*	Intergenic	A/T	60	AS	TRANS	1.05	1.04-1.07	1.80E-10	8.34
rs7304841	12p12	***PDE3A***	Intronic	A/C	59	AIS	TRANS	1.05	1.03-1.07	4.93E-08	5.87
rs35436	12q24	*TBX3*	Intergenic	C/T	62	AS	TRANS	1.05	1.03-1.06	2.87E-08	6.29
rs9526212	13q14	***LRCH1***	Intronic	G/A	76	AS	TRANS	1.06	1.04-1.08	5.03E-10	7.97
rs4932370	15q26	*FURIN-FES*	Intergenic	A/G	33	AIS	TRANS	1.05	1.03-1.07	2.88E-08	6.05
rs11867415	17p13	***PRPF8***	Intronic	G/A	18	AIS	TRANS	1.09	1.06-1.13	4.81E-08	6.06
rs2229383	19p13	***ILF3***-*SLC44A2*	Exonic; synon	T/G	65	AIS	TRANS	1.05	1.03-1.07	4.72E-08	6.02
rs8103309	19p13	*SMARCA4-LDLR*	Intergenic	T/C	65	AS	TRANS	1.05	1.03-1.07	3.40E-08	5.85

Previously known associations											
rs12124533	1p13	*TSPAN2*	Intergenic	T/C	24	LAS	TRANS	1.17	1.11-1.23	1.22E-08	6.60
rs1052053	1q22	***PMF1***-*SEMA4A*	Exonic; nonsyn	G/A	40	AS	TRANS	1.06	1.05-1.08	2.70E-14	11.92
rs13143308	4q25	*PITX2*	Intergenic	T/G	28	CES	TRANS	1.32	1.27-1.37	1.86E-47	45.10
rs4959130	6p25	*FOXF2*	Intergenic	A/G	14	AS	TRANS	1.08	1.05-1.11	1.42E-09	7.52
rs2107595	7p21	*HDAC9-TWIST1*	Intergenic	A/G	24	LAS	TRANS	1.21	1.15-1.26	3.65E-15	12.99
rs635634	9q34	*ABO*	Intergenic	T/C	19	AIS	EUR	1.08	1.05-1.11	9.18E-09	4.99
rs2005108	11q22	*MMP12*	Intergenic	T/C	12	AIS	TRANS	1.08	1.05-1.11	3.33E-08	6.12
rs3184504	12q24	***SH2B3***	Exonic; nonsyn	T/C	45	AIS	TRANS	1.08	1.06-1.10	2.17E-14	12.04
rs12932445	16q22	***ZFHX3***	Intronic	C/T	21	CES	TRANS	1.20	1.15-1.25	6.86E-18	15.49
rs12445022	16q24	*ZCCHC14*	Intergenic	A/G	31	AS	TRANS	1.06	1.04-1.08	1.05E-10	8.57

Table 1 Results from the transethnic and fixed effects (transethnic and Europeans-only) GWAS meta-analyses. For each locus the variant reaching the highest BF in the MANTRA or the lowest p-value in the fixed effects transethnic meta-analysis or the fixed effects Europeans-only meta-analysis, respectively, is shown and the respective stroke phenotype showing the strongest association is specified**.** Gene names in bold indicate that the variant is located within the gene; in other cases the first gene corresponds to the closest gene, whereas additional gene names indicate eQTL signals from multiple studies, or from both eQTLs and meQTLs, or genes previously suspected to be causal (LDLR) with a maximum of two genes reported. Note that the lead SNPs in *ILF3-SLC44A2* and *SMARCA-LDLR* are in low LD (r^2=0.082). Chr, chromosome; TRANS, MANTRA transethnic meta-analysis; EUR, Europeans-only fixed-effects meta-analysis; OR, odds ratio; CI, confidence interval; BF, Bayes factor; NA, not assessed; * rs146390073 did not meet the MAF threshold of 0.01 in samples other than those of European ancestry; **rs7610618: The trans-ethnic meta-analysis results showed high heterogeneity (PP_{het}=0.96) and were thus excluded

ONLINE METHODS

Study design and phenotyping

A detailed description of the study design, participating studies, and phenotype definitions for stroke and stroke subtypes is provided in the **Supplementary Note**. Characteristics of study participants are given in **Supplementary Table 2** for each study. All participants provided written informed consent, and local research ethics committees and institutional review boards approved the individual studies.

Genotyping, imputation and quality control

Genotyping platforms and imputation methods for each participating study are described in **Supplementary Table 2**. All studies used imputed genotypes based on at least the 1000Genomes phase 1 multiethnic reference panel and conducted logistic regression analyses (or Cox regression for longitudinal population-based cohort studies) for five stroke traits (AS, AIS, LAS, CES and SVS) with all measured and imputed genetic variants in dosage format using appropriate software under an additive genetic model with a minimum of sex and age as covariates. Information on additional covariates is given in **Supplementary Table 2**.
Before ancestry-specific meta-analysis, quality control (QC) was performed on each study by two independent researchers following a standardized protocol based on the suggestions of Winkler et al.[72] Marker names and alleles were harmonized across studies. Meta-analyses were restricted to autosomal biallelic markers from the 1000Genomes phase1 v3. Duplicate markers were removed from each study. P-Z plots, QQ-plots and allele-frequency-plots were constructed for each study. After visual inspection, analysis and QC was repeated if deemed necessary. QC was conducted independently for all participating studies in at least two sites. Individual study-level filters were set to remove extreme effect values (beta > 5 or beta < -5), rare SNPs (MAF < 0.01) and variants with low imputation accuracy (oevar_imp or info score < 0.5). Effective allele count was defined as twice the product of minor allele frequency, imputation accuracy (r^2, info score ore oevar_imp), and number of cases. Variants with an effective allele count < 10 were excluded.[72] The number of SNPs passing QC for each study is given in **Supplementary Table 26**.

Genome-wide Association Meta-Analyses

The overall analytical strategy is shown in **Figure 1**. We conducted fixed effects inverse variance weighted meta-analysis with METAL[73], first in each ethnic group (EUR, EAS, AFR, SAS, LAT, and other ASN), followed by meta-analysis of ancestry-specific meta-analysis results. We constructed two versions of each meta-analysis: one with single genomic control (GC) applied and one without GC (for LD score regression analysis).
The EUR specific and transethnic fixed effects meta-analysis were further filtered for heterogeneity (p_het < 5.0×10^{-8}) and for the number of cases included for a specific marker. (< 50% of stroke cases were excluded). In addition, we ran a transethnic GWAS meta-analysis using MANTRA.[15] The latter was based on ancestry-specific meta-analysis results. Final MANTRA results were filtered for a MANTRA posterior probability heterogeneity p-value < 0.95. SNPs with $\log_{10}$(BF) > 6 were considered to be genome-wide significant, whereas SNPs with 6 > $\log_{10}$(BF) > 5 were considered to show suggestive association. We used a method based on summary statistics[74] to estimate the variance in liability explained by each lead variant. Disease prevalence was set to 5.5% for AS, to 4.4% for AIS and to 0.11% for IS subtype in Europeans.[75] Disease prevalence was set to 2.97% for AIS, to 0.91% for LAS, to 0.24% for CES and to 1.76% for SVS in East-Asians (unpublished data from the Hisayama study). We used summary statistics from the Europeans-only fixed-effects meta-

analysis and the East-Asian-only fixed-effects meta-analysis. Genomic inflation was calculated as lambda, using the GenABEL package (available through CRAN repositories). In addition, we calculated the LD score regression intercepts for the Europeans-only fixed effects meta-analysis using European LD scores.

Shared genetic influences of individual loci on mechanistically defined stroke subtypes
We used gwas-pw[22] to detect shared genetic influences of LAS, CES and SVS, aiming to identify genetic variants that influence respective pairs of these traits. Gwas-pw estimates the posterior probability (PPA) for four models. Model 3 is the model where a given genomic region contains a genetic variant that influences both traits. We used the fixed-effects transethnic meta-analysis results as input, transforming results into signed Z scores based on p-value and sign of the log(OR). Chunk size (number of SNPs included in each chunk analyzed) was set automatically using an approximately independent block file (ld-select) as provided by the software. Correlation was set to reflect the overlap in controls. We deemed results of model 3 with a $PPA > 0.9$ as significant.[22]

Conditional analysis
We used GCTA-COJO[76] to perform conditional association analysis in each of the stroke loci in Europeans. We first fit a step-wise joint regression model including all SNPs with joint p-values $< 5.0 \times 10^{-8}$. In instances where regions included only one SNP, we fit a model including the top 2 SNPs from each region. The models made use of (i) summary statistics from the Europeans-only meta-analysis presented herein and (ii) genotype data for 3,291 stroke cases and 11,820 controls of North European ancestry from NINDS-SIGN as an LD-reference for each region.

Gene-based analysis
We performed gene-based tests using the VEGAS approach[77] implemented in the VEGAS2 software.[16] We used 24,769 autosomal refseq genes to perform gene-based association studies. We used 1000 genomes phase 3 super populations African (AFR), East-Asian (EAS), European (EUR), American (AMR) and South-Asian (SAS) as a reference to compute pair-wise LD between variants residing within a gene to perform gene-based association tests. We performed gene-based tests using '-top 10' parameter in VEGAS2, which tests enrichment of top 10% of association p-values within a gene. To maintain specificity whilst including cis-regulatory variants, we included variants that are located within 10kb of a gene's 3' and 5' untranslated region (UTR). We performed 1×10^6 simulations to compute empirical p-values association with each gene. For genes with p-value less than 1×10^{-5} we increased the number of simulations to 1×10^8 to increase the accuracy of the association p-values. For individual stroke subtypes, we performed ancestry-specific gene-based association followed by meta-analysis of gene association p-values using Stouffer's method, based on sample size.

Association of individual stroke risk variants with related vascular traits
We systematically explored genetic overlap with AF, CAD, cIMT, cPL, diastolic BP, systolic BP, HDL-cholesterol levels, LDL-cholesterol levels, triglyceride levels, T2D, VTE and WMH. First, we acquired summary statistics from the appropriate consortia (**Supplementary Table 12**). For each of the non-stroke phenotypes we constructed a SNP set including the index variant of the non-stroke phenotype with p-value $< 1.3 \times 10^{-4}$ plus all variants in high LD (r^2 in 1000G EUR > 0.9 with this index variant). If the MEGASTROKE lead SNP was included in this set of SNPs we deemed the overlap with the non-stroke phenotype to be significant. We show two different tiers: i) variants that showed genome-wide significance in the related vascular trait ($p < 5.0 \times 10^{-8}$) and ii) variants that were not genome-wide significant but passed Bonferroni correction ($p = 1.3 \times 10^{-4}$).

Association of genetic risk scores of related vascular traits with stroke and stroke subtypes

Genetic risk scores generated from variants that are shown to be genome-wide associated with various vascular risk factors (VTE, DBP, SBP, mean arterial pressure [MAP], pulse pressure [PP], HTN, HDL-cholesterol, LDL-cholesterol, TG, T2D, CAD) were used to estimate the overlap between vascular traits and stroke and its subtypes. The effect allele for each risk factor variant was defined as the allele associated with increase in the risk factor levels. Corresponding allele information, beta-coefficient and the standard error from different stroke subtypes was extracted and used as input. Association was tested using the inverse-variance weighting (IVW) method implemented as an R package "gtx V 0.0.8" (available through CRAN repositories).

We further conducted a sensitivity analyses using the MR-Egger method implemented as an R package (TwoSampleMR, available through CRAN repositories),[78] which unlike the IVW method estimates the intercept term as part of the analysis. An intercept term significantly differing from zero suggests the presence of directional pleiotropy. We used a conservative significance threshold of $p<0.05$ for the intercept. In the presence of directional pleiotropy, leave-one-out analysis was carried out by re-testing the association of the vascular GRS with the outcome (stroke) leaving out each SNP in turn, to determine whether a single SNP is driving the association. We manually identified outlier SNPs that may be driving the observed directional pleiotropy and we repeated the analyses (IVW and MR-Egger) after excluding the variants exhibiting directional pleiotropy.

The selection of SNPs for the vascular GRS is based on literature (Pubmed) search and the GWAS catalog (http://www.ebi.ac.uk/gwas/) identifying studies that performed GWAS of the various risk factors. The latest and largest GWAS of each risk factor was selected and the associated variant details were retrieved. For the GRS analysis only independent variants ($r^2<0.01$, based on 1000G EUR panel) were used for the analysis (**Supplementary Table 27**). Risk variant selection for BP traits (SBP, DBP, MAP and PP) was further extended to studies with gene-centric chips. We used beta-coefficients extracted from the summary statistics of the International Consortium of BP GWAS[79,80] as weights for this GRS analysis. A p-value of $< 5.6 \times 10^{-3}$ correcting for 9 independent phenotypes was considered significant. The number of independent vascular phenotypes, taking into account correlation between the phenotypes considered, was estimated based on individual level data from the 3C study using the online tool matSpDlite (http://neurogenetics.qimrberghofer.edu.au/matSpDlite/).

Shared genetic contribution to stroke and related vascular traits at the whole genome level

We used LD score regression to estimate the genetic correlation between stroke and related vascular traits.[33,34] We conducted analyses on the European and East-Asian stroke GWAS summary statistics only. Summary statistics from the GWAS meta-analyses for vascular risk factors and intermediate or related vascular phenotypes (BP, blood lipids, T2D, cIMT, cPL, AF, VTE, CAD, WMH) were acquired from the respective consortia, as detailed in **Supplementary Table 12**. For LD-score regression in East-Asians we further received access to unpublished summary statistics of GWAS for blood lipids conducted in BioBank Japan, as described in the **Supplementary Note**. For each trait, we filtered the summary statistics to the subset of HapMap 3 SNPs to reduce the potential for bias due to poor imputation quality. Analyses were performed separately using summary statistics from the European and East Asian-specific meta-analysis. We used the European or East-Asian LD score files calculated from the 1000G reference panel and provided by the developers. A p-value of $< 5.6 \times 10^{-3}$ correcting for 9 independent phenotypes was considered significant. All analyses were performed using the ldsc package (https://github.com/bulik/ldsc).

Global epigenetic patterns at the 32 stroke risk loci

We used the epigwas tool[35] to test for cell-specific enrichment in chromatin marks that were previously shown to be phenotypically cell-type specific in ENCODE and/or RoadMap epigenome data (H3K4me1, H3K4me3, H3K9ac)[35], leveraging the recent release of ENCODE/RoadMap epigenome data from 127 tissue types.[36] Histone ChIP-seq data for narrow contiguous regions of enrichment was used to calculate the enrichment score (height of the nearest tall peak / distance to the peak) for the lead variant and proxies ($r^2 > 0.8$ in the 1000G cosmopolitan panel). Significance was estimated by examining the relative proximity and specificity of the test SNP set with 10,000 sets (permutation) of matched background. In addition, Bonferroni correction for the number of chromatin marks tested was applied.

Pathway Analyses
To identify pathways overrepresented in the stroke association results we used Data-driven Expression-prioritized Integration for Complex Traits (DEPICT[37]), Ingenuity Pathway Analysis (IPA, https://www.qiagenbioinformatics.com/products/ingenuity-pathway-analysis/), and VEGAS2Pathway.[39] DEPICT version 1 rel 194, was used to identify biological pathways, tissues, and cell types enriched among suggestive associations ($\log_{10}$[BF] > 5) for any stroke and stroke subtypes in the MANTRA transethnic GWAS. Results are presented for the MANTRA transethnic analysis. We deemed DEPICT pathways with an FDR <0.05 as statistically significant.
IPA Pathway analysis was conducted using an extended gene list. The latter comprised genes lying in the boundaries defined by $r^2 > 0.5$ with the lead SNP in Europeans or East-Asians, or being located +50kB from the lead SNP, for all suggestive loci reaching $p < 1.0 \times 10^{-5}$ or $\log_{10}$(BF) > 5, and consisted of 214 genes (**Supplementary Table 25**). This gene list was taken as an input for IPA, using only findings from human and experimentally verified results. Otherwise, standard parameters were used for the analysis. We corrected canonical pathway p-value with the Benjamini-Hochberg method and deemed an FDR < 0.05 as significant.
We performed gene-wide gene-set enrichment analysis using the VEGAS2Pathway approach[39] to test which Biosystem terms[38] are enriched with VEGAS2 derived gene association p-values for stroke subtypes. VEGAS2Pathway performs a competitive gene-set enrichment test, while accounting for gene-density in LD blocks (or correlated association p-values of neighbouring genes), SNP density and pathway size using a resampling strategy. For individual stroke subtypes we performed separate ancestry-specific gene-set enrichment analysis. Next, we combined the gene-set enrichment association p-values across ancestry using Stouffer's method for sample size weighted combination of p-values. For each stroke subtype we tested association of 9,981 Biosystem genesets terms.

Fine-mapping derived from credible SNP set analyses
We implemented the method of Maller et al.[81], converting our ancestry-specific meta-analysis p-values to Bayes factors using Wakefield's approximation[40], in all stroke phenotypes in the EUR only, EAS only and AFR only analysis. We used all SNPs in LD with the lead SNP ($r^2 >$ 0.1, ancestry-specific). The Bayes factors were then used to calculate posterior probabilities, based on the assumption of a single causal SNP in each region. For all regions, we constructed 95% credible sets of potentially causal SNPs.

Investigation of eQTLs, pQTLs, meQTLs and regulatory marks in different tissues
The following datasets, covering a large variety of tissue and cell types were interrogated for eQTLs, pQTLs, and meQTLs:
- The Genotype-Tissue Expression (GTEx-V6) project data providing significant eQTL information from 44 post-mortem tissues (449 individuals)

(http://biorxiv.org/content/early/2016/09/09/074450), significance is based on gene-specific p-value threshold that is permutation-adjusted for multiple SNPs per gene.
- The Genome-wide Repository of Associations between SNPs and Phenotypes build 2.0 (GRASP2),[46,47] as well as a collected expression and epigenetic QTL database of >100 sources covering a wide range of cell and tissue types (**Supplementary Note**), using $p<5x10^{-6}$ as a significance threshold for association with expression of a transcript in the original study
- The Human Genetic Variation Database (HGVD)[48] providing eQTL information from peripheral blood cells in a Japanese population (N=1,208) with significance defined by a FDR < 5%.
- The Biobank-based Integrative Omics Studies (BIOS) providing eQTLs from peripheral blood RNA-seq data in 2,116 unrelated individuals[49], significance is defined by FDR < 5%.
- A subset of the Blueprint epigenome project[50] with eQTL, meQTL and histone modification data (H3K4me1 and H3K27ac) in CD14+ monocytes, CD16+ neutrophils and CD4+ naïve T cells from 197 individuals; these were mapped using the classical QTL association test, allele-specific expression (ASE) test and the combined haplotype test, with significance defined by FDR < 5%.
- The Stockholm-Tartu Atherosclerosis Reverse Networks Engineering Task study (STARNET)[51], providing eQTL data from vascular and metabolic tissues in 600 CAD patients, with association p-values corrected by Benjamini-Hochberg ($p < 0.05$)
- The aortic endothelial cells study[52] providing eQTL data from human aortic endothelial cells in 147 individuals, with Bonferroni multiple testing correction for the number of independent SNPs ($p < 1.0 \times 10^{-4}$)
- The ARIC cohort[53] providing meQTL information from peripheral blood in 794 of European ancestry and 784 of African-American ancestry individuals from, with multiple testing correction for the number of unique CpG probes in the look-up.
- The Cooperative health Research in the region of Augsburg (KORA) cohort with pQTL information from the human blood plasma proteome[54] measuring 1,124 proteins on the SomaSCAN platform in 1,000 participants. Significance for each association was set at $p < 5.0 \times 10^{-8}$.

In each of these datasets we report the most significant *cis* QTL, meQTL, or pQTL surpassing a study-specific predefined significance level or FDR, considering only QTLs in LD with the lead stroke SNP at an $r^2>0.8$ (in 1000G, as well as queries of multiple builds of SNAP[82] and SNiPA[83]), suggesting high concordance. Results are presented grouped per tissue or cell type (**Supplementary Table 23**), or per stroke risk locus (**Supplementary Table 22)**. In addition, we also systematically report the association of the top QTL with stroke risk, and of the lead stroke risk variant with the corresponding transcript expression, methylation level, or protein level (**Supplementary Table 23**).

In addition we used a subset of the Blueprint epigenome project in CD14+ monocytes, CD16+ neutrophils and CD4+ naïve T cells from 197 individuals[50] and Haploreg V4[84] to annotate the lead variants and proxies for enrichment in specific histone modification marks for the chromatin state, based on ChIP-Seq data from multiple cell/tissue types from ENCODE (Encyclopedia of DNA Elements)[85] and NIH RoadMap epigenome.[36] Results for each of the lead SNPs and its proxies are displayed in detail in **Supplementary Table 22**.

Integration of association statistics and in silico functional information using RiVIERA-beta

To identify the most plausible causal variants and genes we used the RiVIERA software[59], which jointly models the summary association statistics and the corresponding epigenetic regulatory information in a Bayesian framework to estimate the PPA. The empirical prior of a variant to be associated with the respective trait through regulatory features was generated using the 848 tissue-specific epigenomic data in 7 chromatin (H3K4me1, H3K4me3, H3K36me3, H3K27me3, H3K9me3, H3K27ac, H3K9ac) and DNA accessibility (DNase I) marks from the ENCODE/RoadMap epigenome data. Binary epigenomic annotation matrices of a variant overlapping the narrow peaks were generated. For inferring the causal region, RiVIERA-beta performs a repeated (n=1,000) random sampling step per locus, with the step size set to 1.0×10^{-4}. Iteration is performed until the convergence (acceptance rate of > 60 %) is achieved, which is critical for the accurate estimation of PPA. We generated 95% credible sets in each region based on the PPA. Regional plots were generated using the association statistics and the PPA. Epigenetic enrichment over a fixed window size (50bp) per tissue group was generated, by taking cumulative sum of empirical prior weighted global epigenetic enrichment. Tissues were grouped into 19 groups as defined in the NIH RoadMap epigenome project.

Scoring method
In an attempt to prioritize the most likely biological candidate genes, we integrated functional and biological information into an empirical score for each of the genes residing in the 32 genome-wide significant loci. These comprised 149 genes within the region defined by an $r^2 > 0.5$ in any of the 1000G European or East-Asian populations or physical distances of ±50 kb from the lead SNP of the respective locus (**Supplementary Table 25**). A score of 1 was assigned for being the nearest gene to the lead SNP, for harboring a missense variant, for harboring histone marks H3K4me3, H3K9ac and H3K4me1 peaks in cells types that showed significant enrichment in epigwas analysis, and functioning as an eGene for an eQTL, meQTL, or pQTL (1 point for each) in at least one study and one tissue type. In addition, a score of 1 was assigned for each stroke phenotype showing evidence of being a drug target gene in the DrugBank database (ATC-C and ATC-B01) and the Therapeutic Target Database (**Supplementary Table 25**), and for overlap with biological pathways in DEPICT, IPA, or VEGAS2 (**Supplementary Tables 18 to 20**).

Drug-Target gene enrichment analysis
For each locus containing a variant with $\log_{10}(BF) > 5$ in the MANTRA analysis, we annotated the genes by considering LD structures ($r^2 > 0.5$ in any of 1KG EUR or ASN populations) or physical distances (±50 kbp) from the lead SNP of the respective locus. Drug target genes were extracted from the DrugBank database[86] (considering those registered as pharmacological "active targets; https://www.drugbank.ca/) and Therapeutic Target Database[87] (TTD; http://bidd.nus.edu.sg/group/cjttd/TTD_HOME.asp) resulting in a list of 1,123 genes (and corresponding proteins) annotated to currently approved drugs indicated for any diseases (**Supplementary Table 25**). Drugs indicated for antithrombotic therapy (n = 69) and cardiovascular diseases (n = 324) were curated from Anatomical Therapeutic Chemical (ATC) codes (**Supplementary Table 25**). Enrichment of overlap between stroke-associated genes with drug targets for antithrombotic therapy and cardiovascular diseases were assessed by Fisher's exact test.

Data availability
The datasets generated and/or analyzed during the current study are available from the corresponding authors upon reasonable request.

METHODS-ONLY REFERENCES

72. Winkler, T.W. *et al.* Quality control and conduct of genome-wide association meta-analyses. *Nat Protoc* **9**, 1192-212 (2014).
73. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-1 (2010).
74. So, H.C., Gui, A.H., Cherny, S.S. & Sham, P.C. Evaluating the heritability explained by known susceptibility variants: a survey of ten complex diseases. *Genet Epidemiol* **35**, 310-7 (2011).
75. Feigin, V.L., Lawes, C.M., Bennett, D.A. & Anderson, C.S. Stroke epidemiology: a review of population-based studies of incidence, prevalence, and case-fatality in the late 20th century. *Lancet Neurol* **2**, 43-53 (2003).
76. Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet* **44**, 369-75, S1-3 (2012).
77. Liu, J.Z. *et al.* A versatile gene-based test for genome-wide association studies. *Am J Hum Genet* **87**, 139-45 (2010).
78. Bowden, J., Davey Smith, G. & Burgess, S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol* **44**, 512-25 (2015).
79. International Consortium for Blood Pressure Genome-Wide Association Studies *et al.* Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* **478**, 103-9 (2011).
80. Wain, L.V. *et al.* Genome-wide association study identifies six new loci influencing pulse pressure and mean arterial pressure. *Nat Genet* **43**, 1005-11 (2011).
81. Wellcome Trust Case Control Consortium *et al.* Bayesian refinement of association signals for 14 loci in 3 common diseases. *Nat Genet* **44**, 1294-301 (2012).
82. Johnson, A.D. *et al.* SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics* **24**, 2938-9 (2008).
83. Arnold, M., Raffler, J., Pfeufer, A., Suhre, K. & Kastenmuller, G. SNiPA: an interactive, genetic variant-centered annotation browser. *Bioinformatics* **31**, 1334-6 (2015).
84. Ward, L.D. & Kellis, M. HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. *Nucleic Acids Res* **44**, D877-81 (2016).
85. Encode Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74 (2012).
86. Wishart, D.S. *et al.* DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res* **34**, D668-72 (2006).
87. Yang, H. *et al.* Therapeutic target database update 2016: enriched resource for bench to clinical drug target and targeted pathway information. *Nucleic Acids Res* **44**, D1069-74 (2016).
88. Kato, N. *et al.* Meta-analysis of genome-wide association studies identifies common variants associated with blood pressure variation in east Asians. *Nat Genet* **43**, 531-8 (2011).
89. Surendran, P. *et al.* Trans-ancestry meta-analyses identify rare and common variants associated with blood pressure and hypertension. *Nat Genet* **48**, 1151-61 (2016).