Stroke genetics informs drug discovery and risk prediction across ancestries

Stéphanie Debette (stephanie.debette@u-bordeaux.fr)
University of Bordeaux, Inserm, Bordeaux Population Health Research Center, UMR 1219
https://orcid.org/0000-0001-8675-7968

Aniket Mishra
University of Bordeaux https://orcid.org/0000-0002-8141-1543

Rainer Malik
Institute for Stroke and Dementia Research (ISD), University Hospital, LMU Munich, Munich, Germany.

Tsuyoshi Hachiya
Disaster Reconstruction Center, Iwate Medical University

Tuuli Jürgenson
Estonian Genome Centre, Institute of Genomics, University of Tartu, Tartu, Estonia

Shinichi Namba
National Cancer Center Research Institute https://orcid.org/0000-0002-7486-3146

Masaru Koido
Division of Molecular Pathology, Institute of Medical Sciences, The University of Tokyo, Tokyo, Japan

Quentin Le Grand
University of Bordeaux, Inserm, Bordeaux Population Health Research Center, team VINTAGE, UMR 1219
https://orcid.org/0000-0002-9299-0747

Frederick Kamanu
TIMI Study Group, Boston, MA, USA

Mingyang Shi
Laboratory of Complex Trait Genomics, Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan

Yunye He
Laboratory of Complex Trait Genomics, Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan

Marios Georgakis
Institute for Stroke and Dementia Research (ISD), University Hospital, LMU Munich, Munich, Germany.

Ilana Caro
University of Bordeaux, Inserm, Bordeaux Population Health Research Center, team VINTAGE, UMR 1219, F-33000 Bordeaux, France

Kristi Krebs
Estonian Genome Centre, Institute of Genomics, University of Tartu, Tartu, Estonia

Felix Vaura
University of Turku https://orcid.org/0000-0002-6036-889X
Naomi Habib
Hebrew University of Jerusalem https://orcid.org/0000-0002-6049-2487
Bendik Winsvold
Oslo University Hospital https://orcid.org/0000-0003-4171-8919
Yon Ho Jee
Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, United States of America
Jesper Qvist Thomassen
Rigshospitalet https://orcid.org/0000-0003-3484-9531
Vida Abedi
University of Memphis
Jara Cárcel-Márquez
Stroke Pharmacogenomics and Genetics Laboratory, Biomedical Research Institute Sant Pau (IIB Sant Pau), Barcelona, Spain
Kuang Lin
University of Oxford
Marianne Nygaard
Danish Twin Registry https://orcid.org/0000-0003-0703-2665
Ganesh Chauhan
Rajendra Institute of Medical Sciences, Ranchi, Jharkhand, INDIA
Hampton Leonard
NIA https://orcid.org/0000-0003-2390-8110
Chaojie Yang
Center for Public Health Genomics, University of Virginia
Ekaterina Yonova-Doing
British Heart Foundation Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK
Maria Knol
Erasmus MC University Medical Center https://orcid.org/0000-0002-3597-1531
Tetsuro Ago
Department of Medicine and Clinical Science, Graduate School of Medical Sciences, Kyushu University
Philippe Amouyel
Institut Pasteur de Lille https://orcid.org/0000-0001-9088-234X
Christopher Anderson
Department of Neurology, Brigham and Women's Hospital, Boston, MA, USA
Nicole Armstrong
Department of Epidemiology, University of Alabama at Birmingham

Mark Bakker
UMC Utrecht Brain Center, Department of Neurology and Neurosurgery, University Medical Center Utrecht, University Utrecht, Utrecht, The Netherlands
Traci Bartz
University of Washington
Joshua Bis
Cardiovascular Health Research Unit https://orcid.org/0000-0002-3409-1110
Constance Bordes
University of Bordeaux, Inserm, Bordeaux Population Health Research Center, team VINTAGE, UMR 1219 https://orcid.org/0000-0003-3367-1778
Sigrid Borte
Institute of Clinical Medicine, University of Oslo https://orcid.org/0000-0001-9540-3256
Anael Cain
The Edmond and Lily Safra Center for Brain Sciences, The Hebrew University of Jerusalem, Israel
Paul Ridker
Division of Preventive Medicine, Brigham and Women's Hospital, Boston, MA 02215
Zhengming Chen
University of Oxford https://orcid.org/0000-0001-6423-105X
Michael Chong
Population Health Research Institute (PHRI), David Braley Cardiac, Vascular and Stroke Research Institute, Hamilton Health Sciences, Thrombosis and Atherosclerosis Research Institute, 237 Barton Str
John Cole
VA Maryland Health Care System, Baltimore, MD, USA
Rafael de Cid
Institute for Health Science Research Germans Trias i Pujol (IGTP) https://orcid.org/0000-0003-3579-6777
Matthias Endres
Charité - Universitätsmedizin Berlin, Klinik und Hochschulambulanz für Neurologie und Centrum für Schlaganfallforschung Berlin (CSB) https://orcid.org/0000-0001-6520-3720
Leslie Ferreira
Post-Graduation Program on Health and Environment, Department of Medicine and Joinville Stroke Biobank, University of the Region of Joinville
Natalie Gasca
Department of Biostatistics, University of Washington
Vilmundur Gudnason
Icelandic Heart Association https://orcid.org/0000-0001-5696-0084
Jun Hata
Department of Epidemiology and Public Health, Graduate School of Medical Sciences, Kyushu University

Aki Havulinna
Finnish Institute for Health and Welfare, Institute for Molecular Medicine Finland
https://orcid.org/0000-0002-4787-8959
Jemma Hopewell
CTSU - Nuffield Department of Population Health
Hyacinth Hyacinth
Department of Neurology and Rehabilitation Medicine, University of Cincinnati College of Medicine, Cincinnati, OH
Michael Inouye
University of Cambridge https://orcid.org/0000-0001-9413-6520
Mina Jacob
Department of Neurology, Donders Center for Medical Neuroscience, Radboud University Medical Center, Nijmegen, The Netherlands
Christina Jeon
Los Angeles County Department of Public Health
Christina Jern
Institute of Biomedicine, Department of Laboratory Medicine, the Sahlgrenska Academy, University of Gothenburg, Box 440, SE-405 30 Gothenburg, Sweden
Masahiro Kamouchi
Department of Health Care Administration and Management, Graduate School of Medical Sciences, Kyushu University
Keith Keene
Department of Biology; Brody School of Medicine Center for Health Disparities, East Carolina University, Greenville, NC
Takanari Kitazono
Department of Medicine and Clinical Science, Graduate School of Medical Sciences, Kyushu University
Steven Kittner
Department of Neurology, University of Maryland School of Medicine, Baltimore, MD, USA
Takahiro Konuma
Department of Statistical Genetics, Osaka University Graduate School of Medicine, Suita, Japan
Amit Kumar
Rajendra Institute of Medical Sciences, Ranchi, Jharkhand, INDIA
Paul Lacaze
Department of Epidemiology and Preventive Medicine, School of Public Health and Preventive Medicine, Monash University, Melbourne, VIC 3004, Australia.
Lenore Launer
National Institute on Aging, National Institutes of Health https://orcid.org/0000-0002-3238-7612
Kaido Lepik
Estonian Genome Centre, Institute of Genomics, University of Tartu, Tartu, Estonia

Jiang Li
https://orcid.org/0000-0002-7006-1285
Liming Li
Peking University https://orcid.org/0000-0001-5873-7089
Ani Manichaikul
University of Virginia School of Medicine https://orcid.org/0000-0002-5998-795X
Hugh Markus
Stroke Research Group, Department of Clinical Neurosciences, University of Cambridge
Nicholas Marston
TIMI Study Group, Boston, MA, USA
Thomas Meitinger
Institute of Human Genetics, Technical University of Munich, 81675 Munich, Germany.
Braxton Mitchell
University of Maryland School of Medicine
Felipe Montellano
Institute of Clinical Epidemiology and Biometry, University of Würzburg https://orcid.org/0000-0002-9438-0854
Takayuki Morisaki
Division of Molecular Pathology, Institute of Medical Sciences, The University of Tokyo, Tokyo, Japan
Thomas Mosley
University of Mississippi Medical Center
Mike Nalls
Center for Alzheimer's and Related Dementias, National Institutes of Health, Bethesda, MD, USA
Børge Nordestgaard
University of Copenhagen https://orcid.org/0000-0002-1954-7220
Martin O'Donnell
College of Medicine Nursing and Health Science NUI Galway, Ireland
Yukinori Okada
Osaka University https://orcid.org/0000-0002-0311-8472
Guillaume Pare
Population Health Research Institute, McMaster University https://orcid.org/0000-0002-6795-4760
Annette Peters
Deutsches Forschungszentrum für Gesundheit und Umwelt (GmbH) https://orcid.org/0000-0001-6645-0985
Bruce Psaty
Cardiovascular Health Research Unit https://orcid.org/0000-0002-7278-2190
Stephen Rich
University of Virginia https://orcid.org/0000-0003-3872-7793
Jonathan Rosand

McCance Center for Brain Health, Massachusetts General Hospital, Boston, MA, USA

Marc Sabatine

TIMI Study Group, Boston, MA, USA

Ralph Sacco

Miller Schoool Of Medcine

Danish Saleheen

Columbia University Medical Center https://orcid.org/0000-0001-6193-020X

Else Charlotte Sandset

Stroke Unit, Department of Neurology, Oslo University Hospital, Oslo, Norway

Muralidharan Sargurupremraj

https://orcid.org/0000-0003-1684-3750

Makoto Sasaki

Iwate Medical University https://orcid.org/0000-0002-3108-4361

Claudia Satizabal

The University of Texas Health Science Center at San Antonio https://orcid.org/0000-0002-1115-4430

Carsten Schmidt

University Medicine Greifswald

Atsushi Shimizu

https://orcid.org/0000-0001-8307-2461

Nicholas Smith

University of Washington

Daniel Strbian

Department of Neurology, Helsinki University Hospital and University of Helsinki

Yoichi Sutoh

Disaster Reconstruction Center, Iwate Medical University

Kozo Tanno

Disaster Reconstruction Center, Iwate Medical University https://orcid.org/0000-0002-1264-0684

Steffen Tiedt

University Hospital, LMU Munich https://orcid.org/0000-0002-8817-8457

Nuria Torres-Aguila

Stroke Pharmacogenomics and Genetics Laboratory, Biomedical Research Institute Sant Pau (IIB Sant Pau), Barcelona, Spain

David-Alexandre Trégouët

INSERM, UMR_S937, ICAN Institute, Université Pierre et Marie Curie

Stella Trompet

Leiden University Medical Center https://orcid.org/0000-0001-5006-0528

Anil Tuladhar

Department of Neurology, Donders Center for Medical Neuroscience, Radboud University Medical Center, Nijmegen, The Netherlands

Anne Tybjærg-Hansen
Rigshospitalet
Marion van Vugt
Division Heart & Lungs, Department of Cardiology, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands.
Riina Vibo
Department of Neurology and Neurosurgery, University of Tartu, Estonia https://orcid.org/0000-0001-9138-3227
Kerri Wiggins
University of Washington https://orcid.org/0000-0003-2749-1279
Daniel Woo
University of Cincinnati College of Medicine
Huichun Xu
University of Maryland School of Medicine
Qiong Yang
Boston University School of Medicine https://orcid.org/0000-0002-3658-1375
Mark Lathrop
Department of Human Genetics, McGill University-Génome Québec Innovation Centre, Québec
Iona Millwood
University of Oxford
Christian Gieger
Helmholtz Zentrum Munich https://orcid.org/0000-0001-6986-9554
Toshiharu Ninomiya
Department of Epidemiology and Public Health, Graduate School of Medical Sciences, Kyushu University
Hans Grabe
University Medicine Greifswald
J Wouter Jukema
Leiden University Medical Center https://orcid.org/0000-0002-3246-8359
Ina Rissanen
Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, The Netherlands
Sudha Seshadri
University of Texas Health Sciences Center https://orcid.org/0000-0001-6135-2622
William Longstreth
University of Washington
Daniel Chasman
Brigham and Women's Hospital https://orcid.org/0000-0003-3357-0862
Joanna Howson
Novo Nordisk Ltd https://orcid.org/0000-0001-7618-0050

Marguerite Irvin
University of Alabama at Birmingham School of Public Health
Hieab Adams
Erasmus University Medical Center https://orcid.org/0000-0003-3687-2508
Sylvia Wassertheil-Smoller
Department of Epidemiology and Population Health, Albert Einstein College of Medicine, New York, NY
Kaare Christensen
University of Southern Denmark https://orcid.org/0000-0002-5429-5292
M. Arfan Ikram
Erasmus University Medical Center https://orcid.org/0000-0003-0372-8585
Tatjana Rundek
Department of Neurology and Evelyn F. McKnight Brain Institute, University of Miami
Jerome Rotter
Lundquist Institute for Biomedical Innovations (Formerly LABioMed) at Harbor-UCLA Medical Center
Moeen Riaz
School of Public Health and Preventive Medicine, Monash University
Eleanor Simonsick
Longitudinal Studies Section, Translational Gerontology Branch
Janika Kõrv
Department of Neurology and Neurosurgery, University of Tartu, Estonia
Paulo França
Post-Graduation Program on Health and Environment, Department of Medicine and Joinville Stroke Biobank, University of the Region of Joinville
Myriam Fornage
University of Texas Health Science Center at Houston https://orcid.org/0000-0003-0677-8158
Ramin Zand
University of Tennessee Health Science Center
Kameshwar Prasad
Rajendra Institute of Medical Sciences, Ranchi, Jharkhand, INDIA
Ruth Frikke-Schmidt
Rigshospitalet
Frank-Erik de Leeuw
Department of Neurology, Donders Center for Medical Neuroscience, Radboud University Medical Center, Nijmegen, The Netherlands
Thomas Liman
Klinik für Neurologie, Carl von Ossietzky University of Oldenburg
Karl Georg Haeusler
Department of Neurology, Universitätsklinikum Würzburg, Germany
Ynte Ruigrok

UMC Utrecht Brain Center, Department of Neurology and Neurosurgery, University Medical Center Utrecht, University Utrecht, Utrecht, The Netherlands

Peter Heuschmann
Institute of Clinical Epidemiology and Biometry, University of Würzburg

Keum Jung
Yonsei University

John-Anker Zwart
https://orcid.org/0000-0001-5721-0154

Teemu Niiranen
Department of Internal Medicine, Turku University Hospital and University of Turku, Turku, Finland

Christian Ruff
TIMI Study Group, Boston, MA, USA

Israel Fernández-Cadenas
Stroke Pharmacogenomics and Genetics Laboratory, Biomedical Research Institute Sant Pau (IIB Sant Pau), Barcelona, Spain

Robin Walters
University of Oxford https://orcid.org/0000-0002-9179-0321

Lili Milani
University of Tartu https://orcid.org/0000-0002-5323-3102

Yoichiro Kamatani
Kyoto University Graduate School of Medicine

Martin Dichgans
Ludwig Maximilian University of Munich https://orcid.org/0000-0002-0654-387X

Biological Sciences - Article

Keywords:

Posted Date: January 4th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1175817/v1

Version of Record: A version of this preprint was published at Nature on September 30th, 2022. See the published version at https://doi.org/10.1038/s41586-022-05165-3.

Stroke genetics informs drug discovery and risk prediction across ancestries

Aniket Mishra[1,153], Rainer Malik[2,153], Tsuyoshi Hachiya[3,153], Tuuli Jürgenson[4,5,153], Shinichi Namba[6], Masaru Koido[7], Quentin Le Grand[1], Frederick K. Kamanu[8,9], Mingyang Shi[10], Yunye He[10], Marios K. Georgakis [2,11,12], Ilana Caro[1], Kristi Krebs[4], Felix C. Vaura[13,14], Naomi Habib[15], Bendik Slagsvold Winsvold[16,17,18], Yon Ho Jee[19], Jesper Qvist Thomassen[20], Vida Abedi[21,22], Jara Cárcel-Márquez[23,24], Kuang Lin[25], Marianne Nygaard[26,27], Ganesh Chauhan[28], Hampton L. Leonard[29,30,31], Chaojie Yang[32,33], Ekaterina Yonova-Doing[34,35], Tetsuro Ago[36], Philippe Amouyel[37,38,39], Christopher D. Anderson[40,11,41], Nicole D. Armstrong[42], Mark K. Bakker[43], Traci M. Bartz [44,45], Joshua C. Bis [44], Constance Bordes[1], Sigrid Børte[46,17,47], Anael Cain[15], Paul M. Ridker[48,49], Zhengming Chen[25], Michael R. Chong [51,52], John W. Cole [53,54], Rafael de Cid [55], Matthias Endres[56,57,58], Leslie E. Ferreira[59], Natalie C. Gasca[60], Vilmundar Gudnason[61,62], Jun Hata[63], Aki S Havulinna[64,65], Jemma C Hopewell[66], Hyacinth I Hyacinth[67], Michael Inouye[68, 69,34,70,71], Mina A. Jacob[72], Christina E. Jeon[73], Christina Jern[74,75], Masahiro Kamouchi[76], Keith L. Keene[77], Takanari Kitazono[36], Steven J. Kittner[54,78], Takahiro Konuma[6,79], Amit Kumar[28], Paul Lacaze[80], Lenore J. Launer[81], Kaido Lepik[4,82,83,84], Jiang Li[21], Liming Li[85], Ani Manichaikul[32], Hugh S. Markus[86], Nicholas A. Marston[8,9], Thomas Meitinger[87,88], Braxton D. Mitchell[89,90], Felipe Montellano[91,92], Takayuki Morisaki[7], Thomas H. Mosley[93], Mike A. Nalls[29,30,31], Børge G. Nordestgaard[94,95], Martin J. O'Donnell[96], Yukinori Okada[6,97,98,99,100], Guillaume Paré[51,52,101], Annette Peters[102,103,104], Bruce M. Psaty[44,105,106], Stephen S. Rich[32], Jonathan Rosand[32,107,41], Marc S. Sabatine[8,9], Ralph L. Sacco[108,109], Danish Saleheen[110], Else Charlotte Sandset[111,112], Muralidharan Sargurupremraj[113], Makoto Sasaki[3], Claudia L. Satizabal[113,114], Carsten O. Schmidt[115], Atsushi Shimizu[3], Nicholas L. Smith[105,116], Daniel Strbian[117], Yoichi Sutoh[3], Kozo Tanno[3], Steffen Tiedt[2], Nuria P. Torres-Aguila[23], David-Alexandre Trégouët[1], Stella Trompet[118,119], Anil Man Tuladhar[72], Anne Tybjærg-Hansen[20,95], Marion van Vugt[120], Riina Vibo[121], Kerri L. Wiggins[44], Daniel Woo[122], Huichun Xu[89], Qiong Yang[123,114], G. Mark Lathrop[124], the COMPASS Consortium[125], the INVENT consortium[125], the Dutch Parelsnoer initiative (PSI) Cerebrovascular Disease Study Group[125], the PRECISE4Q consortium[125], the NINDS Stroke Genetics Network (SiGN) [125], the MEGASTROKE Consortium[125], the China Kadoorie Biobank Collaborative Group[125], the International Stroke Genetics Consortium (ISGC) [125], the CHARGE Consortium[125], the GIGASTROKE Consortium[125], Iona Y Millwood[25,50], Christian Gieger[126], Toshiharu Ninomiya[63], Hans J. Grabe[127,128], J Wouter Jukema[119,129,130], Ina L. Rissanen[131], Sudha

Seshadri[113,114,132], W. T. Longstreth[105,133], Daniel L. Chasman[48,49] , Joanna MM. Howson[34,35], Marguerite R. Irvin[42], Hieab Adams[134], Sylvia Wasssertheil-Smoller[135], Kaare Christensen[26,27,136], Mohammad A. Ikram[137], Tatjana Rundek[108,109], Jerome I. Rotter[138], Moeen Riaz[80], Eleanor M. Simonsick[139], Janika Kõrv[121], Paulo H.C. França[59], Myriam Fornage[140,141], Ramin Zand[142,143], Kameshwar Prasad[28], Ruth Frikke-Schmidt[20, 95], Frank-Erik de Leeuw[72], Thomas Liman[144,57,145], Karl Georg Haeusler[146], Ynte M. Ruigrok[43], Peter Ulrich Heuschmann[91,147,148], Keum Ji Jung [149,25], John-Anker Zwart[16,46,17], Teemu J. Niiranen[13,14], Christian T Ruff[8,9], Israel Fernández-Cadenas[23], Robin G. Walters[25, 50], Lili Milani[4,154], Yoichiro Kamatani[10,154], Martin Dichgans[2,150,151,154*], Stephanie Debette[1,132,152,154*]

1 University of Bordeaux, Inserm, Bordeaux Population Health Research Center, team VINTAGE, UMR 1219, F-33000 Bordeaux, France

2 Institute for Stroke and Dementia Research (ISD), University Hospital, LMU Munich, Munich, Germany.

3 Iwate Tohoku Medical Megabank Organization, Iwate Medical University

4 Estonian Genome Centre, Institute of Genomics, University of Tartu, Tartu, Estonia

5 Institute of Mathematics and Statistics, University of Tartu, Tartu, Estonia

6 Department of Statistical Genetics, Osaka University Graduate School of Medicine, Suita 565-0871, Japan

7 Division of Molecular Pathology, Institute of Medical Sciences, The University of Tokyo, Tokyo, Japan

8 TIMI Study Group, Boston, MA, USA

9 Division of Cardiovascular Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA

10 Laboratory of Complex Trait Genomics, Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan

11 Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA

12 Program in Medical and Population Genetics, Broad Institute of Harvard and the Massachusetts Institute of Technology, Cambridge, MA, USA

13 Department of Internal Medicine, University of Turku, Turku, Finland

14 Department of Public Health and Welfare, Finnish Institute for Health and Welfare, Turku, Finland

15 The Edmond and Lily Safra Center for Brain Sciences, The Hebrew University of Jerusalem, Israel

16 Department of Research and Innovation, Division of Clinical Neuroscience, Oslo University Hospital, Oslo, Norway
17 K. G. Jebsen Center for Genetic Epidemiology, Department of Public Health and Nursing, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology (NTNU), Trondheim, Norway
18 Department of Neurology, Oslo University Hospital, Oslo, Norway
29 Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, United States of America
20 Department of Clinical Biochemistry, Copenhagen University Hospital - Rigshospitalet, Copenhagen, Denmark
21 Department of Molecular and Functional Genomics, Weis Center for Research, Geisinger Health System
22 Department of Public Health Sciences, College of Medicine, The Pennsylvania State University
23 Stroke Pharmacogenomics and Genetics Laboratory, Biomedical Research Institute Sant Pau (IIB Sant Pau), Barcelona, Spain
24 Universitat Autònoma de Barcelona, Departament de Medicina
25 Nuffield Department of Population Health, University of Oxford, Oxford, UK
26 The Danish Twin Registry, Department of Public Health, University of Southern Denmark, Odense, Denmark
27 Department of Clinical Genetics, Odense University Hospital, Odense, Denmark
28 Rajendra Institute of Medical Sciences, Ranchi, Jharkhand, INDIA
29 Center for Alzheimer's and Related Dementias, National Institutes of Health, Bethesda, MD, USA
30 Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA
31 Data Tecnica International LLC, Glen Echo, MD, USA
32 Center for Public Health Genomics, University of Virginia
33 Department of Biochemistry and Molecular Genetics, University of Virginia
34 British Heart Foundation Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK
35 Department of Genetics, Novo Nordisk Research Centre Oxford, Oxford, UK.
36 Department of Medicine and Clinical Science, Graduate School of Medical Sciences, Kyushu University

37 Univ. Lille, U1167 - RID-AGE - LabEx DISTALZ - Risk factors and molecular determinants of aging-related diseases, F-59000 Lille, France
38 Inserm, U1167, F-59000 Lille, France
39 CHU Lille, Public Health Department, F-59000 Lille, France
40 Department of Neurology, Brigham and Women's Hospital, Boston, MA, USA
41 Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, USA
42 Department of Epidemiology, University of Alabama at Birmingham
43 UMC Utrecht Brain Center, Department of Neurology and Neurosurgery, University Medical Center Utrecht, University Utrecht, Utrecht, The Netherlands
44 Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA, USA
45 Department of Biostatistics, University of Washington, Seattle, WA, USA
46 Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway
47 Research and Communication Unit for Musculoskeletal Health (FORMI), Department of Research and Innovation, Division of Clinical Neuroscience, Oslo University Hospital, Oslo, Norway
48 Division of Preventive Medicine, Brigham and Women's Hospital, Boston, MA 02215
49 Harvard Medical School, Boston, MA 02115
50 MRC Population Health Research Unit, University of Oxford, Oxford, UK
51 Population Health Research Institute (PHRI), David Braley Cardiac, Vascular and Stroke Research Institute, Hamilton Health Sciences, Thrombosis and Atherosclerosis Research Institute, 237 Barton Street East, Hamilton, L8L2X2, Ontario, Canada
52 Department of Pathology and Molecular Medicine, Michael G. DeGroote School of Medicine, McMaster University, 1280 Main Street West, Hamilton, L8S 4K1, Ontario, Canada
53 VA Maryland Health Care System, Baltimore, MD, USA
54 Department of Neurology, University of Maryland School of Medicine, Baltimore, MD, USA
55 GenomesForLife-GCAT Lab Group, Germans Trias i Pujol Research Institute (IGTP), Badalona, Spain. On Behalf GCAT study.
56 Klinik und Hochschulambulanz für Neurologie, Charité-Universitätsmedizin Berlin
57 Center for Stroke Research Berlin
58 ExcellenceCluster NeuroCure; German Center for Neurodegenerative Diseases (DZNE), partner site Berlin; German Centre for Cardiovascular Research (DZHK), partner site Berlin

59 Post-Graduation Program on Health and Environment, Department of Medicine and Joinville Stroke Biobank, University of the Region of Joinville
60 Department of Biostatistics, University of Washington
61 Icelandic Heart Association
62 Univ of Icelaand
63 Department of Epidemiology and Public Health, Graduate School of Medical Sciences, Kyushu University
64 Department of Public Health and Welfare, Finnish Institute for Health and Welfare, Helsinki, Finland
65 Institute for Molecular Medicine Finland, FIMM-HiLIFE, Helsinki, Finland
66 CTSU - Nuffield Department of Population Health
67 Department of Neurology and Rehabilitation Medicine, University of Cincinnati College of Medicine, Cincinnati, OH
68 Cambridge Baker Systems Genomics Initiative, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK
69 Cambridge Baker Systems Genomics Initiative, Baker Heart and Diabetes Institute, 75 Commercial Rd, Melbourne 3004, Victoria, Australia
70 British Heart Foundation Centre of Research Excellence, University of Cambridge, Cambridge, UK
71 Health Data Research UK Cambridge, Wellcome Genome Campus and University of Cambridge, Cambridge, UK
72 Department of Neurology, Donders Center for Medical Neuroscience, Radboud University Medical Center, Nijmegen, The Netherlands
73 Los Angeles County Department of Public Health
74 Institute of Biomedicine, Department of Laboratory Medicine, the Sahlgrenska Academy, University of Gothenburg, Box 440, SE-405 30 Gothenburg, Sweden
75 Region Västra Götaland, Sahlgrenska University Hospital, Department of Clinical Genetics and Genomics, Gothenburg, Sweden
76 Department of Health Care Administration and Management, Graduate School of Medical Sciences, Kyushu University
77 Department of Biology; Brody School of Medicine Center for Health Disparities, East Carolina University, Greenville, NC
78 Department of Neurology and Geriatric Research and Education Clinical Center, VA Maryland Health Care System, Baltimore, MD

79 Central Pharmaceutical Research Institute, JAPAN TOBACCO INC., Takatsuki 569-1125, Japan
80 Department of Epidemiology and Preventive Medicine, School of Public Health and Preventive Medicine, Monash University, Melbourne, VIC 3004, Australia.
81 Intramural Rearch Program, National Insttute on Aging, NIH
82 Department of Computational Biology, University of Lausanne, Lausanne, Switzerland
83 Swiss Institute of Bioinformatics, Lausanne, Switzerland
84 University Center for Primary Care and Public Health, Lausanne, Switzerland
85 Department of Epidemiology and Biostatistics, School of Public Health, Peking University Health Science Center, Beijing, China
86 Department of Clinical Neurosciences, Univeristy of Cambridge
87 Institute of Human Genetics, Technical University of Munich, 81675 Munich, Germany
88 Institute of Human Genetics, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany
89 Department of Medicine, University of Maryland School of Medicine, Baltimore, MD USA
90 Geriatrics Research and Education Clinical Center, Baltimore Veterans Administration Medical Center, Baltimore, MD
91 Institute of Clinical Epidemiology and Biometry, University of Würzburg
92 Department of Neurology, University Hospital Würzburg
93 The MIND Center, University of Mississippi Medical Center, Jackson, MS, USA
94 Department of Clinical Biochemistry, Copenhagen University Hospital - Herlev and Gentofte, Copenhagen, Denmark
95 Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark
96 College of Medicine Nursing and Health Science NUI Galway, Ireland
97 Laboratory for Systems Genetics, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan
98 Laboratory of Statistical Immunology, Immunology Frontier Research Center (WPI-IFReC), Osaka University, Suita 565-0871, Japan
99 Integrated Frontier Research for Medical Science Division, Institute for Open and Transdisciplinary Research Initiatives, Osaka University, Suita 565-0871, Japan
100 Center for Infectious Disease Education and Research (CiDER), Osaka University, Suita 565-0871, Japan
101 Department of Health Research Methods, Evidence, and Impact, McMaster University, 1280 Main Street West, Hamilton, L8S 4K1, Ontario, Canada

102 Institute of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany
103 Institute for Medical Information Processing, Biometry and Epidemiology, Ludwig Maximilians University Munich, Munich, Germany
104 German Centre for Cardiovascular Research, DZHK, Partner Site Munich, Munich, Germany
105 Department of Epidemiology, University of Washington, Seattle, WA, USA
106 Department of Health Services, University of Washington, Seattle, WA, USA
107 Center for Genomic Medicine, MGH, Boston, MA, USA. Department of Neurology, MGH, Boston, MA, USA
108 Miller Schoool Of Medcine, Department of Neurology
109 Evelin F. McKnight Brain Institute
110 Division of Cardiology, Department of Medicine, Columbia University, New York, NY, USA
111 Stroke Unit, Department of Neurology, Oslo University Hospital, Oslo, Norway
112 Research and Development, The Norwegian Air Ambulance Foundation, Norway
113 Glenn Biggs Institute for Alzheimer's and Neurodegenerative Diseases, University of Texas Health Science Center at San Antonio
114 Framingham Heart Study, Framingham, MA
115 University Medicine Greifswald, Institute for Community Medicine, SHIP/KEF
116 Seattle Epidemiologic Research and Information Center, Department of Veterans Affairs Office of Research and Development, Seattle, WA
117 Department of Neurology, Helsinki University Hospital and University of Helsinki
118 Department of Internal Medicine, Section of Gerontology and geriatrics, Leiden University Medical Center, Leiden, the Netherlands
119 Department of Cardiology, Leiden University Medical Center, Leiden, the Netherlands
120 Division Heart & Lungs, Department of Cardiology, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands
121 Department of Neurology and Neurosurgery, University of Tartu, Estonia
122 University of Cincinnati College of Medicine
123 Department of Biostatistics, Boston University School of Public Health, Boston, MA
124 McGill Genome Centre, Montreal, QC Canada
125 A list of members and affiliations appears in the Supplementary Note

126 Research Unit Molecular Epidemiology, Institute of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany
127 Department of Psychiatry and Psychotherapy, University Medicine Greifswald, Germany
128 German Center for Neurodegenerative Diseases (DZNE), Site Rostock/ Greifswald, Rostock, Germany
129 Netherlands Heart Institute, Utrecht, the Netherlands
130 Einthoven Laboratory for Experimental Vascular Medicine, LUMC, Leiden, the Netherlands
131 Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, The Netherlands
132 Department of Neurology, Boston University School of Medicine, Boston, MA
133 Department of Neurology, University of Washington, Seattle, Washington, USA
134 Department of Clinical Genetics, Department of Radiology and Nuclear Medicine
135 Department of Epidemiology and Population Health, Albert Einstein College of Medicine, New York, NY
136 Department of Clinical Biochemistry and Pharmacology, Odense University Hospital, Odense, Denmark
137 Department of Epidemiology, Erasmus University Medical Center
138 The Institute for Translational Genomics and Population Sciences, Department of Pediatrics, The Lundquist Institute for Biomedical Innovation at Harbor-UCLA Medical Center
139 Longitudinal Studies Section, Translational Gerontology Branch
140 Brown Foundation Institute of Molecular Medicine, McGovern Medical School, University of Texas Health Science Center at Houston, Houston, TX, USA
141 Human Genetics Center, School of Public Health, University of Texas Health Science Center at Houston, Houston, TX, USA
142 Geisinger Neuroscience Institute, Geisinger Health System, Danville
143 Department of Neurology, College of Medicine, The Pennsylvania State University
144 Klinik für Neurologie, Carl von Ossietzky University of Oldenburg
145 German Center for Neurodegenerative Disease DZNE, partner site Berlin, Germany
146 Department of Neurology, Universitätsklinikum Würzburg, Germany
147 Comprehensive Heart Failure Center, University Hospital Würzburg
148 Center for Clinical Trials, University Hospital Würzburg
149 Institute for Health Promotion, Graduate School of Public Health, Yonsei University, Seoul, Republic of Korea.

150 Munich Cluster for Systems Neurology, Munich 81377, Germany
151 German Center for Neurodegenerative Diseases (DZNE), Munich 81377, Germany
152 CHU de Bordeaux, Department of Neurology, F-33000 Bordeaux, France
153 These authors contributed equally to this work
154 These authors jointly supervised this work

* Corresponding authors:
Stéphanie Debette, Bordeaux Population Health research center, Inserm U1219, and Department of Neurology, Bordeaux University Hospital; University of Bordeaux, 146 rue Léo Saignat, 33076 Bordeaux Cedex, France. Tel: +33 5 57 57 16 59; Fax: +33 5 47 30 42 09. E-mail: stephanie.debette@u-bordeaux.fr
&
Martin Dichgans, LMU Klinikum, Institut für Schlaganfall- und Demenzforschung (ISD), Campus Großhadern | Feodor-Lynen-Straße 17 | 81377 München. Tel. +49 (0)89 4400-46019 | Fax +49 (0)89 4400-46010. E-Mail: Martin.Dichgans@med.uni-muenchen.de

Summary

Previous genome-wide association studies (GWAS) of stroke, the second leading cause of death, have been conducted in populations of predominantly European ancestry.[1,2] We undertook cross-ancestry GWAS meta-analyses of stroke and its subtypes in 110,182 stroke patients (33% non-European) and 1,503,898 control individuals of five ancestries from population- and clinic-based studies, nearly doubling the number of cases in previous stroke GWAS. We identified association signals at 89 independent loci, of which 61 were novel. Effect sizes were overall highly correlated across ancestries. Cross-ancestry fine-mapping, *in silico* mutagenesis analysis using a novel machine-learning approach,[3] transcriptome and proteome-wide association analyses revealed putative causal genes (e.g. *SH3PXD2A* and *FURIN*) and variants (e.g. at *GRK5* and *NOS3*). Using a novel three-pronged approach,[4] we provided genetic evidence for putative drug effects, highlighting *F11*, *KLKB1*, *PROC*, *GP1BA*, and *VCAM1* as possible targets, with drugs already under investigation for stroke for *F11* and *PROC*. A polygenic score integrating cross-ancestry and ancestry-specific stroke GWAS with vascular risk factor GWAS (iPGS) showed strong prediction of ischemic stroke risk in European and, for the first time, East-Asian populations.[5,6] The iPGS performed better than stroke PGS alone and better than previous best iPGS, in Europeans and East-Asians. Transferability of European-specific iPGS to East-Asians was limited. Stroke genetic risk scores were predictive of ischemic stroke independent of clinical risk factors in 52,600 clinical trial participants with cardiometabolic disease and performed considerably better than previous scores, both in Europeans and East-Asians. Altogether our results provide critical insight to inform biology, reveal potential drug targets for intervention, and provide genetic risk prediction tools across ancestries for targeted prevention.

Introduction

Stroke is the second leading cause of death worldwide, responsible for approximately 12% of total deaths, with an increasing burden particularly in low-income countries.[7] Characterized by a neurological deficit of sudden onset, stroke is predominantly caused by cerebral ischemia (of which the main etiological subtypes are large-artery atherosclerotic stroke [LAS], cardioembolic stroke [CES], and small-vessel stroke [SVS]) and, less often, by intracerebral hemorrhage (ICH). The frequency of stroke subtypes differs between ancestry groups as exemplified by a higher prevalence of SVS and ICH in Asian and African compared with European populations. Most genetic loci associated with stroke have been identified in populations of European ancestry. The largest published genome-wide association study (GWAS) meta-analysis to date (67,162 cases and 454,450 controls, MEGASTROKE) reported 32 stroke risk loci.[1,8] To identify new genetic associations and provide insight into stroke pathogenesis and putative drug targets, we first performed cross-ancestry GWAS on 1,614,080 participants including 110,182 stroke patients. We then characterized identified stroke risk loci by leveraging expression and protein quantitative trait loci, cross-ancestry fine-mapping, and shared genetic variation with other traits. Finally, we used a series of approaches for genomics-driven drug discovery for stroke prevention and treatment, and explored the prediction of stroke with polygenic scores across ancestries in the setting of both population-based studies and clinical trials.

Results

Genetic discovery from association analyses

We performed a fixed-effect inverse-variance weighted (IVW) GWAS meta-analysis on 29 population-based cohorts or biobanks with incident stroke ascertainment and 25 clinic-based case-control studies, comprising up to 110,182 stroke patients and 1,503,898 controls (of which 45.5% in longitudinal cohorts or biobanks), nearly doubling the number of cases in previous stroke GWAS (the GIGASTROKE initiative, **Supplementary Table 1**, **Extended**

Data Fig. 1). Genome-wide genotyping and imputation characteristics are described in **Supplementary Table 2**. The cohorts included individuals of European (EUR, 66.7% of stroke patients), East-Asian (EAS, 24.8%), African-American (AFR, 3.7%), South-Asian (SAS, 3.3%), and Hispanic (HIS, 1.4%) ancestry. Analyses were performed for any stroke (AS: comprising ischemic stroke, ICH, and stroke of unknown or undetermined type), any ischemic stroke regardless of subtype (AIS, N=86,668), and ischemic stroke subtypes (LAS, N=9,219; CES, N=12,790; SVS, N=13,620). We also conducted separate GWAS of incident AS and AIS (N=32,903 and 16,863) in longitudinal population-based cohort studies.

We tested up to ~7,588,359 single nucleotide polymorphisms (SNPs) with minor allele frequency (MAF) ≥0.01 for association with stroke. The LD score intercepts for our ancestry-specific GWAS meta-analyses ranged between 0.91 and 1.12, suggesting no systematic inflation of association statistics (**Supplementary Table 3**). We identified variants associated with stroke at genome-wide significance ($p<5\times10^{-8}$) at 60 loci, of which 33 were novel (**Fig. 1**, **Supplementary Table 4**). Lead variants at all novel loci were common (MAF≥0.05), except for low-frequency intronic variants in *THAP5* (MAF=0.02, in complete association [$r^2=1$] with variants in the 5'UTR of *NRCAM*) associated with cross-ancestry incident AS/AIS, and in *COBL* (MAF=0.04) associated with AS/AIS in South-Asians. Using conditional and joint analysis (GCTA-COJO),[9] we confirmed three independent signals at *PITX2* and two at *SH3PXD2A* (CES in EUR, **Supplementary Table 5**).[1] Cross-ancestry gene-based association analyses using VEGAS[10] revealed 158 gene-wide significant associations ($p<2.63\times10^{-6}$) in 34 loci, of which 7 were in 4 novel loci not reaching genome-wide significance in the single-variant analyses (*AGAP5/SYNPO2L/SEC24C/CHCHD1, USP34, USP38,* and *MAMSTR*, **Supplementary Table 6-7**). Next, we conducted a cross-ancestry meta-analysis with MR-MEGA,[11] which accounts for the allelic heterogeneity between ancestries. We identified three additional genome-wide significant loci for AS (all novel), near *TSPAN19*, and in introns of *DAZL* and *SHOC1*, all showing high heterogeneity in allelic effects across ancestries (Heterogeneity P-value<0.01, **Supplementary Table 8**).

Overall, the largest number of genome-wide significant associations was identified for AS (50 loci, 27 novel [27]) and AIS (45 loci, [19]), of which one with incident AIS only. While AIS subtypes were not available in some population-based cohorts (**Supplementary Table 1**), genome-wide significance was reached for 3 loci ([1]) for LAS, 7 ([5]) for CES, and 7 ([2]) for SVS (**Supplementary Table 4**). To further enhance statistical power for AIS subtypes, we conducted multi-trait analyses of GWAS (MTAG)[12] in Europeans and East-Asians, including

traits correlated with specific stroke subtypes, namely: (i) coronary artery disease (CAD) for LAS, both caused by atheroma (ii) atrial fibrillation (AF) for CES, as its main underlying cause, and (iii) white matter hyperintensity volume (WMH, an MRI-marker of cerebral small vessel disease) for SVS (available in Europeans only). In Europeans, 11 [10] additional loci were associated with LAS (10 novel), 3 with SVS (all reported in a recent SVS GWAS[2]), and 5 with CES (all novel, **Supplementary Tables 9-11**). Moreover, 18 and 15 additional genome-wide significant associations were identified for AS and AIS, respectively (all novel) using MTAG with WMH, CAD, and AF (**Supplementary Tables 12-13**). In East-Asians, one locus was associated with AS (*FGF5*) and one with LAS (*HDAC9*, novel in EAS) using MTAG. This brings the number of identified stroke risk loci to 89 [61] in total, of which 68 [45] associated with AS, 50 [35] with AIS, 14 [11] with LAS, 12 [10] with CES, and 10 [2] with SVS (**Fig. 1, Supplementary Table 4, 8, and 9-14**).

Comparing effects across ancestries and cross-ancestry fine-mapping

---

To our knowledge, our results include the most comprehensive and largest description of stroke genetic risk variants to date in each of the five represented ancestries. In cross-ancestry meta-analyses (IVW and MR-MEGA) 56 loci reached genome-wide significance, while 39 loci were genome-wide significant in Europeans, 6 in East-Asians (4 shared with Europeans), 1 in South-Asians, and 2 in African-Americans (at *3p21* and *PTCH1* [SVS], **Supplementary Table 4**).

For the 60 stroke risk loci derived from the IVW meta-analyses we compared the per-allele effect size across the three ancestries with the largest sample size (EUR, EAS, AFR). Correlations of per-allele effect sizes of index variants varied from r=0.55 (EUR with AFR) to 0.66 (EUR with EAS) and 0.74 (EAS with AFR, **Fig. 2a**).

To identify putative causal variants at stroke risk loci identified through IVW meta-analyses, we performed multiple-causal-variant fine-mapping using SuSiE,[13] separately in Europeans and East-Asians (**Methods**). Across stroke types we identified 110 and 16 95% credible set (CS)-trait pairs in EUR and EAS respectively, each of which having a 95% posterior probability of containing a causal variant, with multiple CS identified at 6 (EUR) and one (EAS) stroke risk loci (**Supplementary Tables 15-17).** Within the CS identified in EUR, 17 variants were found to have a posterior inclusion probability (PIP) > 0.9. We found overlapping CS between Europeans and East-Asians at *SH3PXD2A* (19 overlapping variants), suggesting cross-ancestry shared genetic architecture at this locus (**Fig. 2b**). Two loci had CS

with a single variant (rs10886430 at *GRK5* [PIP= 0.999], associated with *GRK5* platelet gene expression and thrombin-induced platelet aggregation,[14] and rs1549758 at *NOS3*, PIP= 0.995), likely representing strong targets for functional validation.

Although there were six nonsynonymous variants among CS (rs671 [*ALDH2*], rs8071623 [*SEPT4*], rs35212307 [*WDR12*], rs72932557 [*CARF*], rs11906160 [*MYH7B*], and rs2501968 [*CENPQ*]), exonic variants for coding RNA within CS were few (1.2%). To detect putative causal regulatory variants, we conducted *in silico* mutagenesis analysis using MENTR, a machine-learning method to pin-point prediction of causal variants on transcriptional changes.[3] From CS, we obtained 78 robust predictions of variant-transcript-model sets comprising 13 variants and 19 transcripts (**Supplementary Table 18**). In particular, rs12476527 (5'UTR of *KCNK3*, also a blood pressure locus[15]) was predicted to increase *KCNK3* expression in kidney cortex tubule cells, despite no eQTL of this variant being reported in GTEx (v8) or eQTLgen (2019-12-23). Furthermore, three variants (rs12705390 at *PIK3CG*, rs2282978 at *CDK6,* rs2483262 at *PRDM16*) were predicted to affect expression of a long non-coding RNA and enhancer RNAs, in endothelial cells, umbilical vein, and visceral preadipocytes respectively.

Characterization of stroke-associated loci

VEGAS2Pathway[16] analysis revealed significant enrichment ($P<5.01\times10^{-6}$) of stroke risk loci in pathways involved in (i) carboxylation of amino-terminal glutamate residues required for activation of proteins involved in blood clot formation and regulation, (ii) negative regulation of coagulation, and (ii) angiopoietin receptor Tie2-mediated signaling, involved in angiogenesis (**Supplementary Table 19**).

We explored shared genetic variation with 12 (in Europeans) and 6 (in East-Asians) vascular risk factor and disease traits (**Methods, Supplementary Methods**). In Europeans, the lead variants for stroke at 57 of the 88 risk loci (64.8%) were associated ($P<5\times10^{-8}$) with at least one vascular trait, most frequently blood pressure (34 loci, 38.6%, **Extended Data Figure 2, Supplementary Table 20)**. Following correction for multiple testing (**Methods**, $p<4.17\times10^{-3}$) all vascular risk traits except LDL-cholesterol showed significant genetic correlation with at least one stroke type, the strongest correlations being for CAD and LAS (r_g=0.73), AF and CES (r_g=0.63), and SBP with all stroke types (r_g ranging from 0.21 for CES to 0.49 for LAS and SVS, **Extended Data Fig. 3, Supplementary Table 21**) Using two-sample Mendelian randomization (MR) we found evidence for a causal association for every vascular risk trait

except triglycerides with at least one stroke type ($p<4.17x10^{-3}$), with some subtype-specific association patterns. Genetically predicted WMH was associated with increased risk of SVS but not other stroke subtypes, while genetically predicted venous thromboembolism (VTE) was associated with AS, AIS, CES, and LAS, but not SVS (**Extended Data Fig. 3, Supplementary Table 22**). In East-Asians, SBP, DBP, and BMI showed significant genetic correlation with any stroke (r_g=0.45, 0.39 and 0.24 vs. r_g=0.36, 0.21, and 0.22 in Europeans), with evidence for a causal association of SBP and DBP with AS, AIS, and SVS (**Extended Data Fig. 4**, **Supplementary Tables 21-22**).

Next, to generate hypotheses of target genes and directions of effect, we conducted transcriptome-wide association studies using TWAS-Fusion[17] and expression quantitative trait loci (eQTL) based on RNA sequencing in different tissues.[18-21] We identified 27 genes whose genetically regulated expression associated with stroke and its subtypes at the transcriptome-wide level and colocalized in at least one tissue (10 genes in arteries and heart; 6 genes in brain tissue; 17 genes across tissues), of which 18 overlapped with 11 genome-wide significant stroke risk loci (**Extended Data Fig. 5, Supplementary Table 23)**. For loci where bulk tissue expression levels of several genes showed evidence for association with stroke, human single-cell sequencing data of vascular-related brain cells in the dorsolateral prefrontal cortex (dPFC) showed distinct cell-specific gene expression patterns suggesting that multiple genes could be involved via different cell types (**Extended Data Fig. 6**). Further, using proteome-wide association studies (PWAS) in dPFC brain tissue we found evidence for association of ICA1L with AS and AIS through its cis-regulated protein abundance, with colocalization evidence (**Extended Data Fig. 7, Supplementary Table 24**). In both TWAS and PWAS, lower *ICA1L* transcript or protein abundance in the dPFC was associated with higher risk of stroke.

Genomics-driven drug discovery

We used a three-pronged approach for genomics-driven discovery of drugs for prevention or treatment of stroke (**Methods**, **Fig. 3**).[4] First, using GREP[22] we observed significant enrichment of stroke-associated genes (MAGMA[23] false discovery rates [FDR] <0.05) in drug-target genes for blood and blood-forming organs (Anatomical Therapeutic Chemical Classification System [ATC] B drugs, for AS, AIS, and CES). This encompasses the previously described *PDE3A* and *FGA* genes,[24] encoding targets for cilostazol (antiplatelet agent) and alteplase (thrombolytic drug), respectively, as well as *F11*, *KLKB1*, and *MUT*

encoding targets for conestat alfa, ecallantide (both used for hereditary angioedema) and vitamin B12, respectively (**Supplementary Table 25**). Second, we used Trans-Phar[25] to test the negative correlations between genetically determined case-control gene expression associated with stroke (TWAS using all GTEX v7 tissues[18]) and compound-regulated gene expression profiles. We observed significant negative correlations for BRD.A22514244 (for SVS; drug target unknown) and GR.32191 (for CES, **Supplementary Table 26**). GR-32191 is a Thromboxane A2 receptor antagonist proposed as an alternative antiplatelet therapy for stroke prevention,[26] and further drugs of this class are under development.[27] We note that one of those drugs, Terutroban, was evaluated in a Phase III study but failed to show non-inferiority against Aspirin.[28] Third, we used protein quantitative trait loci (pQTL) for 218 drug-target proteins as instruments for MR and found evidence for causal associations of 9 plasma proteins with stroke risk (4 cis-pQTL, 6 trans-pQTL), of which 6 were supported by colocalization analyses, with no evidence for reverse causation using the Steiger test (PROC, VCAM1, F11, KLKB1, MMP12, and GP1BA, **Supplementary Table 27**). Using public drug databases we curated drugs targeting those proteins in a direction compatible with a beneficial therapeutic effect against stroke based on MR estimates: such drugs were identified for PROC, VCAM1, F11, KLKB1, and GP1BA (**Supplementary Table 28**). Drugs targeting F11 (NCT04755283, NCT04304508, NCT03766581) and PROC (NCT02222714) are currently under investigation for stroke, and our results provided genetic support for this. Of note, *F11* and *KLKB1* are adjacent genes with a long range linkage disequilibrium pattern and complex co-regulation,[29] as illustrated here by the presence of a shared trans-pQTL in *KNG1* (**Supplementary Table 27**). Additional studies are needed to disentangle causal associations and the most appropriate drug target in this region.[30,31] To further validate the candidate drugs and estimate their potential side effects, we investigated whether the drug-target genes were associated with stroke-related phenotypes using a phenome-wide association study (PheWAS) approach.[32] We conducted PheWAS in Estonian Biobank (EstBB) for the pQTL variants and rare deleterious variants in *PROC*, *VCAM1*, *F11*, *KLKB1*, and *GP1BA* genes (**Supplementary Table 29**). Rs2289252, a cis-pQTL for F11, was associated with higher risk of venous thromboembolic disorders ($p<5.37\times10^{-6}$), as previously described,[33] and showed suggestive association ($p=4.23\times10^{-3}$) with cerebral infarction (I63, **Extended Data Fig. 8**). Conversely, we observed no significant association with non-stroke-related phenotypes, suggesting the safety of targeting F11. Similar profiles were observed in UK Biobank and FinnGen (https://r5.finngen.fi/variant/4-186286227-C-T), with no significant associations with other disorders and no overlap of subthreshold signals with side-effects reported in clinical trials.[34]

Overall, combining evidence from genomics-driven drug discovery approaches, characterization of stroke risk loci, and prior knowledge from monogenic disease models and experimental data, we found evidence for potential functional implication of 47 genes to be prioritized for further functional follow-up, with evidence from multiple approaches for 17 genes (**Supplementary Table 30**).

Polygenic risk prediction in the population

We explored the risk prediction potential of stroke GWAS, alone and in combination with vascular risk trait GWAS, in Europeans and East-Asians, using ancestry-specific polygenic scores (PGS). PGS were based on ancestry-specific and cross-ancestry GWAS summary statistics. We first derived single PGS (sPGS) models from single stroke GWAS summary data (**Supplementary Table 31**). We then constructed integrative PGS (iPGS) models, which combined multiple GWAS summary data of different traits into a PGS using elastic-net logistic regression (**Extended Data Fig. 9-10**).[6] The iPGS analysis used two datasets for each ancestry for model training and evaluation, respectively. Participants in the training and evaluation datasets did not overlap and were not included in the input GWAS summary data. For Europeans, we constructed the iPGS model using 1,003 prevalent AIS cases and 8,997 controls, followed by evaluation of the model using 1,128 incident AIS cases among 102,099 participants, all from EstBB. The improvement in predictive ability (ΔC-index) was assessed over a base model including age, sex, and the top 5 principal components for population stratification (PCs). The iPGS model for Europeans incorporated 10 GIGASTROKE GWAS (all stroke types, using the European and cross-ancestry analysis) and 14 vascular risk trait GWAS (**Extended Data Fig. 9**, **Supplementary Table 32**). The iPGS model achieved a ΔC-index of 0.022 (**Figure 4a** and **Supplementary Table 33**), 58% higher than that for a previously constructed iPGS model for Europeans, derived from 5 MEGASTROKE GWAS and the same vascular risk trait GWAS (ΔC-index=0.014).[6] The age-, sex-, and top 5 PC-adjusted hazard ratio (HR) per standard deviation (SD) of the PGS was 1.25 (95% confidence interval [CI], 1.18–1.32; $P=8.2\times10^{-14}$) for the GIGASTROKE-based iPGS model compared to 1.19 (95%CI, 1.12–1.26; $P=4.2\times10^{-9}$) for the MEGASTROKE-based iPGS model (**Fig. 4a**). For East-Asians, we derived the iPGS model using 577 prevalent AIS cases and 9,232 controls, and evaluated the model using 1,470 prevalent AIS cases and 40,459 controls, from Biobank Japan (BBJ). A base model including age, sex, and top 5 PCs showed an area under the curve (AUC) of 0.634. The iPGS model was constructed by integrating 10

GIGASTROKE GWAS and 37 vascular risk trait GWAS (**Extended Data Fig. 10**, **Supplementary Table 34**). The iPGS model for East-Asians showed an improvement in AUC (ΔAUC) of 0.020 (**Figure 4a** and **Supplementary Table 35**). The age-, sex-, and top 5 PC-adjusted odds ratio (OR) per SD of PGS was 1.33 (95%CI, 1.26–1.40; $P=2.3\times10^{-26}$) for the iPGS model. The MEGASTROKE- and GIGASTROKE-based iPGS models for Europeans achieved lower AUC improvement (ΔAUC=0.007 and 0.014, respectively) than the GIGASTROKE-based iPGS model for East-Asians. While this suggests that the transferability of iPGS models for Europeans to East-Asians might be limited (**Supplementary Table 35**), it does indicate that an ancestry-specific stroke iPGS approach yields similar improvement in predictive ability relative to their base models.
For Europeans (**Figure 4b** and **Supplementary Table 36**), compared to those in the middle 10% (45–55%) of the GIGASTROKE-based iPGS, those in the top 1% showed a >2.6-fold higher hazard of ischemic stroke (HR=2.61 [95%CI, 1.72–3.96]; $P=1.1\times10^{-6}$), while those in the top 0.1% showed a >3.6-fold higher risk (HR=3.65 [95%CI, 1.28–10.38]; P=0.02). For East-Asians (**Figure 4c** and **Supplementary Table 37**), those in the top 1% of the iPGS showed >2.1-fold higher odds of ischemic stroke (OR=2.11 [95% CI, 1.37–3.25]; $P=6.7\times10^{-4}$) and the risk of those in the top 0.1% was >3.1-fold higher (OR=3.11 [95% CI, 1.08–8.92]; P=0.04) than the middle 10%. Although caution is warranted when interpreting risk estimates in the highest PGS groups due to wide confidence intervals, these results indicate that GIGASTROKE-based iPGS models may be useful to stratify individuals exposed to genetically high risk of ischemic stroke, not only for Europeans but also for East-Asians.

Risk prediction in a clinical trial setting

We further explored whether a genetic risk score (GRS) based on genome-wide significant risk loci from the cross-ancestry IVW any stroke (AS) meta-analyses could identify individuals at higher risk for AIS after accounting for established risk factors in 5 clinical trials[35-39] across the spectrum of cardiometabolic disease. The primary analysis was conducted in 51,288 European participants of whom 960 developed an incident ischemic stroke (AIS) over 3 years follow-up. In a Cox model adjusted for age, sex, and vascular risk factors (**Methods**), a higher GIGASTROKE GRS was significantly associated with increased risk for AIS in Europeans (adjusted hazard ratio [HR] of 1.17 [95%CI, 1.09-1.24] per standard deviation [SD] increase, $P=2x10^{-6}$, **Supplementary Table 38**). This association was substantially stronger than the association with the earlier MEGASTROKE GRS based on 32

genome-wide significant stroke risk loci (HR=1.07 [1.00-1.14], P=0.036).[1,40] Compared with patients in the lowest GIGASTROKE GRS tertile, patients in the top GRS tertile had an adjusted HR of 1.35 (1.16-1.58) for developing AIS whereas those in the middle tertile had an adjusted HR of 1.13 (0.96-1.33, $P_{trend}=1.4x10^{-4}$, **Fig. 4**). The performance of the GRS was stronger in individuals without previous stroke (N=44,095; adjusted HR of top versus lowest tertile, 1.37 [1.14-1.65]) than in those with a previous stroke (N=7,193; adjusted HR, 1.15 [0.87-1.54]). Similar associations were observed when using effect estimates from stroke GWAS meta-analyses in Europeans or for AIS (**Supplementary Table 38**). In secondary analyses we explored the association of the GIGASTROKE cross-ancestry AS GRS with incident AIS in the much smaller East-Asian sample (1,312 participants of whom 27 developed an incident stroke over 3 years follow-up), and found consistent associations (HR=1.49 [1.00-2.21] per SD increase, P=0.048, **Supplementary Table 38**), while the MEGASTROKE GRS was not associated with incident AIS in East-Asians (HR=0.82 [0.55-1.23], P=0.34).

Discussion

Our GWAS meta-analyses gathering over 110,000 stroke patients from five different ancestries identified 61 novel risk loci for stroke and stroke subtypes and suggest substantial shared susceptibility to stroke across ancestries, with strong correlation of effect sizes. Effect estimates for variants that were common across ancestries were typically similar, while, expectedly, variants that were rare or low frequency (MAF≤0.05) in one or more populations showed differences in effect size, e.g. at *PROCR*, *TAP1*, or *BNCZ-CNTLN* (MAF≤0.05 in EAS), or at *GRK5*, *FOXF2*, or *COBL* (MAF≤0.05 in AFR). Ancestry-specific meta-analyses in non-European populations detected fewer loci than in Europeans (likely due to smaller sample sizes), which were nevertheless biologically plausible, e.g. *3p21* and *PTCH1 f*or SVS in AFR. Rare variants at *3p21* were recently shown to be associated with WMH volume,[41] whereas common variants at *PTCH1* were associated with functional outcome after ischemic stroke (in EUR).[42] Novel association signals from cross-ancestry GWAS included for instance variants at *PROCR, GRK5* and *F11* (thrombosis), *LPA* and *ATP2B1* (lipid metabolism, hypertension, and atherosclerosis), *SWAP70* (membrane ruffling), and *LAMC1* (cerebrovascular matrisome).

Extensive bioinformatics analyses highlight genes for prioritization in further functional follow-up (**Supplementary Table 30**). For example, a promoter variant of *SH3PXD2A*, encoding an adaptor protein involved in extracellular matrix degradation via invadopodia and podosome formation,[43] was predicted to modulate its expression in macrophages. As another example, *FURIN* expression levels across tissues were associated with increased stroke risk. *FURIN* is expressed in brain endothelial cells,[44] has previously been implicated in coronary artery disease,[45] and FURIN inhibition reduces vascular remodeling and atherosclerotic lesion progression in mice.[46] FURIN also plays a key role in SARS-CoV-2 infectivity,[47] and COVID-19 patients are at increased risk of AIS, especially LAS.[48] The *FURIN* locus was predominantly associated with LAS in our data (**Supplementary Table 39**).

Our results provide genetic evidence for putative drug effects using three independent approaches, with converging results from two methods (gene enrichment analysis and pQTL-based MR) for drugs targeting F11 and KLKB1. F11 and F11a inhibitors (e.g. abelacimab, BAY 2433334, BMS-986177) are currently explored in phase-2 trials for primary or secondary stroke prevention (NCT04755283, NCT04304508, NCT03766581). Additional

evidence from pQTL-based MR suggested PROC, GP1BA, and VCAM1 as potential drug targets for stroke. A recombinant variant of human activated protein C (encoded by *PROC*) was found to be safe for the treatment of acute ischemic stroke following thrombolysis, mechanical thrombectomy or both in phase 1 and 2 trials (3K3A-APC, NCT02222714),[49,50] and is poised for an upcoming phase 3 trial. 3K3A-APC is proposed as a neuroprotectant, with evidence for protection of white matter tracts and oligodendrocytes from ischemic injury in mice.[51] Anfibatide, a GPIbα antagonist, reduced blood-brain barrier disruption following ischemic stroke in mice[52] and is being tested as an antiplatelet drug in myocardial infarction (NCT01585259). While specific VCAM1 inhibitors are not available, probucol, a lipid lowering drug with pleiotropic effects including VCAM1 inhibition was tested for secondary prevention of atherosclerotic events in CAD patients (PROSPECTIVE, UMIN000003307).[53]

We improved polygenic risk prediction of stroke and importantly pioneered the exploration of stroke PGS across ancestries. Polygenic scores integrating cross-ancestry and ancestry-specific stroke GWAS with vascular risk factor GWAS (iPGS) showed strong prediction of ischemic stroke risk in European and, importantly, for the first time, in East-Asians where stroke incidence is highest.[7] The iPGS performed better than stroke PGS alone and better than previous best iPGS in Europeans.[6] We obtained similar improvement in predictive ability of ancestry-specific and cross-ancestry iPGS relative to base models in Europeans and East-Asians, whereas, in contrast to the approach we develop, transferability of European-specific iPGS to East-Asians was limited. Individuals in the top 0.1% of the PGS distribution had a more than 3-fold risk of ischemic stroke in both EUR and EAS compared to those in the middle 10%. Our results indicate that GIGASTROKE-based iPGS models may be useful to stratify individuals exposed to genetically high risk of ischemic stroke. They highlight the importance of ancestry-specific and cross-ancestry genomic studies for the transferability of genomic risk prediction across populations, and the urgent need to vastly increase the diversity of participants in genomic studies to avoid exacerbation of health disparities in the era of precision medicine and precision public health.[54-56]

Finally, leveraging data from 5 clinical trials in 52,600 patients with established cardiometabolic disease, we showed that a cross-ancestry genetic risk score predicted ischemic stroke, independently of the presence of clinical risk factors while outperforming previous genetic risk evaluation.[5,40] Notably, although the trials included predominantly European participants, consistent results were observed, also for the first time, in participants with East-Asian ancestry.

While non-European ancestry-specific stroke GWAS are limited by sample availability, our study includes by far the largest contribution of non-European stroke genetics resources (N>310,000 for the GWAS and >55,000 for the PGS/GRS studies). Although the lack of suitable additional datasets precludes direct replication efforts, we provide validation of our findings in independent population studies and major clinical trials. The muted risk prediction in participants with previous stroke history possibly points to the impact of selection or index event biases and secondary prevention therapy.[57]

In conclusion, our results provide critical insight to inform future biological research into the pathogenesis of stroke and its subtypes, highlight potential drug targets for intervention, and provide genetic risk prediction tools across ancestries for targeted prevention.[58,59]

References

1 Malik, R. *et al.* Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nature Genetics* **50**, 524-537, doi:10.1038/s41588-018-0058-3 (2018).

2 Traylor, M. *et al.* Genetic basis of lacunar stroke: a pooled analysis of individual patient data and genome-wide association studies. *Lancet Neurol*, doi:10.1016/s1474-4422(21)00031-4 (2021).

3 Koido, M. *et al.* Predicting cell-type-specific non-coding RNA transcription from genome sequence. (bioRxiv 2020.03.29.011205; doi: https://doi.org/10.1101/2020.03.29.011205).

4 Namba, S. *et al.* A practical guideline of genomics-driven drug discovery in the era of global biobank meta-analysis. *medRxiv* doi:https://doi.org/10.1101/2021.12.03.21267280 (2021.12.03.21267280).

5 Rutten-Jacobs, L. C. *et al.* Genetic risk, incident stroke, and the benefits of adhering to a healthy lifestyle: cohort study of 306 473 UK Biobank participants. *BMJ* **363**, k4168, doi:10.1136/bmj.k4168 (2018).

6 Abraham, G. *et al.* Genomic risk score offers predictive performance comparable to clinical risk factors for ischaemic stroke. *Nat Commun* **10**, 5819, doi:10.1038/s41467-019-13848-1 (2019).

7 Global, regional, and national burden of stroke and its risk factors, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet Neurol* **20**, 795-820, doi:10.1016/s1474-4422(21)00252-0 (2021).

8 Malik, R. *et al.* Genome-wide meta-analysis identifies 3 novel loci associated with stroke. *Annals of Neurology* **84**, 934-939, doi:10.1002/ana.25369 (2018).

9 Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet* **44**, 369-375, S361-363, doi:10.1038/ng.2213 (2012).

10 Mishra, A. & Macgregor, S. VEGAS2: Software for More Flexible Gene-Based Testing. *Twin Res Hum Genet* **18**, 86-91, doi:10.1017/thg.2014.79 (2015).

11 Magi, R. *et al.* Trans-ethnic meta-regression of genome-wide association studies accounting for ancestry increases power for discovery and improves fine-mapping resolution. *Hum Mol Genet* **26**, 3639-3650, doi:10.1093/hmg/ddx280 (2017).

12 Turley, P. *et al.* Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nat Genet* **50**, 229-237, doi:10.1038/s41588-017-0009-4 (2018).

13 Wang, G., Sarkar, A., Carbonetto, P. & Stephens, M. A simple new approach to variable selection in regression, with application to genetic fine mapping. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **82**, 1273-1300, doi:10.1111/rssb.12388 (2020).

14 Rodriguez, B. A. T. *et al.* A Platelet Function Modulator of Thrombin Activation Is Causally Linked to Cardiovascular Disease and Affects PAR4 Receptor Signaling. *Am J Hum Genet* **107**, 211-221, doi:10.1016/j.ajhg.2020.06.008 (2020).

15 Evangelou, E. *et al.* Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. *Nat Genet* **50**, 1412-1425, doi:10.1038/s41588-018-0205-x (2018).

16 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, doi:10.1017/thg.2016.100 (2017).

17 Gusev, A. *et al.* Integrative approaches for large-scale transcriptome-wide association studies. *Nat Genet* **48**, 245-252, doi:10.1038/ng.3506 (2016).

18 Consortium, G. *et al.* Genetic effects on gene expression across human tissues. *Nature* **550**, 204-213, doi:10.1038/nature24277 (2017).

19 Fromer, M. *et al.* Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nature Neuroscience* **19**, 1442-1453, doi:10.1038/nn.4399 (2016).

20 Gusev, A. *et al.* Integrative approaches for large-scale transcriptome-wide association studies. *Nature Genetics* **48**, 245-252, doi:10.1038/ng.3506 (2016).

21 Wright, F. A. *et al.* Heritability and genomics of gene expression in peripheral blood. *Nature Genetics* **46**, 430-437, doi:10.1038/ng.2951 (2014).

22 Sakaue, S. & Okada, Y. GREP: genome for REPositioning drugs. *Bioinformatics (Oxford, England)* **35**, 3821-3823, doi:10.1093/bioinformatics/btz166 (2019).

23 de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS computational biology* **11**, e1004219, doi:10.1371/journal.pcbi.1004219 (2015).

24 Malik, R. *et al.* Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nat Genet* **50**, 524-537, doi:10.1038/s41588-018-0058-3 (2018).

25 Konuma, T., Ogawa, K. & Okada, Y. Integration of genetically regulated gene expression and pharmacological library provides therapeutic drug candidates. *Human Molecular Genetics* **30**, 294-304, doi:10.1093/hmg/ddab049 (2021).

26 Chamorro, A. TP receptor antagonism: a new concept in atherothrombosis and stroke prevention. *Cerebrovasc Dis* **27 Suppl 3**, 20-27, doi:10.1159/000209262 (2009).

27 Yan, A. *et al.* Thromboxane A2 receptor antagonist SQ29548 reduces ischemic stroke-induced microglia/macrophages activation and enrichment, and ameliorates brain injury. *Sci Rep* **6**, 35885, doi:10.1038/srep35885 (2016).

28 Bousser, M. G. *et al.* Terutroban versus aspirin in patients with cerebral ischaemic events (PERFORM): a randomised, double-blind, parallel-group trial. *Lancet* **377**, 2013-2022, doi:10.1016/s0140-6736(11)60600-4 (2011).

29 Safdar, H. *et al.* Regulation of the F11, Klkb1, Cyp4v3 gene cluster in livers of metabolically challenged mice. *PLoS One* **8**, e74637, doi:10.1371/journal.pone.0074637 (2013).

30 de Haan, H. G. *et al.* Targeted sequencing to identify novel genetic risk factors for deep vein thrombosis: a study of 734 genes. *J Thromb Haemost* **16**, 2432-2441, doi:10.1111/jth.14279 (2018).

31 Rohmann, J. L. *et al.* Genetic determinants of activity and antigen levels of contact system factors. *J Thromb Haemost* **17**, 157-168, doi:10.1111/jth.14307 (2019).

32 Nguyen, P. A., Born, D. A., Deaton, A. M., Nioi, P. & Ward, L. D. Phenotypes associated with genes encoding drug targets are predictive of clinical trial side effects. *Nat Commun* **10**, 1579, doi:10.1038/s41467-019-09407-3 (2019).

33 Lindstrom, S. *et al.* Genomic and transcriptomic association studies identify 16 novel susceptibility loci for venous thromboembolism. *Blood* **134**, 1645-1657, doi:10.1182/blood.2019000435 (2019).

34 Verhamme, P. *et al.* Abelacimab for Prevention of Venous Thromboembolism. *N Engl J Med* **385**, 609-617, doi:10.1056/NEJMoa2105872 (2021).

35 Sabatine, M. S. *et al.* Evolocumab and Clinical Outcomes in Patients with Cardiovascular Disease. *N Engl J Med* **376**, 1713-1722, doi:10.1056/NEJMoa1615664 (2017).

36 Bonaca, M. P. *et al.* Long-term use of ticagrelor in patients with prior myocardial infarction. *N Engl J Med* **372**, 1791-1800, doi:10.1056/NEJMoa1500857 (2015).

37 Scirica, B. M. *et al.* Saxagliptin and cardiovascular outcomes in patients with type 2 diabetes mellitus. *N Engl J Med* **369**, 1317-1326, doi:10.1056/NEJMoa1307684 (2013).

38 O'Donoghue, M. L. *et al.* Effect of darapladib on major coronary events after an acute coronary syndrome: the SOLID-TIMI 52 randomized clinical trial. *JAMA* **312**, 1006-1015, doi:10.1001/jama.2014.11061 (2014).

39 Giugliano, R. P. *et al.* Edoxaban versus warfarin in patients with atrial fibrillation. *N Engl J Med* **369**, 2093-2104, doi:10.1056/NEJMoa1310907 (2013).

40 Marston, N. A. *et al.* Clinical Application of a Novel Genetic Risk Score for Ischemic Stroke in Patients With Cardiometabolic Disease. *Circulation* **143**, 470-478, doi:10.1161/circulationaha.120.051927 (2021).

41 Malik, R. *et al.* Whole-exome sequencing reveals a role of HTRA1 and EGFL8 in brain white matter hyperintensities. *Brain* **144**, 2670-2682, doi:10.1093/brain/awab253 (2021).

42 Söderholm, M. *et al.* Genome-wide association meta-analysis of functional outcome after ischemic stroke. *Neurology* **92**, e1271-e1283, doi:10.1212/wnl.0000000000007138 (2019).

43 Zagryazhskaya-Masson, A. *et al.* Intersection of TKS5 and FGD1/CDC42 signaling cascades directs the formation of invadopodia. *The Journal of Cell Biology* **219**, e201910132, doi:10.1083/jcb.201910132 (2020).

44 Yang, A. *et al.* A human brain vascular atlas reveals diverse cell mediators of Alzheimer's disease risk. (bioRxiv 2021.04.26.441262; doi: https://doi.org/10.1101/2021.04.26.441262).

45 Yang, X. *et al.* FURIN Expression in Vascular Endothelial Cells Is Modulated by a Coronary Artery Disease-Associated Genetic Variant and Influences Monocyte Transendothelial Migration. *J Am Heart Assoc* **9**, e014333, doi:10.1161/jaha.119.014333 (2020).

46 Yakala, G. K. *et al.* FURIN Inhibition Reduces Vascular Remodeling and Atherosclerotic Lesion Progression in Mice. *Arterioscler Thromb Vasc Biol* **39**, 387-401, doi:10.1161/atvbaha.118.311903 (2019).

47 Cantuti-Castelvetri, L. *et al.* Neuropilin-1 facilitates SARS-CoV-2 cell entry and infectivity. *Science* **370**, 856-860, doi:10.1126/science.abd2985 (2020).

48 Nannoni, S., de Groot, R., Bell, S. & Markus, H. S. Stroke in COVID-19: A systematic review and meta-analysis. *Int J Stroke* **16**, 137-149, doi:10.1177/1747493020972922 (2021).

49 Lyden, P. *et al.* Phase 1 safety, tolerability and pharmacokinetics of 3K3A-APC in healthy adult volunteers. *Curr Pharm Des* **19**, 7479-7485, doi:10.2174/13816128196661312301314 54 (2013).

50 Lyden, P. *et al.* Final Results of the RHAPSODY Trial: A Multi-Center, Phase 2 Trial Using a Continual Reassessment Method to Determine the Safety and Tolerability of 3K3A-APC, A Recombinant Variant of Human Activated Protein C, in Combination with Tissue Plasminogen Activator, Mechanical Thrombectomy or both in Moderate to Severe Acute Ischemic Stroke. *Ann Neurol* **85**, 125-136, doi:10.1002/ana.25383 (2019).

51 Huuskonen, M. T. *et al.* Protection of ischemic white matter and oligodendrocytes in mice by 3K3A-activated protein C. *J Exp Med* **219**, doi:10.1084/jem.20211372 (2022).

52 Chu, W. *et al.* Blockade of platelet glycoprotein receptor Ib ameliorates blood-brain barrier disruption following ischemic stroke via Epac pathway. *Biomed Pharmacother* **140**, 111698, doi:10.1016/j.biopha.2021.111698 (2021).

53 Yamashita, S. *et al.* Probucol Trial for Secondary Prevention of Atherosclerotic Events in Patients with Coronary Heart Disease (PROSPECTIVE). *J Atheroscler Thromb* **28**, 103-123, doi:10.5551/jat.55327 (2021).

54 Khoury, M. J. & Holt, K. E. The impact of genomics on precision public health: beyond the pandemic. *Genome Med* **13**, 67, doi:10.1186/s13073-021-00886-y (2021).

55 Ben-Eghan, C. *et al.* Don't ignore genetic data from minority populations. *Nature* **585**, 184-186, doi:10.1038/d41586-020-02547-3 (2020).

56 Martin, A. R. *et al.* Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat Genet* **51**, 584-591, doi:10.1038/s41588-019-0379-x (2019).

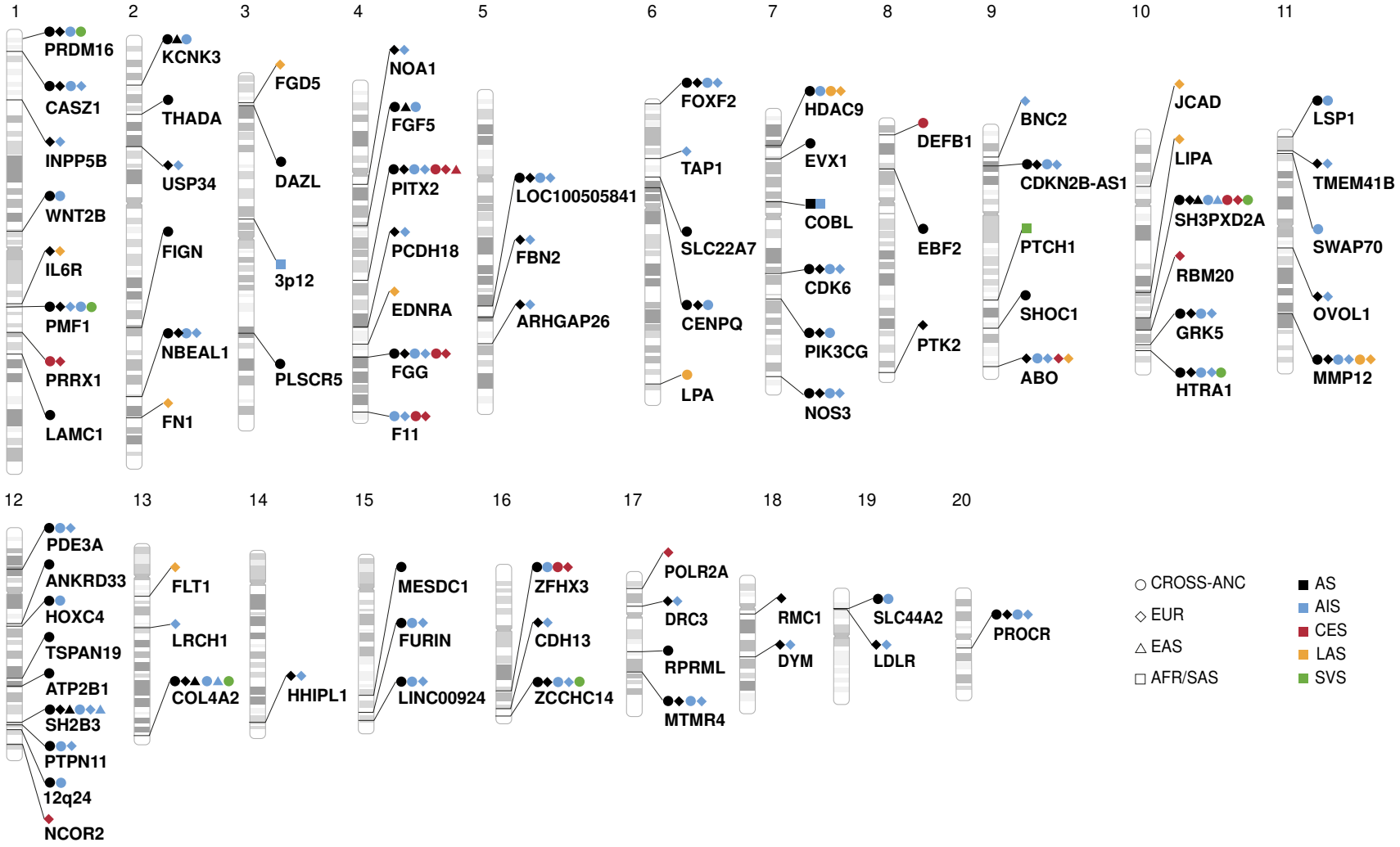
57 Dudbridge, F. *et al.* Adjustment for index event bias in genome-wide association studies of subsequent events. *Nat Commun* **10**, 1561, doi:10.1038/s41467-019-09381-w (2019).

58 Responsible use of polygenic risk scores in the clinic: potential benefits, risks and gaps. *Nat Med* **27**, 1876-1884, doi:10.1038/s41591-021-01549-6 (2021).

59 Wand, H. *et al.* Improving reporting standards for polygenic scores in risk prediction studies. *Nature* **591**, 211-219, doi:10.1038/s41586-021-03243-6 (2021).

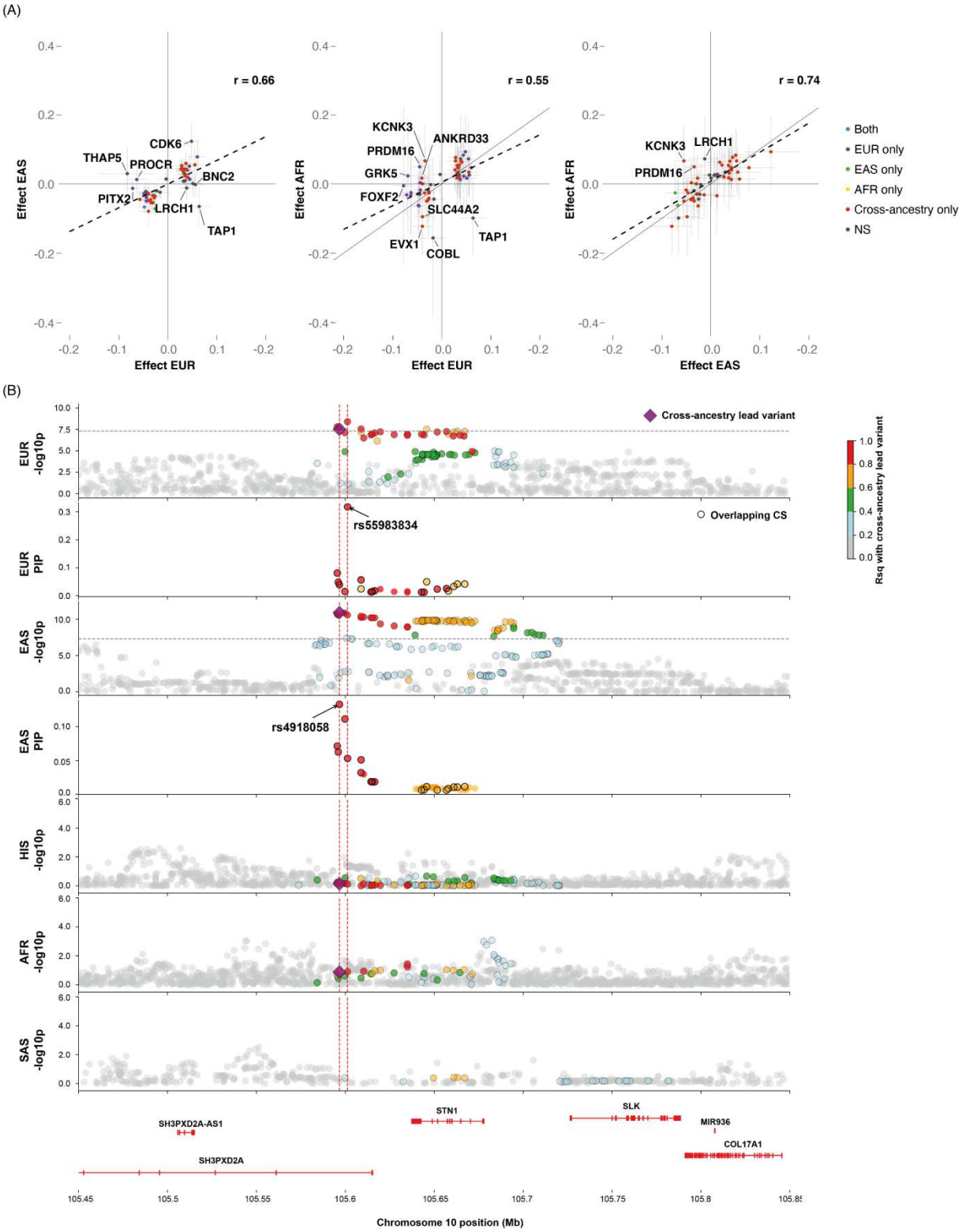
Figures

Figure 1: Identifying genetic variants influencing stroke risk



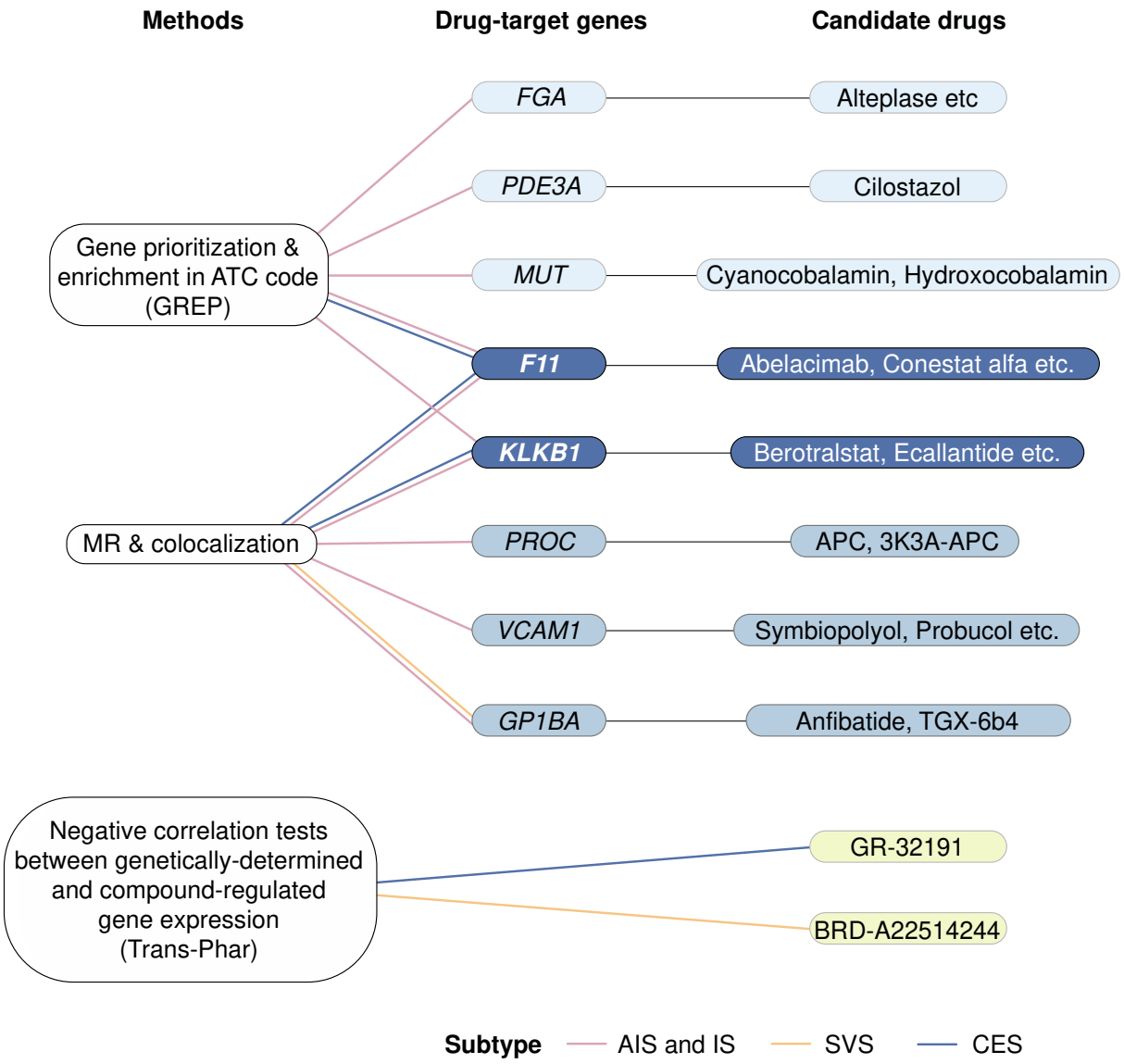
Ideogram of 89 genomic regions influencing stroke risk; circles represent genome-wide significant (GWS) loci in cross-ancestry analyses, diamonds GWS loci in Europeans, triangles GWS loci in East-Asians, and squares GWS loci in African-Americans or South-Asians; colors correspond to associated stroke types (green, AS; red, AIS; light blue, SVS; dark blue, CES; purple, LAS); nearest genes to lead variants are displayed (red: new loci; blue: known loci) loci;

Figure 2: Effect size comparison across ancestry groups of lead variants identified in stroke GWAS and cross-ancestry fine-mapping.



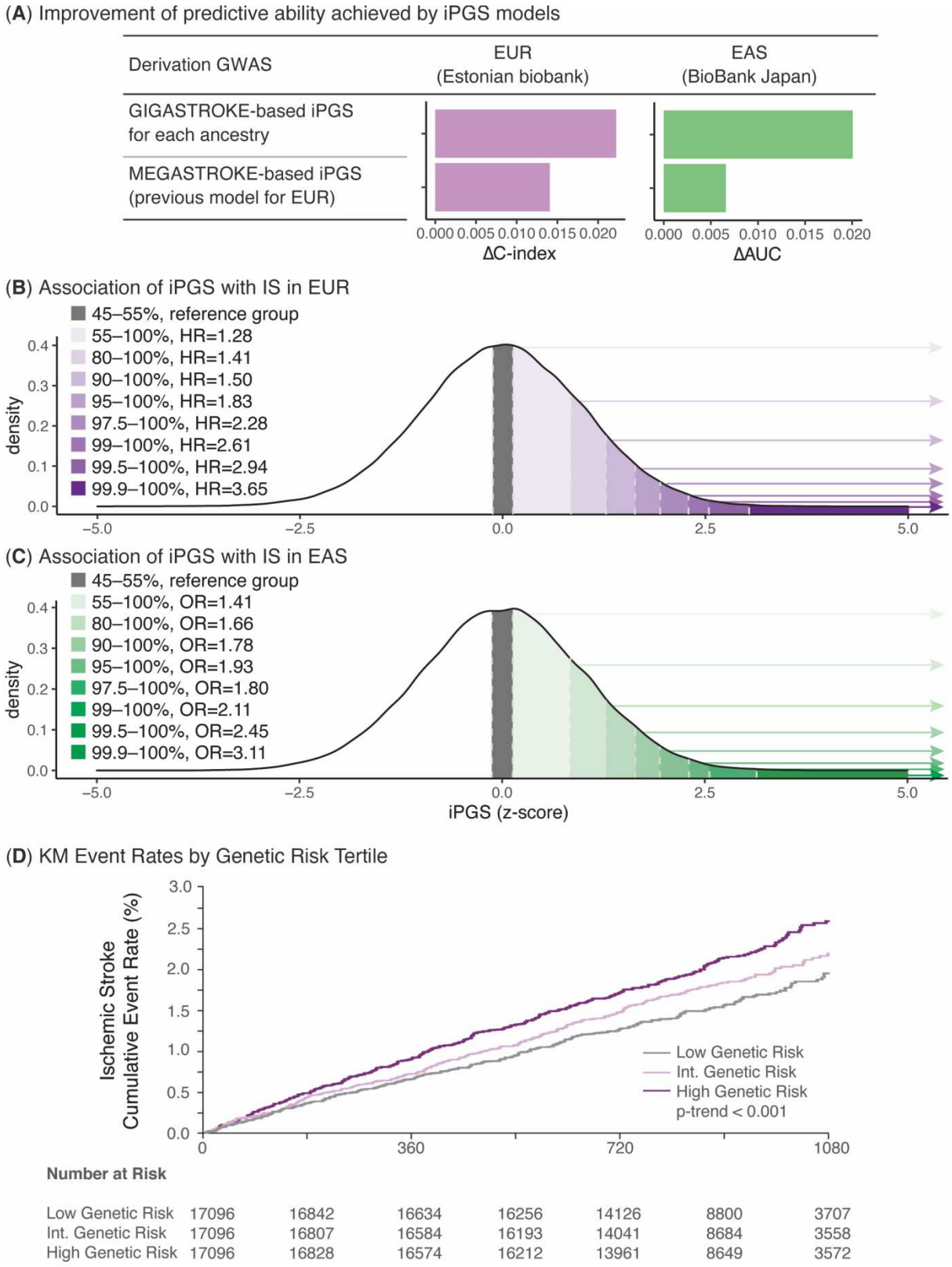
a: Per-allele effect sizes (β) of the 60 lead variants in European ancestry any stroke GWAS meta-analysis (x axis) are plotted against per-allele effect sizes from the East-Asian stroke GWAS meta-analysis (y axis) (left); European per-allele effect sizes (x axis) are plotted against African-American per-allele effect sizes (y axis) (middle); East-Asian per-allele effect sizes (x axis) are plotted against African-American per-allele effect sizes (y axis) (right). Each dot denotes the per-allele effect size; purple (EUR), significant ($p<5\times10^{-8}$) in Europeans only (± cross-ancestry); green (EAS), significant ($p<5\times10^{-8}$) in East-Asians only (± cross-ancestry); yellow (AFR), significant ($p<5\times10^{-8}$) in African-Americans only (± cross-ancestry); blue (both), significant ($p<5\times10^{-8}$) in both plotted ancestries; red (cross-ancestry only), significant ($p<5\times10^{-8}$) in cross-ancestry analyses and not in the two plotted ancestries; grey (NS), non-significant ($p\geq5\times10^{-8}$) in cross-ancestry analyses and in the two plotted ancestries. For SNPs showing a difference in effect size (absolute value) between pairs of ancestries > 0.05, the nearest gene is indicated. r corresponds to the Pearson correlation coefficient between effect sizes across ancestries. **b:** Locus plots of the variants at *SH3PXD2A* locus in 5 ancestries. Fine-mapped variants are only shown in EUR and EAS (insufficient power in other ancestries). Variants are colored by their LD level with the cross-ancestry lead variant (rs4918058) in purple diamond. In fine-mapping panel only variants in CS are shown. Shared variants between CS of EUR and EAS are in black circle. The red vertical lines represent the position of lead variant in EUR (rs55983834) and EAS (rs4918058). The gray horizontal line represents p-value of 5×10^{-8}. LD of each ancestry were derived from 1000 Genomes Project. EUR: European, EAS: East-Asian, HIS: Hispanic, AFR: African, SAS: South Asian, PIP: posterior inclusion probability, CS: 95% credible set of SuSiE.

Figure 3: Genomics driven drug discovery



Top: overlap enrichment analysis using GREP[22]; middle: integrating Mendelian randomization results using cis- and trans-pQTLs as instrumental variables with data from drug databases; bottom: negative correlation tests between compound-regulated gene expression profiles and genetically determined case-control gene expression profiles using Trans-Phar.

Figure 4: Risk prediction in a population and trial setting



(A) Predictive ability and association of polygenic scores for Europeans and East-Asians: improvement of predictive ability achieved by integrative polygenic score (iPGS) is shown. The GIGASTROKE-based iPGS model for each ancestry was compared to a previously constructed MEGASTROKE-based iPGS model for Europeans.[6] (B,C) Association of iPGS for Europeans (B) and East-Asians (C) with ischemic stroke is shown. Compared to the middle decile (45–55%) of the population as a reference group, the risk of high iPGS groups with varying percentile thresholds was estimated using a Cox proportional hazards model for Europeans and logistic regression models for East-Asians with the adjustments for age, sex, and top 5 genetic principal components; (D) Kaplan-Meier event rates for ischemic stroke in EUR in 5 clinical trials (**Methods**), by tertile of genetic risk score at 3 years (the genetic risk score uses effect estimates of the cross-ancestry AS GWAS as weights). Int. indicates intermediate; and KM, Kaplan-Meier, AUC indicates area under the curve; EAS, East-Asians; EUR, Europeans; GWAS, genome-wide association study; iPGS, integrative polygenic score; IS, ischemic stroke.

Methods

Study design and phenotypes

Information on participating studies, study design, and definition of stroke and stroke subtype is provided in the **Supplementary Appendix**. Population characteristics of individual studies are provided in **Supplementary Table 1**. Relevant research and medical ethics committees approved individual studies. All participants or their next-of-kin signed an informed consent.

Genotyping, imputation and genome-wide association testing

Genotyping methods, pre-imputation quality control (QC) of genotypes and imputation methods of individual cohorts are presented in **Supplementary Table 2**. High quality samples and SNPs underwent imputation using mostly Haplotype Reference Consortium (HRC) or 1000 genomes phase 1/3 reference panels and more seldom TOPMed, HapMap or biobank specific reference panels. Individual studies performed a genome-wide association study (GWAS) using logistic regression (or cox regression in some longitudinal population-based cohorts) testing association of genotypes with five stroke phenotypes (AS, AIS, CES, LAS, and SVS) under an additive effect model, adjusting for age, sex, principal components of population stratification, and study-specific covariates when needed (**Supplementary Table 2**).

The R package EasyQC along with in-house custom harmonization scripts were used to perform the QC of individual GWAS summary results. Marker names and alleles were harmonized across studies. Meta-analyses were restricted to autosomal biallelic SNPs from the HRC panel. Duplicate markers were removed. Prior to meta-analysis we removed variants with extreme effect size values ($\log(OR)>5$ or $\log(OR)<-5$), minor allele frequency (MAF) <0.01, imputation quality score less than 0.50 and effective allele count (EAC= 2 × Number of cases × MAF × imputation quality score) less than 6.

The overall analytical strategy is shown in **Extended Data Fig. 1**. We conducted ancestry-specific fixed-effect inverse-variance weighted (IVW) meta-analyses in EUR, EAS, AFR, HIS, and EAS populations, followed by cross-ancestry meta-analyses, using METAL.[1] In each meta-

analysis we removed variants with heterogeneity P-value $<1\times10^{-6}$ and variants available in less than 1/3rd of the total number of cases and less than 1/3rd of the total number of contributing studies. We applied the covariate adjusted LD score regression (cov-LDSC) method to ancestry-specific GWAS meta-analyses without GC correction to test for genomic inflation and to compute robust SNP-heritability estimates in admixed populations.[2]

We applied the conditional and joint analysis approach[3] implemented in the Genome-wide Complex Trait Analysis software[4] (GCTA-COJO) to identify potentially independent signals within the same genomic region. We performed GCTA-COJO analyses on 1) EUR GWAS meta-analysis summary statistics using HRC imputed data of 6,489 French participants from the 3C-study as a reference[5] and 2) EAS ancestry specific GWAS meta-analysis summary statistics using Biobank Japan data as reference (**Supplementary Appendix**).

We additionally performed a cross-ancestry meta-regression using MR-MEGA.[6] Prior to meta-analysis using MR-MEGA we applied the 'genomic inflation' correction option to all input files, and removed variants with extreme effect size values (log(OR)>5 or log(OR)<–5), MAF<0.01, imputation quality score less than 0.50 and effective allele count (EAC= 2 × Number of cases × MAF × imputation quality score) less than 6. Post-meta-analysis we considered loci to be genome-wide significant for MR-MEGA $P<5\times10^{-8}$ and showing nominal association (P <0.05) in at least 1/3rd of studies in any individual ancestral group (EUR, EAS, AFR, HIS, SAS).

Multi-trait association study

To identify additional stroke risk loci we conducted multi-trait analyses of GWAS (MTAG)[7] in Europeans and East-Asians, including traits correlated with specific stroke subtypes, namely coronary artery disease (CAD) for LAS, atrial fibrillation (AF)[8] for CES, and white matter hyperintensity volume[9] (WMH, an MRI-marker of cerebral small vessel disease, available in Europeans only) for SVS. We also ran an MTAG analysis of AS and AIS, including all three correlated traits (CAD, AF, WMH [EUR]). In Europeans we used summary statistics of published GWAS for CAD,[10] AF,[8] and WMH.[9] In East-Asians we used the independent Tohoku Medical Megabank cohort to generate GWAS of AF and CAD (**Supplementary Appendix**). Associations were retained when the following three conditions were verified: (i) MTAG p-value

for stroke $<5\times10^{-8}$; (ii) p-value for stroke <0.05 in the univariate GWAS; and (iii) MTAG p-value for stroke less than the p-value for any of the included traits in univariate GWAS.

Gene and pathway-based analyses

We performed gene-based tests of common variant associations using the VEGAS2 software.[11] All variants in the gene or within 10kb on either side of a gene's transcription site were used to compute a gene-based p-value. We performed analyses using the '-top 10' parameter that tests enrichment of the top 10% variants assigned to a gene accounting for LD between variants and total number of variants within a gene. We used 1000 Genomes phase 3 continental reference samples European, East-Asian, African, South-Asian and South-American (for our Hispanic samples), to compute LD between variants for respective ancestry-specific gene-based analyses. We then meta-analyzed ancestry-specific gene-based results, using Stouffer's method for sample size weighted combination of P-values. Gene-wide significance was defined as $p<2.72\text{x}10^{-6}$, correcting for 18,371 autosomal protein-coding genes tested.

Next, we used the ancestry-specific gene-based association p-values to perform pathway analyses for individual ancestral groups, testing enrichment of gene-based p-values in Biosystems pathways with VEGAS2Pathway.[11,12] For each stroke phenotype, we meta-analysed the ancestry-specific pathway association p-values using Stouffer's method. Pathway-wide significance was defined at $p<5.01\text{x}10^{-6}$ correcting for 9,977 Biosystems pathways tested.

Shared genetic variation, genetic correlation, Mendelian randomization with vascular risk traits

We explored shared genetic variation with 12 vascular risk factor and related disease traits in Europeans using summary statistics of GWAS on systolic blood pressure (SBP),[13] diastolic blood pressure (DBP),[13] body mass index (BMI) and waist-to-hip ratio (WHR),[14] high density lipoprotein (HDL) cholesterol,[15] low density lipoprotein (LDL) cholesterol,[15] triglycerides,[15] type 2 diabetes,[16] WMH volume,[9] atrial fibrillation,[8] coronary artery disease,[10] and venous thromboembolism (VTE).[17] We extracted sentinel stroke risk variants (or a proxy [$r^2>0.9$]) that showed genome-wide significant association ($p<5\text{x}10^{-8}$) with the aforementioned vascular risk factors.

We then systematically explored genetic correlations and potentially causal associations between vascular risk traits and risk of stroke using LD score regression (LDSC) and Mendelian randomization (MR) analyses, with 12 (in Europeans) and 6 (in East-Asians) vascular risk traits. In individuals of European ancestry, we used summary statistics of the aforementioned GWAS.[8-10,13-17] For the analysis in East-Asians we used unpublished GWAS for SBP, DBP, LDL and HDL cholesterol, triglycerides, and BMI in up to 53,323 participants of the independent Tohoku Medical Megabank Project (**Supplementary Appendix**)

We used cov-LDSC to compute genetic correlations between stroke and vascular risk traits, using European and East-Asian GWAS summary files and 1000Gp3v5 reference data of respective continental ancestries (considering the recommended subset of high quality HapMap3 SNPs only).

For MR analyses, we constructed genetic instruments for each vascular risk trait based on genome-wide significant associations ($p<5x10^{-8}$) after clumping for LD at $r^2<0.01$ (based on European and East-Asian 1000G). We applied two-sample MR analyses in the GIGASTROKE summary statistics separately for individuals of EUR and EAS ancestry based on variant associations derived from the aforementioned sources. Following extraction of the association estimates and harmonization of their direction-of-effect alleles, we computed MR estimates with fixed-effects inverse-variance weighted (IVW) analyses.[18] As a measure of pleiotropy, we assessed heterogeneity across the MR estimates for each instrument in the IVW MR analyses with Cochran's Q statistic ($p<0.05$ was considered significant).[19] We further applied alternative MR methods that are more robust to the use of pleiotropic instruments: the weighted median estimator allows the use of invalid instruments under the assumption that at least half of the instruments used in the MR analysis are valid;[20] MR-Egger regression allows for the estimation of an intercept term, provides less precise estimates and relies on the assumption that the strengths of potential pleiotropic instruments are independent of their direct associations with the outcome.[21] The intercept obtained from MR-Egger regression was used as a measure of directional pleiotropy ($p<0.05$ indicated significance).[21] MR analyses were performed in R v4.1.1 using the MendelianRandomization package.

For all genetic correlation and MR analyses, we set statistical significance at a Bonferroni-corrected p-value$<4.17x10^{-3}$ in EUR (correcting for 12 vascular risk traits) and $<8.33x10^{-3}$ in EAS (correcting for 6 vascular risk traits).

Cross-ancestry fine mapping

Fine-mapping was performed separately for Europeans and East-Asians with susieR v.0.9.1[22] on all variants within 3Mb of the lead variant of each genomic risk locus (60 loci reaching genome-wide significance in the IVW meta-analysis). Unrelated individuals from UK Biobank (UKB, N=420,000) and Biobank Japan (BBJ, N=170,000) were used as ancestry-matched LD reference panels that fulfill the sample size requirement.[23] After extracting variants present in the LD reference panel, default settings of susieR were used while allowing a maximum of 10 putative causal variants in each locus. We checked the loci harboring multiple 95% confident credible sets and removed likely false positive signals from the cross-ancestry analysis by checking LD pattern. We compared the variants in CS of the same loci between EUR and EAS.

To detect putative causal regulatory variants in the credible sets, we conducted an *in silico* mutagenesis analysis using MENTR (Mutation Effect prediction on Non-coding RNA (ncRNA) TRanscription; https://github.com/koido/MENTR), a quantitative machine-learning framework that predicts the effect of genetic variants on transcription, including transcription of ncRNAs, in a tissue- or cell-type-dependent manner.[24,25] The *in silico* mutations predicted to have strong effects are highly concordant with the observed effects of known variants in a cell-type-dependent manner. Furthermore, MENTR does not use population datasets and therefore is less susceptible to LD-dependent association signals, allowing to pin-point prediction of causal variants on transcriptional changes. From 1,274 variants in the credible sets from the EUR and EAS fine-mapping, we searched FANTOM5 promoters and enhancers, obtained by cap analysis of gene expression (CAGE), within +/- 100-kb from each variant. As a result, we found 37,878 variant-transcript pairs comprising 1,270 variants and 2,350 transcripts. We used MENTR with the pre-trained FANTOM5 347 cell/tissue models + LCL models[26-29] and extracted reliable predictions using the pre-determined robust threshold (absolute *in silico* mutation effects ≥0.1, achieving >90% concordance for predicting effects on expression).

Transcriptome-wide and proteome-wide association studies

We performed transcriptome-wide association studies (TWAS) using TWAS-Fusion[30] to identify genes whose expression is significantly associated with stroke risk. We restricted the analysis to tissues considered relevant for cerebrovascular disease, and used precomputed functional weights

from 21 publicly available expression quantitative trait loci (eQTL) reference panels from blood (Netherlands Twin Registry, NTR; Young Finns Study, YFS)[30,31], arterial and heart (Genotype-Tissue Expression version 7 [GTEx v7]),[32] and brain tissues (GTEx v7, CommonMind Consortium [CMC]).[32,33] In addition, we used the newly developed cross-tissue weights generated in GTEx v8 using sparse canonical correlation analysis (sCCA) across 49 tissues available on the TWAS-Fusion website, including gene expression models for the first 3 canonical vectors (sCCA1-3), which were shown to capture most of the gene expression signal.[34] TWAS-Fusion was then used to estimate the TWAS association statistics between predicted gene expression and stroke by integrating information from expression reference panels (SNP-expression weights), GWAS summary statistics (SNP-stroke effect estimates), and LD reference panels (SNP correlation matrix).[30] Transcriptome-wide significant genes (eGenes) and the corresponding eQTLs were determined using Bonferroni correction, based on the average number of features (5005.8 genes) tested across all reference panels and correcting for the 5 stroke phenotypes ($p<2.0x10^{-6}$). eGenes were then tested in conditional analysis as implemented in the Fusion software.[30] To ensure observed associations does not reflect random correlation between gene expression and non-causal variants associated with stroke, we performed a colocalization analysis (COLOC) on the conditionally significant genes ($p<0.05$) to estimate the posterior probability of a shared causal variant between the gene expression and trait association (PP4).[35] We used a prior probability of $p<2.0x10^{-6}$ for the stroke association. Genes presenting a PP4$\geq$0.75, for which eQTLs did not reach genome-wide significance in association with stroke, and were not in LD ($r^2<0.01$) with any of the lead SNPs of genome-wide significant risk loci for stroke, were considered as novel.

Using similar parameters in TWAS-Fusion,[36] we also performed a proteome-wide association study. For this analysis we used the precomputed weights for protein expression in dorsolateral prefrontal cortex (dPFC)[37] from the ROS/MAP study (n=376, proteins=1,475)[38] and the Banner Sun Health Institute study (n=152, proteins=1,145).[39] Proteome-wide significant genes and the corresponding pQTLs were determined using Bonferroni correction, on the number of proteins tested across the reference panel and correcting for the 5 stroke phenotypes ($p<1.7x10^{-4}$ for ROS/MAP and $p<2.2x10^{-8}$ for the Banner Sun Health Institute). We then followed the same method as described for the TWAS

Brain single-cell expression of TWAS genes

Single-nucleus RNA-sequencing data of the dLPFC region of 24 aging individuals chosen to represent the range of pathologic and clinical diagnoses of AD dementia, from the ROS/MAP cohorts, was obtained.[40] RNA profiles of cells annotated as endothelial, pericytes or smooth muscle cells and vascular leptomeningeal cells (VLMC) were used, and a pseudo-bulk RNA profile was generated for each cell type, by averaging the expression of all genes across the cells. Average expression level and percentage of expressed genes were calculated for genes of interest using the DotPlot function from the Seurat package V4.0.4 in R V.4.1.1.

Genomics-driven drug discovery

We used three methodologies for in-depth genomics-driven drug discovery as described previously:[41] (i) an overlap enrichment analysis of disease-risk genes in drug-target genes in medication categories, (ii) negative correlation tests between genetically determined case-control gene expression profiles and compound-regulated gene expression profiles, and (iii) endophenotype Mendelian randomization (MR). The detail of the methods is described in the following sections. For the overlap enrichment analysis and the endophenotype MR nominated drug targets we curated drug candidates from four major drug databases, DrugBank,[42] Therapeutic Target Database (TTD),[43] PharmGKB,[44] and Open Target Platform.[45] As for the endophenotype MR, we curated drugs with opposite effects against the signs of the MR effect sizes. On the other hand, the negative correlation tests directly prioritized candidate compounds. We manually curated supporting evidence for candidate drugs and compounds.

Overlap enrichment analysis of disease-risk genes in drug-target genes in medication categories

We ran MAGMA[46] to summarize variant-level p-values into gene-level and used the genes with false discovery rates (FDR) less than 0.05 as the disease-risk genes. We then used GREP[47] to perform a series of Fisher's exact tests for the enrichment of the disease-risk genes in the drug-target genes involved in the drug indication categories, Anatomical Therapeutic Chemical Classification System (ATC) codes.

Negative correlation tests between genetically determined and compound-regulated gene expression profiles

We nominated the compounds with inverse effects on gene expression against genetically determined gene expression by using Trans-Phar[48]. In brief, genetically determined case-control gene expression was inferred for 44 tissues in the Genotype-Tissue Expression project v7[32] with FOCUS[49], and the genes in the top decile for the absolute value of the Z-score were used for the following correlation analysis. The Library of Integrated Network-based Cellular Signatures project (LINCS) CMAP L1000 library data[50] was used for the compound library. After matching the tissues in GTEx with the cell lines in the LINCS L1000 library, we performed a series of Spearman's rank correlation tests for 308,872 pairs of genetically determined gene expression and the compound-perturbed cell-type gene expression profiles. We prioritized compounds with FDR<0.1 as we previously showed that the compounds with FDR < 0.1 contained plausible therapeutic targets with literature supports.[41]

Endophenotype Mendelian randomization

To pin-point the disease-causing proteins that were targeted by existing drugs, we performed MR analysis (specifically, Wald ratio test) by using lead variants in protein quantitative trait loci (pQTL) as instrumental variables. We used the tier 1 lead variants defined by Zheng *et al.*[51] to avoid confounding by horizontal pleiotropy. The tier 1 variants were summarized from five pQTL studies[52-56] and excluded the variants with heterogeneous effect sizes among the studies or the number of associated proteins larger than five. We restricted the lead variants to the variants associated with drug-target proteins. For the lead variants of pQTL that were missing in the stroke GWAS summary statistics, the proxy variants with the largest R^2 were used if the R^2 was greater than 0.8. In total, we used 277 lead variants for 218 drug-target proteins for MR. Weused the "TwoSampleMR" R package[57] for MR analysis. As post-MR quality controls, we performed (i) directionality check of causal relationships by Steiger filtering[58] and (ii) colocalization analysis for the proteins with FDR < 0.05. To examine colocalization assuming multiple causal variants per locus, coloc[35] was applied to the decomposed signals by SuSiE[22] for the variants within 500 kb upstream and downstream of the lead variants (coloc + SuSiE).[59] If SuSiE did not converge after 10,000 iterations, coloc was used instead. Coloc + SuSiE and coloc were run with their respective default parameters. For the two pQTL studies without public summary statistics,[52,56] we compared the R^2 between the lead variants of the pQTL study and the stroke GWAS. We considered that colocalization occurred when the maximum posterior probability (i.e., PP.H4) was greater than 0.8 or R^2 was greater than 0.8.

PheWAS

We conducted a phenome-wide association study (PheWAS),[60] in Estonian Biobank (EstBB) for the pQTL variants and rare deleterious variants in identified drug target genes using the R software (4.0.3). We tested the association between ICD10 main codes and genetic variants using logistic regression adjusting for sex, birth year and 10 genotype PCs. All ICD10 codes with number of cases<100 and all variants with MAF<0.001 were removed from the analysis. We applied Bonferroni correction to select statistically significant associations (number of tested ICD main codes:1,034, number of tested SNPs:7 corrected p-value threshold:$0.05/1034*7=6.91\times10^{-6}$). Results were visualized using the PheWas library (https://github.com/PheWAS/PheWAS).

Polygenic risk prediction

We constructed integrative polygenic scores (iPGS) models for stroke in Europeans and East-Asians (**Extended Data Fig.9-10**). For each ancestry, two independent datasets were used for model training and evaluation, respectively. We used as input summary statistics data of multiple GWAS for stroke outcomes and vascular risk traits in order to derive iPGS models. We denote the number of input GWAS as N. For each of the N GWAS summary data, 37 candidate single trait polygenic score (sPGS) models were generated using P+T,[61,62] LDpred,[63] and PRScs[64] algorithms with an ancestry-specific LD reference panel from the 1000 Genomes Project (**Supplementary Methods**).[65] The plink (v1.90b6.8),[66] LDpred (v.1.0.11),[63] and PRScs.py (Jun 5, 2021)[64] programs were used to compute P+T, LDpred, and PRScs models, respectively. Among the 37 candidate models, subsequently, the best sPGS model, which was defined as the model that showed a maximal improvement in AUC over a base model (age, sex, and top 5 PCs were included in the base model), was selected using the model training dataset.[62,67] Then, N best sPGS models were selected from the N input GWASs.

Each best sPGS was z-transformed (zero mean and unit SD) over the model training dataset, followed by elastic-net logistic regression[68] to model the associations between the N sPGS variables and IS with the adjustments for age, sex, and top 5 genetic PCs. Two regularization parameters (α and λ) were optimized using 10-fold cross-validation. Then, coefficients (weights) for the N sPGS models were determined by the elastic-net logistic regression with optimal regularization parameters, followed by the integration of N sPGS models into a single iPGS

model according to the formula presented in a previous study.[67] The elastic-net regression was performed using the glmnet R package.[69]

The predictive ability of the iPGS model was estimated using the model evaluation dataset, where we evaluated the improvement in C-index for a prospective cohort dataset (Europeans) or AUC for a case-control dataset (East-Asians) over a base model that includes age, sex, and top 5 genetic PCs.

We used EstBB data for the model training and evaluation of iPGS model in Europeans . The model training dataset was composed of 1,003 prevalent IS cases at baseline and 8,997 controls. The control subjects were randomly selected among EstBB participants who had no history of AS at baseline and who did not develop AS during follow-up. The remaining 102,099 EstBB subjects were used for the model evaluation (mean±SD age at baseline, 44.0±15.7 years; 37.8% men). Among the subjects in the model evaluation dataset, 1,128 incident IS cases were observed during 4.6±4.8 years. To derive the European iPGS model, we incorporated 5 ancestry-specific and 5 cross-ancestry stroke GWAS (AS, AIS, LAS, SVS, and CES) from the GIGASTROKE project, and 14 GWASs of vascular risk traits from other groups (**Extended Data Fig.9**). To avoid the overlap of subjects across datasets, the GWAS summary statistics for stroke outcomes were re-calculated for the iPGS analysis by excluding the EstBB from the meta-analysis of GIGASTROKE studies. To allow comparison with a previous European iPGS model based on the MEGASTROKE GWAS,[67] we selected the best sPGS model from 10 GWASs of vascular risk traits (T2D, SBP, DBP, TC, LDL-C, HDL-C, TG, BMI, height, and smoking)[15,70-74] using the model training dataset. The 10 selected sPGS models and 4 pre-computed sPGS models (one AF[75] and three CAD models[10,76,77] provided by the authors of the previous study[67]) were incorporated into the GIGASTROKE-based iPGS model as vascular risk traits.

For the East-Asian iPGS model we used BBJ data for the model training and evaluation. The model training dataset was composed of 577 IS cases and 9,232 controls, whereas there were 1,470 IS cases and 40,459 controls in the model training dataset. The mean±SD of age at recruitment was 69.2±10.8 years old for cases and 66.5±12.5 for controls in the model evaluation dataset. The percentage of males was 70.0% for cases and 53.1% for controls. The two case-control datasets were not included in the meta-analysis of GIGASTROKE studies, and therefore, the overlap of subjects across datasets was avoided. To derive the East-Asian iPGS model, we incorporated 5 ancestry-specific and 5 cross-ancestry stroke GWAS (AS, AIS, LAS, SVS, and

CES) from the GIGASTROKE project, and 37 GWAS of vascular risk traits from other groups (**Extended Data Fig. 10**). Among the 37 GWAS, 21 were Japanese-ancestry GWAS[78-84] and 16 were cross-ancestry GWAS.[85]

Genetic risk score in a clinical trial setting

Subjects who had consented for genetic testing and who were of European ancestry from the ENGAGE AF-TIMI (Effective Anticoagulation with Factor Xa Next Generation in Atrial Fibrillation),[86] SOLID-TIMI (Stabilization of Plaques Using Darapladib),[87] SAVOR-TIMI (Saxagliptin Assessment of Vascular Outcomes Recorded in Patients with Diabetes Mellitus),[88] PEGASUS-TIMI (Prevention of Cardiovascular Events in Patients With Prior Heart Attack Using Ticagrelor Compared to Placebo on a Background of Aspirin),[89] and FOURIER (Further Cardiovascular Outcomes Research With PCSK9 Inhibition in Patients With Elevated Risk)[90] trials were included in this analysis. Methods for genotyping and imputation have previously been published and are summarized in **Supplementary Table 2**).[91,92] A set of 58 sentinel variants at stroke risk loci identified in IVW meta-analysis was used to calculate a GRS in each trial participant and identify tertiles of genetic risk (**Supplementary Table 40**). A Cox model was used to estimate hazard ratios for ischemic stroke associated with the quantitative GRS and across genetic risk groups, adjusted for clinical risk factors (age, sex, hypertension, hyperlipidemia, diabetes, smoking, CAD, AF, and congestive heart failure) and the first 5 principal components of population stratification. Analyses were conducted primarily in participants of European ancestry (N=51,288, with 960 incident strokes), with secondary analyses in the much smaller East-Asian (N=1,312, with 27 incident strokes) ancestry subset, using AS cross-ancestry IVW meta-analysis effect estimates as weights for the primary analysis and ancestry-specific and AIS effect estimates for secondary analyses. We also looked separately at associations with incident stroke in participants with and without previous stroke.

Methods References

1 Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-2191, doi:10.1093/bioinformatics/btq340 (2010).

2 Luo, Y. *et al.* Estimating heritability and its enrichment in tissue-specific gene sets in admixed populations. *Hum Mol Genet* **30**, 1521-1534, doi:10.1093/hmg/ddab130 (2021).

3 Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet* **44**, 369-375, S361-363, doi:10.1038/ng.2213 (2012).

4 Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* **88**, 76-82, doi:10.1016/j.ajhg.2010.11.011 (2011).

5 3C Study Group. Vascular factors and risk of dementia: design of the Three-City Study and baseline characteristics of the study population. *Neuroepidemiology* **22**, 316-325 (2003).

6 Magi, R. *et al.* Trans-ethnic meta-regression of genome-wide association studies accounting for ancestry increases power for discovery and improves fine-mapping resolution. *Hum Mol Genet* **26**, 3639-3650, doi:10.1093/hmg/ddx280 (2017).

7 Turley, P. *et al.* Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nat Genet* **50**, 229-237, doi:10.1038/s41588-017-0009-4 (2018).

8 Nielsen, J. B. *et al.* Biobank-driven genomic discovery yields new insight into atrial fibrillation biology. *Nat Genet* **50**, 1234-1239, doi:10.1038/s41588-018-0171-3 (2018).

9 Sargurupremraj, M. *et al.* Cerebral small vessel disease genomics and its implications across the lifespan. *Nat Commun* **11**, 6285, doi:10.1038/s41467-020-19111-2 (2020).

10 Nikpay, M. *et al.* A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat Genet* **47**, 1121-1130, doi:10.1038/ng.3396 (2015).

11 Mishra, A. & Macgregor, S. VEGAS2: Software for More Flexible Gene-Based Testing. *Twin Res Hum Genet* **18**, 86-91, doi:10.1017/thg.2014.79 (2015).

12 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, doi:10.1017/thg.2016.100 (2017).

13 Evangelou, E. *et al.* Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. *Nat Genet* **50**, 1412-1425, doi:10.1038/s41588-018-0205-x (2018).

14 Pulit, S. L. *et al.* Meta-analysis of genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry. *Hum Mol Genet* **28**, 166-174, doi:10.1093/hmg/ddy327 (2019).

15 Willer, C. J. *et al.* Discovery and refinement of loci associated with lipid levels. *Nat Genet* **45**, 1274-1283, doi:10.1038/ng.2797 (2013).

16 Xue, A. *et al.* Genome-wide association analyses identify 143 risk variants and putative regulatory mechanisms for type 2 diabetes. *Nat Commun* **9**, 2941, doi:10.1038/s41467-018-04951-w (2018).

17 Lindstrom, S. *et al.* Genomic and transcriptomic association studies identify 16 novel susceptibility loci for venous thromboembolism. *Blood* **134**, 1645-1657, doi:10.1182/blood.2019000435 (2019).

18 Burgess, S., Butterworth, A. & Thompson, S. G. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet Epidemiol* **37**, 658-665, doi:10.1002/gepi.21758 (2013).

19 Bowden, J., Hemani, G. & Davey Smith, G. Invited Commentary: Detecting Individual and Global Horizontal Pleiotropy in Mendelian Randomization-A Job for the Humble Heterogeneity Statistic? *Am J Epidemiol* **187**, 2681-2685, doi:10.1093/aje/kwy185 (2018).

20 Hartwig, F. P., Davey Smith, G. & Bowden, J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. *Int J Epidemiol* **46**, 1985-1998, doi:10.1093/ije/dyx102 (2017).

21 Bowden, J., Davey Smith, G. & Burgess, S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol* **44**, 512-525, doi:10.1093/ije/dyv080 (2015).

22 Wang, G., Sarkar, A., Carbonetto, P. & Stephens, M. A simple new approach to variable selection in regression, with application to genetic fine mapping. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **82**, 1273-1300, doi:10.1111/rssb.12388 (2020).

23 Benner, C. *et al.* Prospects of Fine-Mapping Trait-Associated Genomic Regions by Using Summary Statistics from Genome-wide Association Studies. *Am J Hum Genet* **101**, 539-551, doi:10.1016/j.ajhg.2017.08.012 (2017).

24 Koido, M. *et al.* Predicting cell-type-specific non-coding RNA transcription from genome sequence. (bioRxiv 2020.03.29.011205; doi: https://doi.org/10.1101/2020.03.29.011205).

25 Zhou, J. *et al.* Deep learning sequence-based ab initio prediction of variant effects on expression and disease risk. *Nat Genet* **50**, 1171-1179, doi:10.1038/s41588-018-0160-6 (2018).

26 Andersson, R. *et al.* An atlas of active enhancers across human cell types and tissues. *Nature* **507**, 455-461, doi:10.1038/nature12787 (2014).

27 Forrest, A. R. *et al.* A promoter-level mammalian expression atlas. *Nature* **507**, 462-470, doi:10.1038/nature13182 (2014).

28 Hon, C. C. *et al.* An atlas of human long non-coding RNAs with accurate 5' ends. *Nature* **543**, 199-204, doi:10.1038/nature21374 (2017).

29 Garieri, M. *et al.* The effect of genetic variation on promoter usage and enhancer activity. *Nat Commun* **8**, 1358, doi:10.1038/s41467-017-01467-7 (2017).

30 Gusev, A. *et al.* Integrative approaches for large-scale transcriptome-wide association studies. *Nature Genetics* **48**, 245-252, doi:10.1038/ng.3506 (2016).

31 Wright, F. A. *et al.* Heritability and genomics of gene expression in peripheral blood. *Nature Genetics* **46**, 430-437, doi:10.1038/ng.2951 (2014).

32 Consortium, G. *et al.* Genetic effects on gene expression across human tissues. *Nature* **550**, 204-213, doi:10.1038/nature24277 (2017).

33 Fromer, M. *et al.* Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nature Neuroscience* **19**, 1442-1453, doi:10.1038/nn.4399 (2016).

34 Feng, H. *et al.* Leveraging expression from multiple tissues using sparse canonical correlation analysis and aggregate tests improves the power of transcriptome-wide association studies. *PLoS genetics* **17**, e1008973, doi:10.1371/journal.pgen.1008973 (2021).

35 Giambartolomei, C. *et al.* Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS genetics* **10**, e1004383, doi:10.1371/journal.pgen.1004383 (2014).

36 Gusev, A. *et al.* Integrative approaches for large-scale transcriptome-wide association studies. *Nat Genet* **48**, 245-252, doi:10.1038/ng.3506 (2016).

37 Wingo, A. P. *et al.* Integrating human brain proteomes with genome-wide association data implicates new proteins in Alzheimer's disease pathogenesis. *Nat Genet* **53**, 143-146, doi:10.1038/s41588-020-00773-z (2021).

38 Bennett, D. A. *et al.* Religious Orders Study and Rush Memory and Aging Project. *J Alzheimers Dis* **64**, S161-S189, doi:10.3233/jad-179939 (2018).

39 Beach, T. G. *et al.* Arizona Study of Aging and Neurodegenerative Disorders and Brain and Body Donation Program. *Neuropathology* **35**, 354-389, doi:10.1111/neup.12189 (2015).

40 Cain, A. *et al.* Multi-cellular communities are perturbed in the aging human brain and with Alzheimer's disease. (bioRxiv 2020.12.22.424084; doi: https://doi.org/10.1101/2020.12.22.424084).

41 Namba, S. *et al.* A practical guideline of genomics-driven drug discovery in the era of global biobank meta-analysis. *medRxiv* doi:https://doi.org/10.1101/2021.12.03.21267280 (2021.12.03.21267280).

42 Wishart, D. S. *et al.* DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Research* **46**, D1074-D1082, doi:10.1093/nar/gkx1037 (2018).

43 Chen, X., Ji, Z. L. & Chen, Y. Z. TTD: Therapeutic Target Database. *Nucleic Acids Research* **30**, 412-415, doi:10.1093/nar/30.1.412 (2002).

44 Whirl-Carrillo, M. *et al.* An Evidence-Based Framework for Evaluating Pharmacogenomics Knowledge for Personalized Medicine. *Clinical Pharmacology and Therapeutics* **110**, 563-572, doi:10.1002/cpt.2350 (2021).

45 Ochoa, D. *et al.* Open Targets Platform: supporting systematic drug-target identification and prioritisation. *Nucleic Acids Research* **49**, D1302-D1310, doi:10.1093/nar/gkaa1027 (2021).

46 de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS computational biology* **11**, e1004219, doi:10.1371/journal.pcbi.1004219 (2015).

47 Sakaue, S. & Okada, Y. GREP: genome for REPositioning drugs. *Bioinformatics (Oxford, England)* **35**, 3821-3823, doi:10.1093/bioinformatics/btz166 (2019).

48 Konuma, T., Ogawa, K. & Okada, Y. Integration of genetically regulated gene expression and pharmacological library provides therapeutic drug candidates. *Human Molecular Genetics* **30**, 294-304, doi:10.1093/hmg/ddab049 (2021).

49 Mancuso, N. *et al.* Probabilistic fine-mapping of transcriptome-wide association studies. *Nature Genetics* **51**, 675-682, doi:10.1038/s41588-019-0367-1 (2019).

50 Subramanian, A. *et al.* A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. *Cell* **171**, 1437-1452.e1417, doi:10.1016/j.cell.2017.10.049 (2017).

51 Zheng, J. *et al.* Phenome-wide Mendelian randomization mapping the influence of the plasma proteome on complex diseases. *Nature Genetics* **52**, 1122-1131, doi:10.1038/s41588-020-0682-6 (2020).

52 Emilsson, V. *et al.* Co-regulatory networks of human serum proteins link genetics to disease. *Science (New York, N.Y.)* **361**, 769-773, doi:10.1126/science.aaq1327 (2018).

53 Folkersen, L. *et al.* Mapping of 79 loci for 83 plasma protein biomarkers in cardiovascular disease. *PLoS genetics* **13**, e1006706, doi:10.1371/journal.pgen.1006706 (2017).

54 Suhre, K. *et al.* Connecting genetic risk to disease end points through the human blood plasma proteome. *Nature Communications* **8**, 14357, doi:10.1038/ncomms14357 (2017).

55 Sun, B. B. *et al.* Genomic atlas of the human plasma proteome. *Nature* **558**, 73-79, doi:10.1038/s41586-018-0175-2 (2018).

56 Yao, C. *et al.* Genome-wide mapping of plasma protein QTLs identifies putatively causal genes and pathways for cardiovascular disease. *Nature Communications* **9**, 3268, doi:10.1038/s41467-018-05512-x (2018).

57 Hemani, G. *et al.* The MR-Base platform supports systematic causal inference across the human phenome. *eLife* **7**, e34408, doi:10.7554/eLife.34408 (2018).

58 Hemani, G., Tilling, K. & Davey Smith, G. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. *PLoS genetics* **13**, e1007081, doi:10.1371/journal.pgen.1007081 (2017).

59 Wallace, C. A more accurate method for colocalisation analysis allowing for multiple causal variants. *PLoS genetics* **17**, e1009440, doi:10.1371/journal.pgen.1009440 (2021).

60 Nguyen, P. A., Born, D. A., Deaton, A. M., Nioi, P. & Ward, L. D. Phenotypes associated with genes encoding drug targets are predictive of clinical trial side effects. *Nat Commun* **10**, 1579, doi:10.1038/s41467-019-09407-3 (2019).

61 Purcell, S. M. *et al.* Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748-752, doi:10.1038/nature08185 (2009).

62 Khera, A. V. *et al.* Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet* **50**, 1219-1224, doi:10.1038/s41588-018-0183-z (2018).

63 Vilhjálmsson, B. J. *et al.* Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. *Am J Hum Genet* **97**, 576-592, doi:10.1016/j.ajhg.2015.09.001 (2015).

64 Ge, T., Chen, C.-Y., Ni, Y., Feng, Y.-C. A. & Smoller, J. W. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nature Communications* **10**, 1776, doi:10.1038/s41467-019-09718-5 (2019).

65 Genomes Project, C. *et al.* A global reference for human genetic variation. *Nature* **526**, 68-74, doi:10.1038/nature15393 (2015).

66 Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7, doi:10.1186/s13742-015-0047-8 (2015).

67 Abraham, G. *et al.* Genomic risk score offers predictive performance comparable to clinical risk factors for ischaemic stroke. *Nat Commun* **10**, 5819, doi:10.1038/s41467-019-13848-1 (2019).

68 Zou, H. & Hastie, T. Regularization and variable selection via the elastic net. *J. R. Statist. Soc. B* **67, Part 2**, 301–320 (2005).

69 Friedman, J., Hastie, T. & Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw* **33**, 1-22 (2010).

70 Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat Genet* **42**, 441-447, doi:10.1038/ng.571 (2010).

71 Wood, A. R. *et al.* Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat Genet* **46**, 1173-1186, doi:10.1038/ng.3097 (2014).

72 Locke, A. E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197-206, doi:10.1038/nature14177 (2015).

73 Scott, R. A. *et al.* An Expanded Genome-Wide Association Study of Type 2 Diabetes in Europeans. *Diabetes* **66**, 2888-2902, doi:10.2337/db16-1253 (2017).

74 Wain, L. V. *et al.* Novel Blood Pressure Locus and Gene Discovery Using Genome-Wide Association Study and Expression Data Sets From Blood and the Kidney. *Hypertension*, doi:10.1161/hypertensionaha.117.09438 (2017).

75 Weng, L. C. *et al.* Genetic Predisposition, Clinical Risk Factor Burden, and Lifetime Risk of Atrial Fibrillation. *Circulation* **137**, 1027-1038, doi:10.1161/circulationaha.117.031431 (2018).

76 Abraham, G. *et al.* Genomic prediction of coronary heart disease. *Eur Heart J* **37**, 3267-3278, doi:10.1093/eurheartj/ehw450 (2016).

77 Inouye, M. *et al.* Genomic Risk Prediction of Coronary Artery Disease in 480,000 Adults: Implications for Primary Prevention. *J Am Coll Cardiol* **72**, 1883-1893, doi:10.1016/j.jacc.2018.07.079 (2018).

78 Ishigaki, K. *et al.* Large-scale genome-wide association study in a Japanese population identifies novel susceptibility loci across different diseases. *Nat Genet* **52**, 669-679, doi:10.1038/s41588-020-0640-3 (2020).

79 Low, S. K. *et al.* Identification of six new genetic loci associated with atrial fibrillation in the Japanese population. *Nat Genet* **49**, 953-958, doi:10.1038/ng.3842 (2017).

80 Akiyama, M. *et al.* Genome-wide association study identifies 112 new loci for body mass index in the Japanese population. *Nat Genet* **49**, 1458-1467, doi:10.1038/ng.3951 (2017).

81 Kanai, M. *et al.* Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. *Nat Genet* **50**, 390-400, doi:10.1038/s41588-018-0047-6 (2018).

82 Akiyama, M. *et al.* Characterizing rare and low-frequency height-associated variants in the Japanese population. *Nat Commun* **10**, 4393, doi:10.1038/s41467-019-12276-5 (2019).

83 Matoba, N. *et al.* GWAS of smoking behaviour in 165,436 Japanese people reveals seven new loci and shared genetic architecture. *Nat Hum Behav* **3**, 471-477, doi:10.1038/s41562-019-0557-y (2019).

84 Matoba, N. *et al.* GWAS of 165,084 Japanese individuals identified nine loci associated with dietary habits. *Nat Hum Behav* **4**, 308-316, doi:10.1038/s41562-019-0805-1 (2020).

85 Sakaue, S. *et al.* A cross-population atlas of genetic associations for 220 human phenotypes. *Nat Genet* **53**, 1415-1424, doi:10.1038/s41588-021-00931-x (2021).

86 Giugliano, R. P. *et al.* Edoxaban versus warfarin in patients with atrial fibrillation. *N Engl J Med* **369**, 2093-2104, doi:10.1056/NEJMoa1310907 (2013).

87 O'Donoghue, M. L. *et al.* Effect of darapladib on major coronary events after an acute coronary syndrome: the SOLID-TIMI 52 randomized clinical trial. *JAMA* **312**, 1006-1015, doi:10.1001/jama.2014.11061 (2014).

88 Scirica, B. M. *et al.* Saxagliptin and cardiovascular outcomes in patients with type 2 diabetes mellitus. *N Engl J Med* **369**, 1317-1326, doi:10.1056/NEJMoa1307684 (2013).

89 Bonaca, M. P. *et al.* Long-term use of ticagrelor in patients with prior myocardial infarction. *N Engl J Med* **372**, 1791-1800, doi:10.1056/NEJMoa1500857 (2015).

90 Sabatine, M. S. *et al.* Evolocumab and Clinical Outcomes in Patients with Cardiovascular Disease. *N Engl J Med* **376**, 1713-1722, doi:10.1056/NEJMoa1615664 (2017).

91 Marston, N. A. *et al.* Clinical Application of a Novel Genetic Risk Score for Ischemic Stroke in Patients With Cardiometabolic Disease. *Circulation* **143**, 470-478, doi:10.1161/circulationaha.120.051927 (2021).

92 Marston, N. A. *et al.* Predicting Benefit From Evolocumab Therapy in Patients With Atherosclerotic Disease Using a Genetic Risk Score: Results From the FOURIER Trial. *Circulation* **141**, 616-623, doi:10.1161/circulationaha.119.043805 (2020).

ACKNOWLEDGEMENTS

Detailed acknowledgements are included in the **Supplementary Appendix**. We thank participants and staff of contributing studies.

AUTHOR CONTRIBUTIONS

S.D., M.D., Y.K., and L.M. jointly supervised research. A.M., R.M., T.H., and T.J. contributed equally. S.D. and M.D. designed and conceived the study. D.C., M.F., M.N., S.N., T.K., Y.O., J.Q.T., R.F-S., S.T., J.B., T.B., K.W., M.R., Y-H.J., B.W., S.B., H.L., M.N., C.Y., A.M., S.R., J.R., M.C., F.K., T.H., Y.S., A.S., G.C., A.K., D.S., Q.Y., F.V., J.L., A.C., N.H., T.J., K.K., K.L., J.C-M., N.P.T-A., R.M., M.G., J.H., E.Y-D., M.S., Y.H., M.K., A.M., Q.L-G., A.C., M.v.V., R.W., K.L., contributed to bioinformatics analyses. Y.R., M.B, C.A., D.W., P.R., T.M., K.C., B.N., A.T-H., R.F-S., J.W.J, M.E., T.B., K.W., M.J., F-E.d.L., P.L., M.R., L.F., P.F., C.J., K.K., H.H., Y-H.J., C.J., J.K., R.V., B.W., S.B., E.C.S., J-A.Z., H.S.M., N.G., M.C., G.P., M.O'D., N.M., F.K., M.S., C.R., K.T., M.S., K.P., D.S., J.H., S.S., A.H., L.L., G.G., N.S., D-A.T., R.S., T.R., H.A., M.A.I., P.H., K-G.H., F.M., V.A., R.Z., S.W-S., M.I., S.K., B.M., H.X., J.C., C-O.S., L.M., J.R., D.S., R.d.C., J.H., T.N., T.A., M.K., T.K., S.T., M.D., C.G., A.P., T.M., T.M., M.K., Y.K., S.D., I.R., M.I., N.A., I.M. contributed samples and phenotyping. S.D., M.D., Y.K., L.M., A.M., R.M., T.H., T.J., Y.H., M.K., M.S., K.K., and S.N. wrote and edited the manuscript. All authors provided critical revision.

COMPETING INTERESTS

C.A. has received sponsored research support from Bayer AG, and has consulted for ApoPharma; T.K. is an employee of JAPAN TOBACCO INC; M.E. reports grants from Bayer and fees paid to the Charité from AstraZeneca, Bayer, Boehringer Ingelheim, BMS, Daiichi Sankyo, Amgen, GSK, Sanofi, Covidien, Novartis, Pfizer, all outside the submitted work; B.P. serves on the Steering Committee of the Yale Open Data Access Project funded by Johnson & Johnson; P.A. works with Fondation Alzheimer (non profit foundation) and Genoscreen (Biotech Company); H.L.L's participation in this project was part of a competitive contract awarded to Data Tecnica International LLC by the National Institutes of Health to support open science research; M.A.N.'s participation in this project was part of a competitive contract awarded to Data Tecnica International LLC by the National Institutes of Health to support open science research, he also

currently serves on the scientific advisory board for Clover Therapeutics and is an advisor to Neuron23 Inc; N.M. declares institutional research grants to the TIMI Study Group at Brigham and Women's Hospital from Amgen, Pfizer, Ionis, Novartis, AstraZeneca, and NIH . The TIMI Study Group has received institutional research grant support through Brigham and Women's Hospital from Abbott, Amgen, Anthos Therapeutics, ARCA Biopharma, Inc., AstraZeneca, Daiichi-Sankyo, Eisai, Intarcia, Ionis Pharmaceuticals, Inc., MedImmune, Merck, Novartis, Pfizer, Regeneron Pharmaceuticals, Inc., Roche, The Medicines Company, Zora Biosciences, Janssen Research and Development, LLC, Siemens Healthcare Diagnostics, Inc., Softcell Medical Limited; M.S. has consultancies with Althera, Amgen, Anthos Therapeutics, AstraZeneca, Beren Therapeutics, Bristol-Myers Squibb, DalCor, Dr. Reddy's Laboratories, Fibrogen, IFM Therapeutics, Intarcia, Merck, Moderna, Novo Nordisk, Silence Therapeutics and research grant support through Brigham and Women's Hospital from Abbott, Amgen, Anthos Therapeutics, AstraZeneca, Bayer, Daiichi-Sankyo, Eisai, Intarcia, Ionis, Medicines Company, MedImmune, Merck, Novartis, Pfizer, Quark Pharmaceuticals. C.R. has consultancies with Anthos, Bayer, Bristol Myers Squibb, Boehringer Ingelheim, Daiichi Sankyo, Janssen, Pfizer. Institutional research grant to the TIMI Study Group at Brigham and Women's Hospital from Anthos, AstraZeneca, Boehringer Ingelheim, Daiichi Sankyo, Janssen, National Institutes of Health, Novartis. Consultancies with Anthos, Bayer, Bristol Myers Squibb, Boehringer Ingelheim, Daiichi Sankyo, Janssen, Pfizer. T.H. receives personal fees from Genome Analytics Japan, Inc.; J.C.H is supported by a personal fellowship from the British Heart Foundation (FS/14/55/30806), and acknowledges additional support from the Nuffield Department of Population Health (NDPH), University of Oxford, the British Heart Foundation Centre for Research Excellence, Oxford, and the Oxford Biomedical Research Centre. JCH holds Steering Committee and DSMB positions for various cardiovascular randomized controlled trials, and is a PI/co-PI of research grants from industry related to cardiovascular clinical trials and observational studies that are governed by University of Oxford contracts that protect personal independence. NDPH also has a staff policy of not taking personal payments from industry. Further details can be found at https://www.ndph.ox.ac.uk/files/about/ndph-independence-of-research-policy-jun-20.pdf/@@download; S.S. has consultancies with Biogen; P.H. reports grants from German Ministry of Research and Education, during the conduct of the study; research grants from German Ministry of Research and Education, European Union, Charité – Universitätsmedizin Berlin, Berlin

Chamber of Physicians, German Parkinson Society, University Hospital Würzburg, Robert Koch Institute, German Heart Foundation, Federal Joint Committee (G-BA) within the Innovationfond, German Research Foundation, Bavarian State (ministry for science and the arts), German Cancer Aid, Charité – Universitätsmedizin Berlin (within Mondafis; supported by an unrestricted research grant to the Charité from Bayer), University Göttingen (within FIND-AF randomized; supported by an unrestricted research grant to the University Göttingen from Boehringer-Ingelheim), University Hospital Heidelberg (within RASUNOA-prime; supported by an unrestricted research grant to the University Hospital Heidelberg from Bayer, BMS, Boehringer-Ingelheim, Daiichi Sankyo), outside the submitted work; K.G.H. reports a study grant by Bayer, lecture fees/advisory board fees from Abbott, Alexion, AMARIN, AstraZeneca, Bayer, Biotronik, Boehringer Ingelheim, Bristol-Myers-Squibb, Daiichi Sankyo, Edwards Lifesciences, Medtronic, Pfizer, Premier Research, SUN Pharma and W. L. Gore & Associates; H.J.G has received travel grants and speakers honoraria from Fresenius Medical Care, Neuraxpharm, Servier and Janssen Cilag as well as research funding from Fresenius Medical Care; J.H. is full time employee of Novo Nordisk; E.Y-D. is full time employee of Novo Nordisk

DATA AVAILABILITY

Summary statistics for the GWAS meta-analysis of stroke will be deposited in a public repository and made available by the time of publication. All other data supporting the findings of this study are available either within the article, the supplementary information and supplementary data files, or from the corresponding authors upon reasonable request.

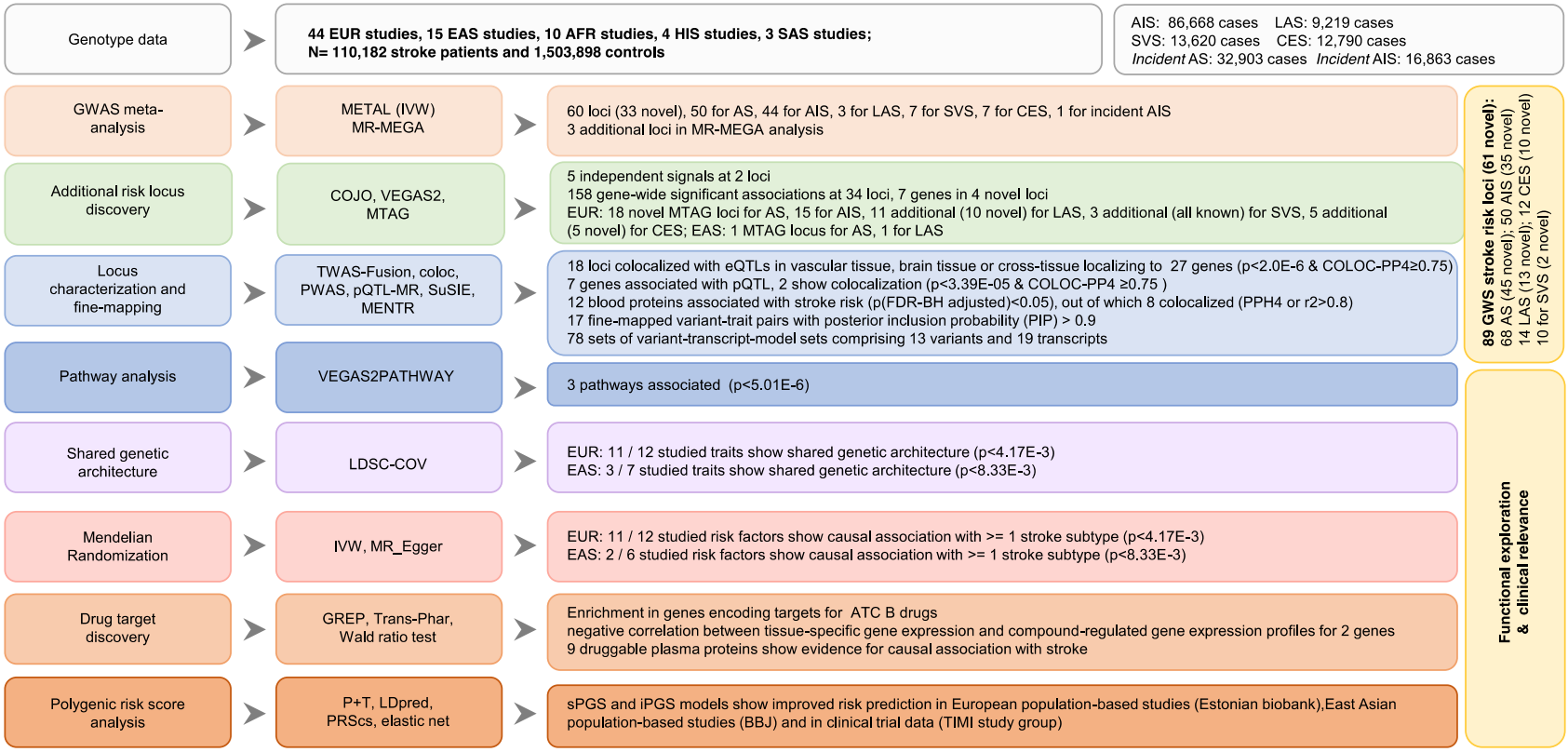
SUPPLEMENTARY MATERIAL

Supplementary Appendix (separate word document)

Supplementary Tables (separate excel file)

EXTENDED DATA FIGURES

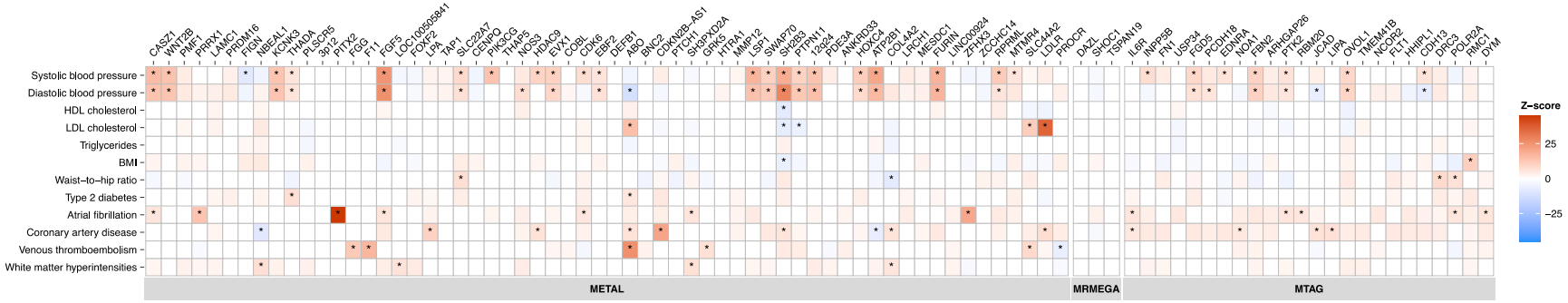
Extended Data Fig. 1: GIGASTROKE study workflow



Study workflow and rationale. EUR: European; EAS: East-Asian; AFR: African; HIS: Hispanic; SAS: South Asian; AS: any stroke; AIS: any ischemic stroke; LAS: large artery stroke; CES: cardioembolic stroke; SVS: small vessel stroke; GWAS: genome-wide association study; IVW: inverse-variance weighted; MR-MEGA: meta-regression of multi-ethnic genetic association; COJO:conditional and joint analysis; VEGAS2:versatile gene-based association study 2; MTAG: multi-trait analysis of GWAS; TWAS: Transcriptome-wide association study ; coloc: Colocalisation Test; PWAS: Proteome-wide association studies;pQTL-MR:

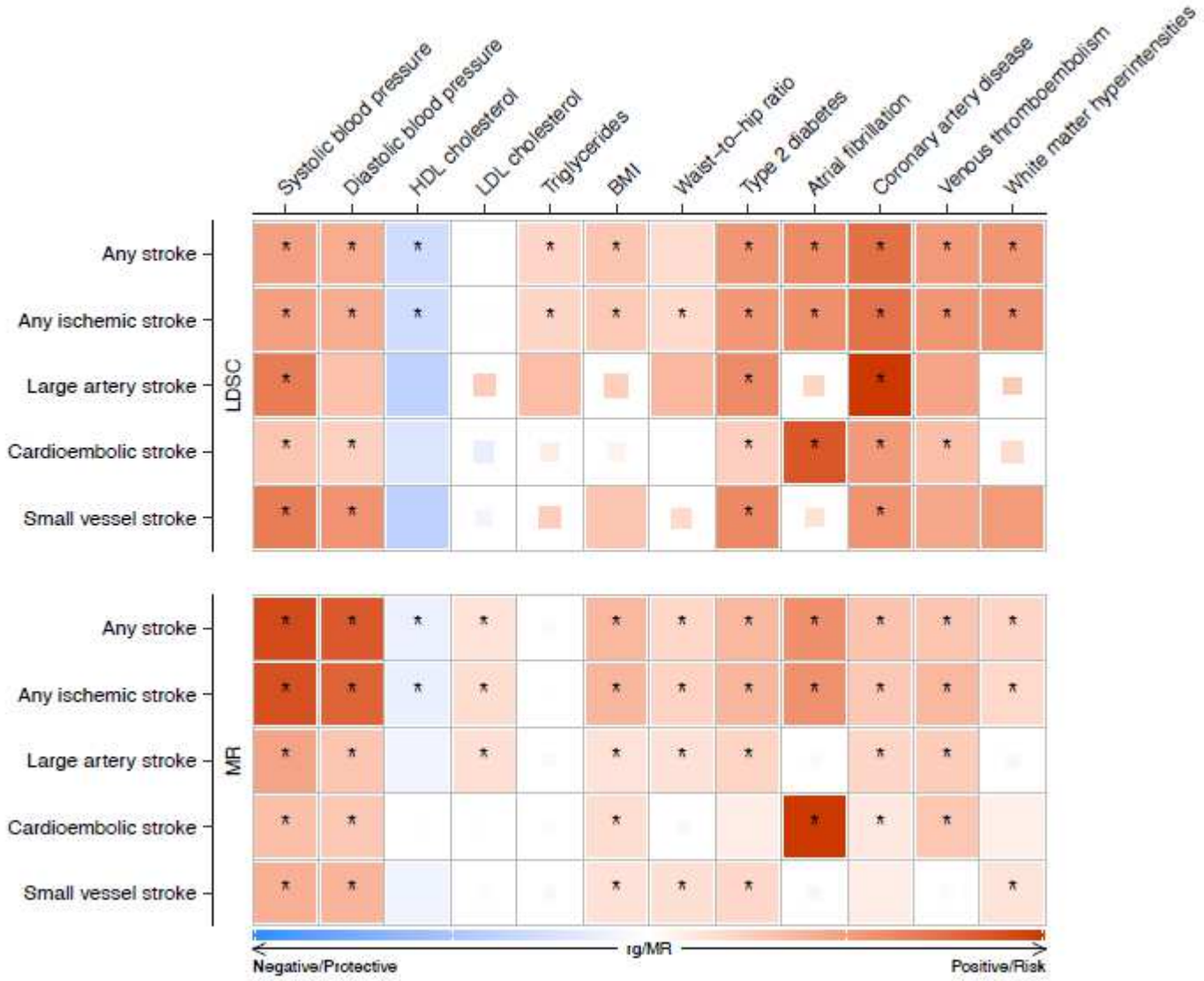
protein quantitative trait loci Mendelian Randomization; SuSIE: sum of single effects model; MENTR; PIP: posterior probability; FDR: false discovery rate; LDSC-COV: covariate-adjusted LD score regression; MR-Egger: Mendelian randomization-Egger; GREP: genome for REPositioning drugs; ATC: Anatomical Therapeutic Chemical; P+T: pruning and thresholding; PRScs: polygenic risk score under continuous shrinkage; BBJ: Biobank Japan; TIMI: thrombolysis in myocardial infarction

Extended Data Fig. 2: Association of stroke risk variants with vascular risk traits



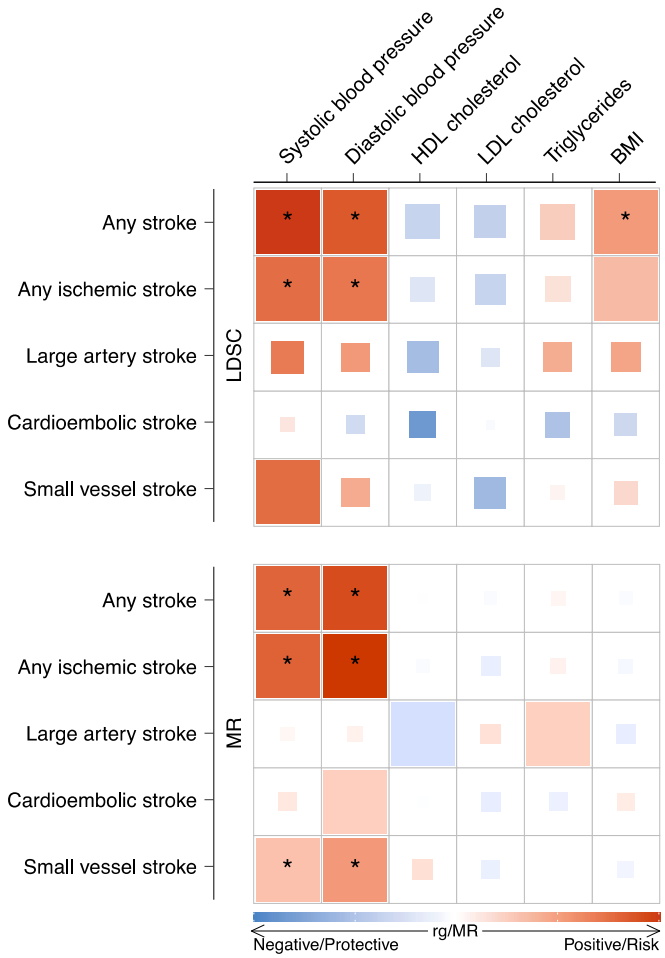
We report only associations for which the stroke lead variant of a proxy in very high LD ($r^2 > 0.9$) showed genome-wide significant association with the vascular risk trait in a prior GWAS. Colors represent the Z-scores of association of stroke risk increasing alleles with the trait

Extended Data Fig. 3: Genetic correlations and Mendelian randomization (MR) causal estimates of 12 vascular risk factor and disease traits with stroke (any and stroke subtypes), in European ancestry participants



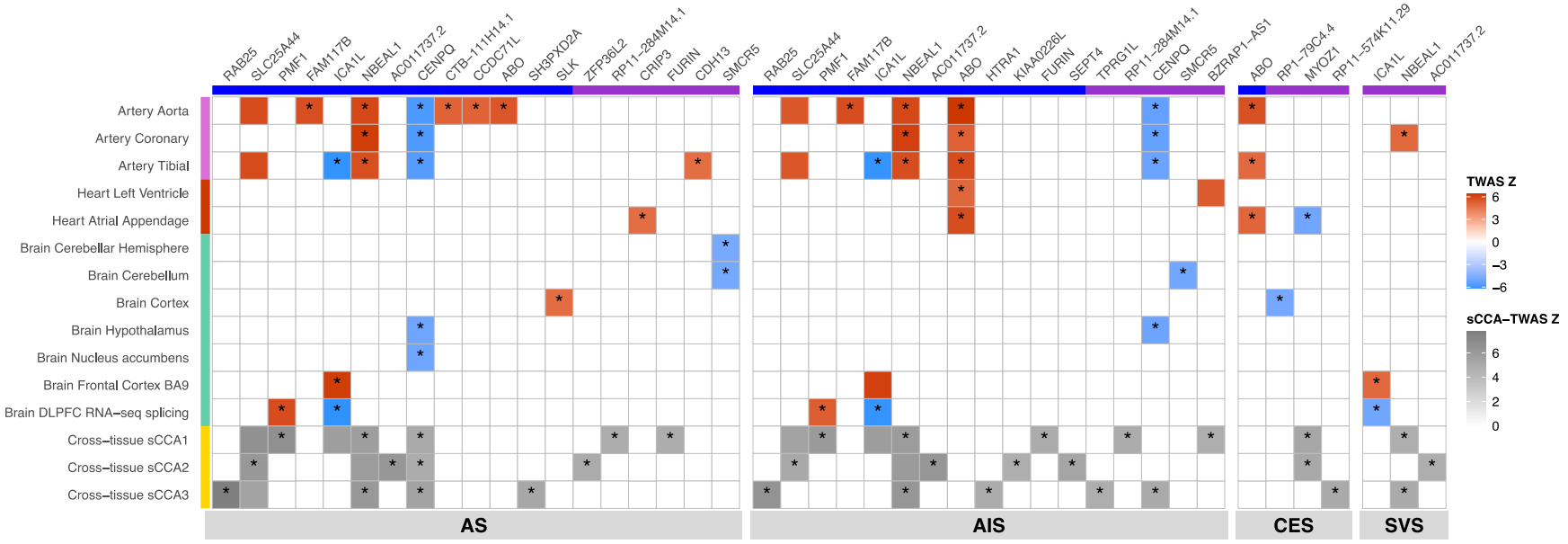
Larger squares correspond to more significant P-values, with genetic correlations or MR causal estimates (expressed in Z-scores) significantly different from zero at a $P<0.05$ shown as a full-sized square. Genetic correlations or causal estimates that are significant after multiple testing Bonferroni correction ($P<4.17\times10^{-3}$) are marked with an asterisk. Two-sided P-values were calculated using LD score regression for genetic correlations and inverse variance weighted analysis for MR.

Extended Data Fig. 4: Genetic correlations and Mendelian randomization (MR) causal estimates of 6 vascular risk factor and disease traits with stroke (any and stroke subtypes), in East-Asian ancestry participants



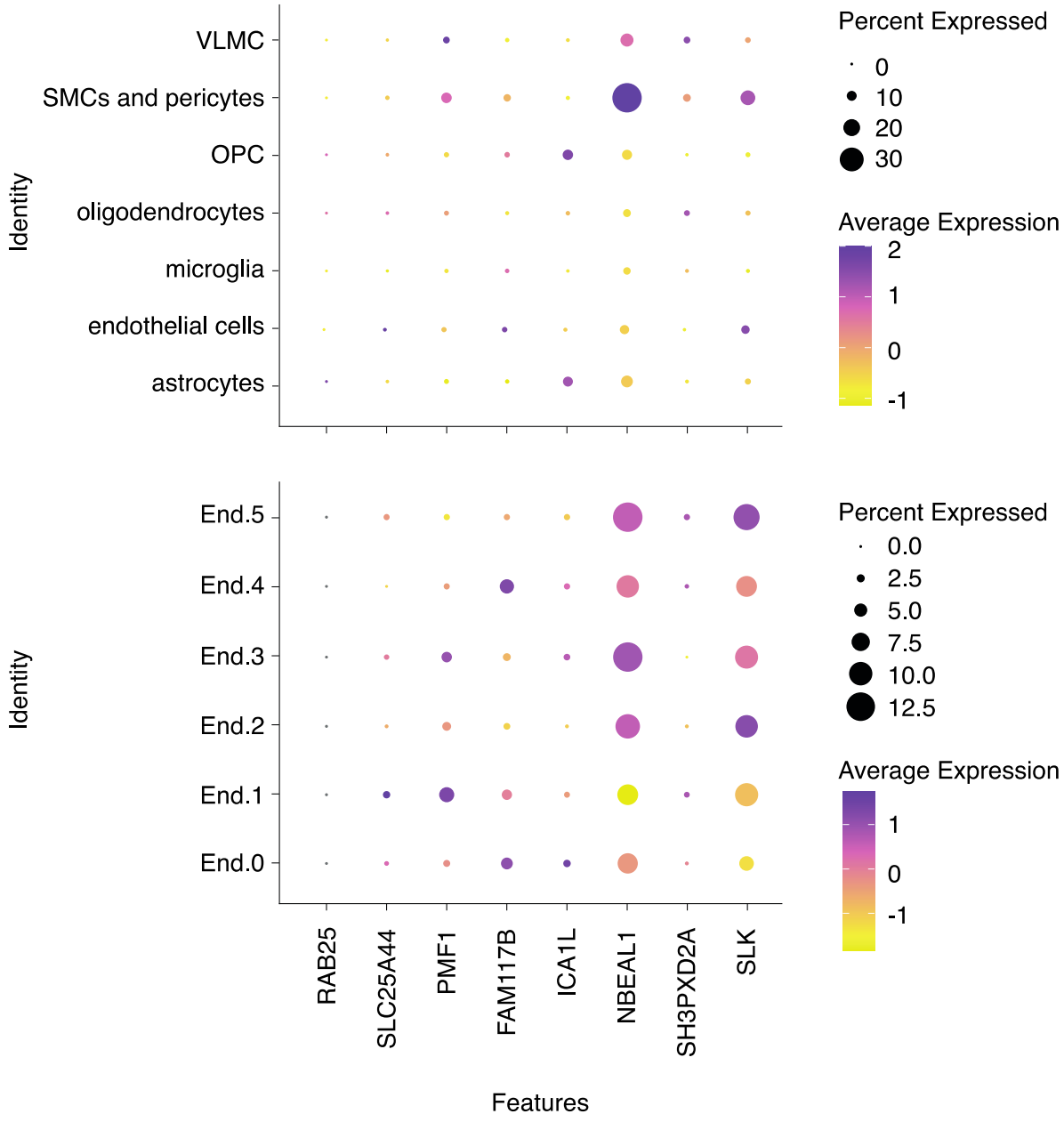
Larger squares correspond to more significant P-values, with genetic correlations or MR causal estimates significantly different from zero at a P<0.05 shown as a full-sized square. Genetic correlations or causal estimates (expressed in Z-scores) that are significant after multiple testing Bonferroni correction ($P<8.33x10^{-3}$) are marked with an asterisk. Two-sided P-values were calculated using LD score regression for genetic correlations and inverse variance weighted analysis for MR.

Extended Data Fig. 5: Transcriptome-wide association study (TWAS) of stroke in multiple tissues



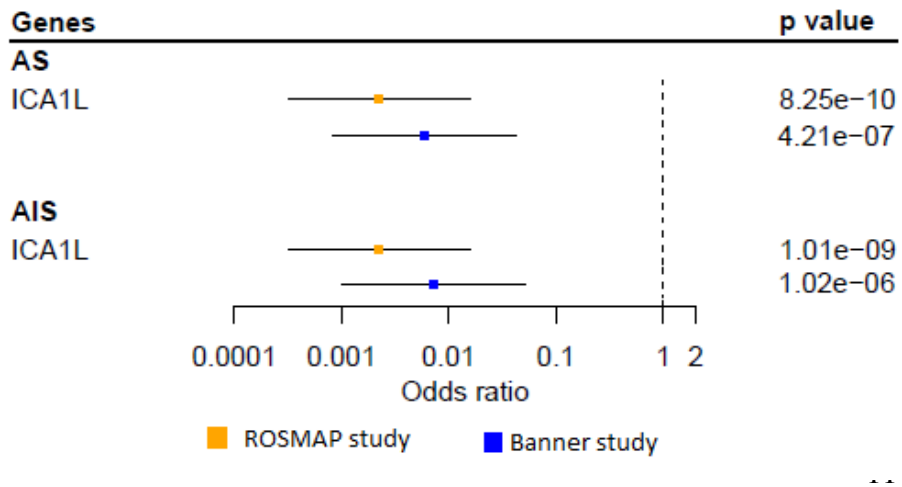
Heatmap of the transcriptome-wide association studies of stroke (any stroke and stroke subtypes) reaching transcriptome wide significance and colocalized in GIGASTROKE; Colored squares are TWAS significant ($p<2.0x10^{-6}$); * Conditionally significant ($p<0.05$) and COLOC PP4 ≥ 0.75; Genes are presented on the x-axis, those underlined in blue are in a GWAS locus, those underlined in purple are not within a genome-wide significant stroke risk locus (**Methods**); Tissue types are on the y-axis (blue: cross-tissue weights; pink: arterial; orange: heart; green: brain)

Extended Data Fig. 6: Single-cell gene expression data of TWAS-COLOC genes



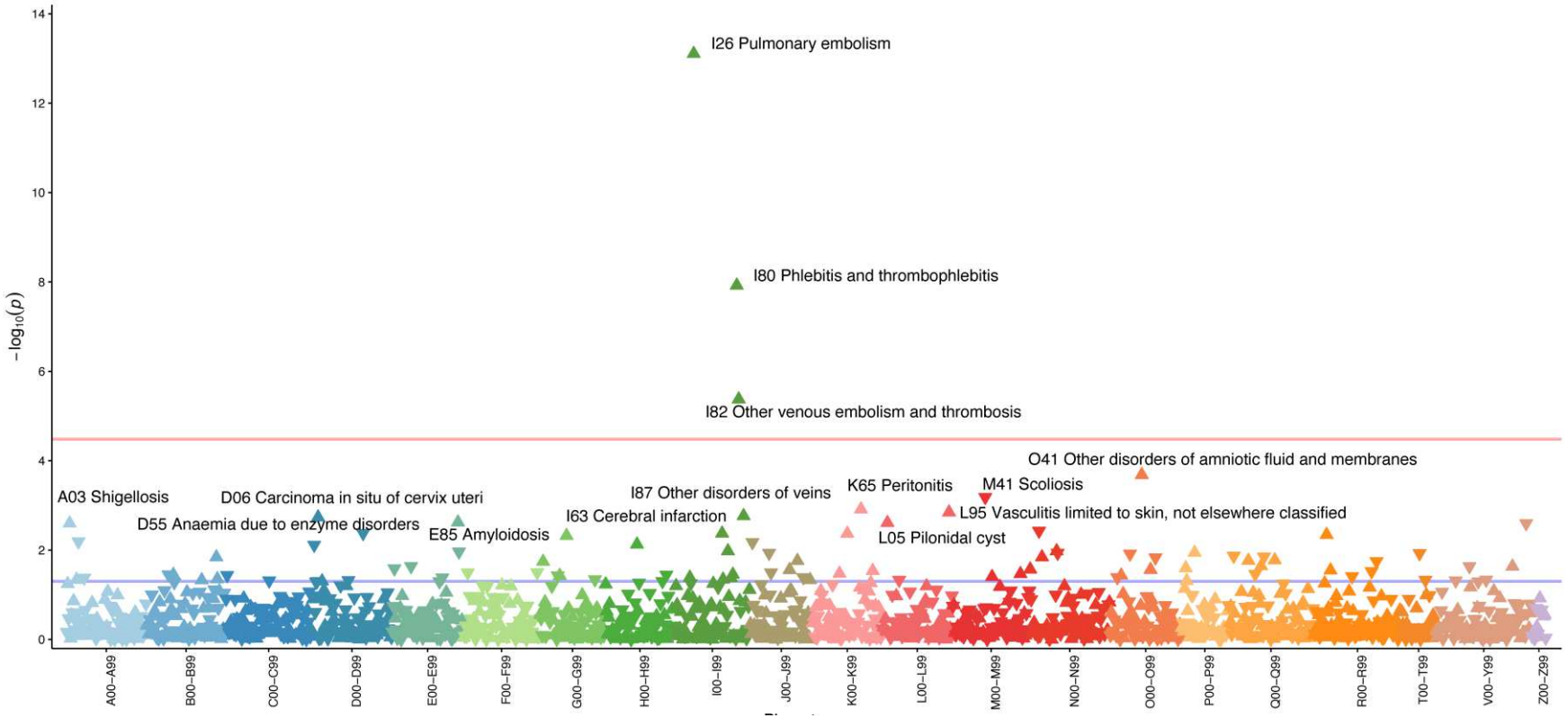
Dot plot of the mean expression level in expressing cells (color) and percent of expressing cells (circle size) of selected genes across different cell types (top) and endothelial subsets (bottom).

Extended Data Fig. 7: Proteome-wide association study (PWAS) of stroke in brain tissue



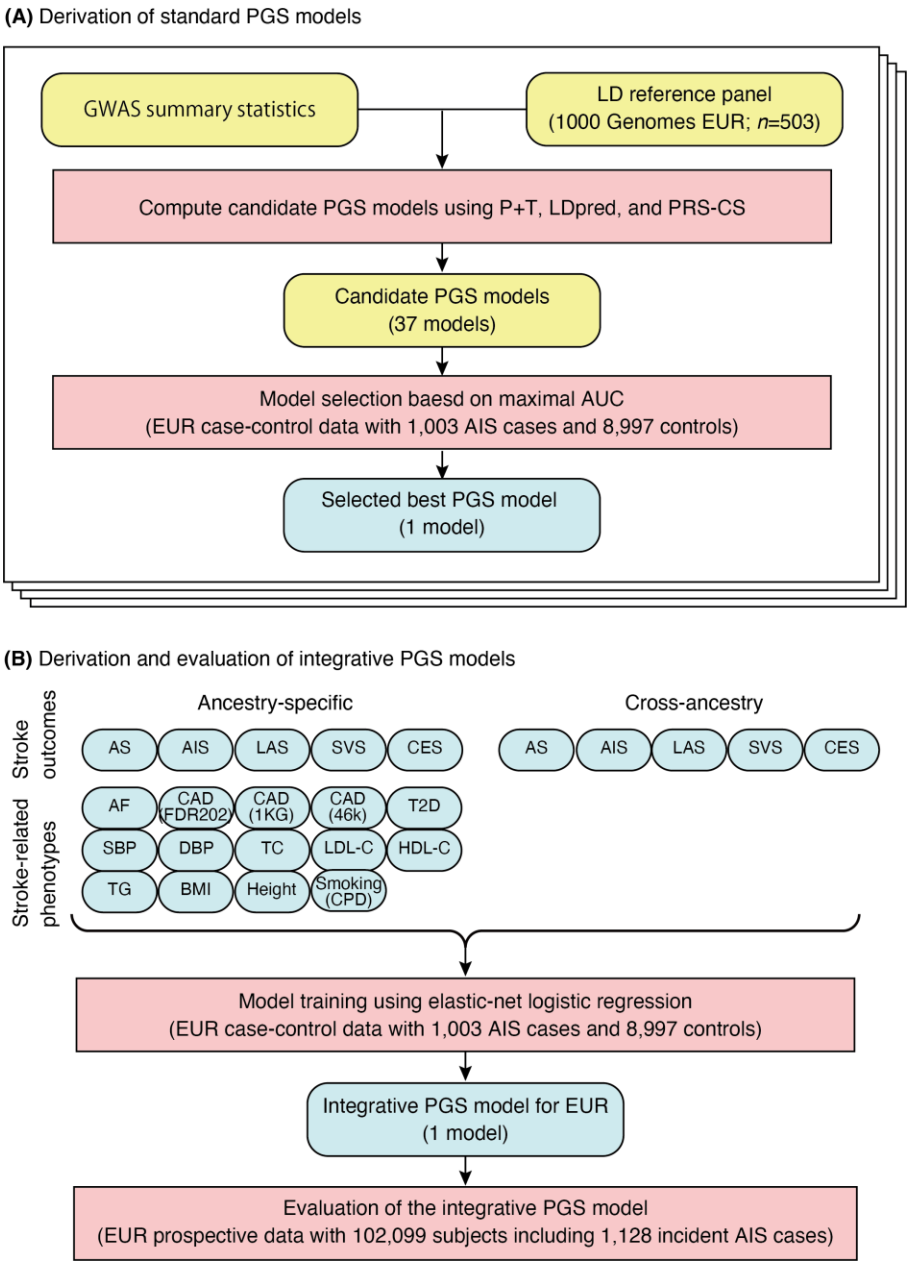
Association of ICA1L protein abundance in dorsolateral prefrontal cortex with risk of AS and AIS, using proteome-wide association studies and colocalization.

Extended Data Fig. 8: Drug target pQTL PheWAS



PheWAS in Estonian biobank for pQTL of drug targets identified as being putative drug targets for stroke in the Mendelian randomization analysis, for which associations reached phenome-wide significance ($p=6.91\times10^{-6}$): top, PheWAS for rs2289252, a cis-pQTL for F11. Each triangle in the plot represents one ICD10 main code and the direction of the triangle represents direction of effect.

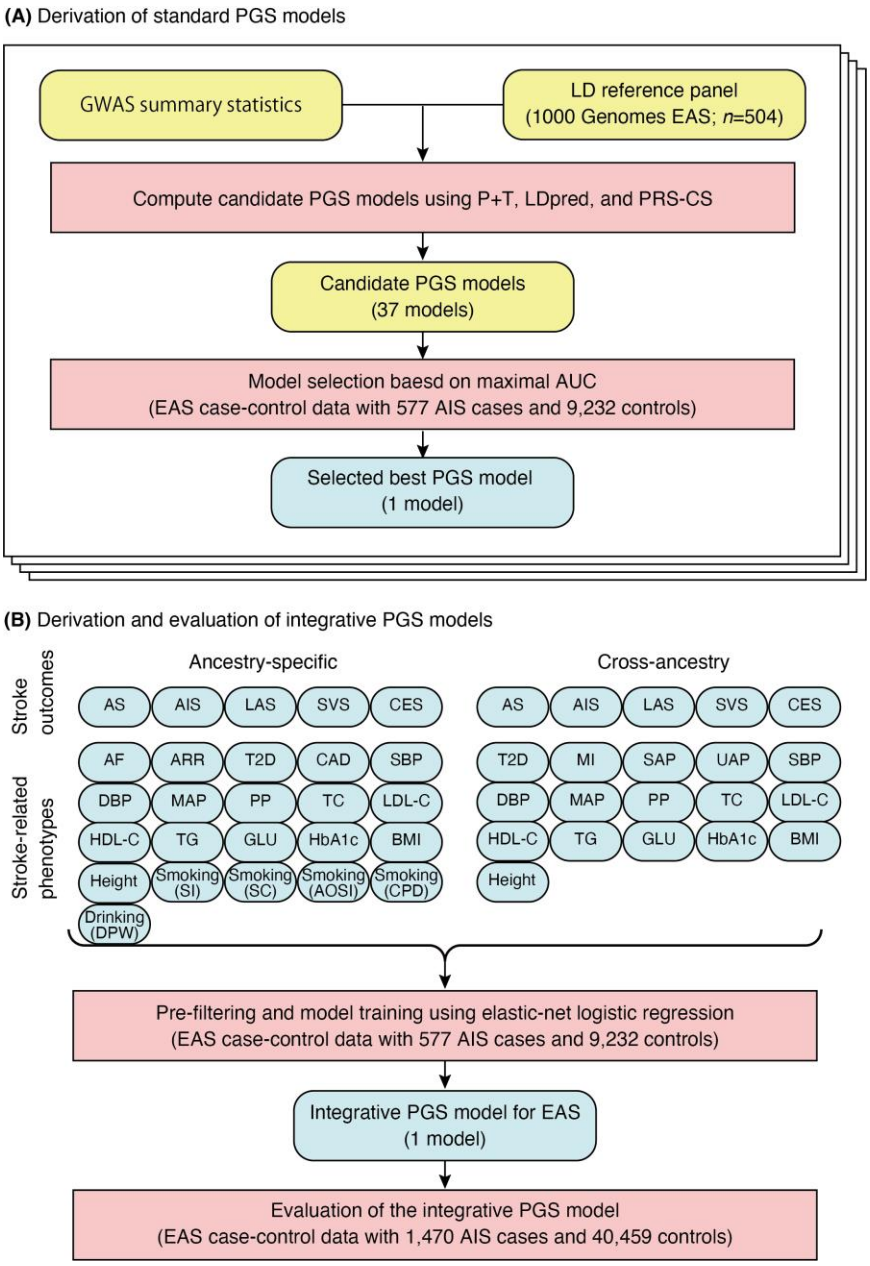
Extended Data Fig. 9: Derivation and evaluation of an integrative polygenic score models for Europeans



The integrative PGS (iPGS) model for Europeans was derived from 10 GIGASTROKE GWASs and 14 GWASs of vascular risk traits. (A) From the genome-wide summary statistics for each GWAS and a linkage disequilibrium (LD) reference panel of the European subjects (n=503) from the 1000 Genomes Project, 37 candidate PGS models were computed using P+T, LDpred, and

PRScs algorithms. Then, the best PGS model was selected for each GWAS, where the best model was defined as the model that showed the maximal area under the curve (AUC) in the model training dataset (a European case-control data with 1,003 ischemic stroke [IS] cases and 8,997 controls). (B) The 24 selected PGS models derived from the 24 GWASs were used as the variables for elastic-net logistic regression and the weights for the variables were trained using the model training dataset. By combining the 24 PGS models using the weights, the iPGS model consisting of 7,010,016 variants was constructed. The iPGS model was evaluated in the model evaluation dataset (a European prospective cohort data with 102,099 subjects including 1,128 incident IS cases); AS indicates any stroke; AIS, any ischemic stroke; LAS, large artery stroke; SVS, small vessel stroke; CES, cardioembolic stroke; AF, atrial fibrillation; CAD, coronary artery disease; T2D, type 2 diabetes; SBP, systolic blood pressure; DBP, diastolic blood pressure; TC, total cholesterol; LDL-C, low-density lipoprotein cholesterol; HDL-C, high-density lipoprotein cholesterol; TG, triglyceride; BMI, body mass index; CPD, cigarettes per day; AUC indicates area under the curve; EUR, Europeans; GWAS, genome-wide association study; LD, linkage disequilibrium; PGS, polygenic score

Extended Data Fig. 10: Derivation and evaluation of an integrative polygenic score models for East-Asians



The integrative PGS (iPGS) model for East-Asians was derived from 10 GIGASTROKE GWASs and 37 GWASs of vascular risk traits. (A) From the genome-wide summary statistics for each GWAS and a linkage disequilibrium (LD) reference panel of the East-Asian subjects (n=504) from the 1000 Genomes Project, 37 candidate PGS models were computed using P+T, LDpred,

and PRScs algorithms. Then, the best PGS model was selected for each GWAS, where the best model was defined as the model that showed the maximal area under the curve (AUC) in the model training dataset (an East-Asian case-control data with 577 ischemic stroke [IS] cases and 9,232 controls). (B) Among the 47 selected PGS models derived from the 47 GWASs, 12 were significantly associated with IS in the model training dataset (Bonferroni-corrected $P<0.05$). The significant PGS models were used as the variables for elastic-net logistic regression and the weights for the variables were trained using the model training dataset. By combining the 12 significant PGS models using the weights, the iPGS model consisting of 8,544,464 variants was constructed. The iPGS model was evaluated in the model evaluation dataset (an East-Asian case-control data with 1,470 IS cases and 40,459 controls).

AS indicates any stroke; AIS, any ischemic stroke; LAS, large artery stroke; SVS, small vessel stroke; CES, cardioembolic stroke; AF, atrial fibrillation; ARR, Arrhythmia; T2D, type 2 diabetes; CAD, coronary artery disease; SBP, systolic blood pressure; DBP, diastolic blood pressure; MAP, mean arterial pressure; PP, pulse pressure; TC, total cholesterol; LDL-C, low-density lipoprotein cholesterol; HDL-C, high-density lipoprotein cholesterol; TG, triglyceride; GLU, glucose; BMI, body mass index; SI, smoking initiation; SC; smoking cessation; AOSI, age of smoking initiation; CPD, cigarettes per day; DPW, drinks per week; MI, myocardial infarction; SAP, stable angina pectoris; UAP, unstable angina pectoris; AUC indicates area under the curve; EAS, East-Asian; GWAS, genome-wide association study; LD, linkage disequilibrium; PGS, polygenic score

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryAppendixGIGASTROKE15122021.pdf
- ExtendedDataFig.6.pdf
- ExtendedDataFig.4.pdf
- ExtendedDataFig.2.pdf
- GIGASTROKESuppl.Tables151221.xlsx
- ExtendedDataFig.9.png
- ExtendedDataFig.1.pdf
- ExtendedDataFig.3.pdf
- ExtendedDataFig.5.pdf
- ExtendedDataFig.7.bmp
- ExtendedDataFig.10.png
- ExtendedDataFig.8.tif