

## Genetic Contributions to Early and Late Onset Ischemic Stroke

Authors: Thomas Jaworek, MS; Huichun Xu, MD, PhD; Brady J. Gaynor, MS; John W. Cole, MD, MS; Kristiina Rannikmäe, MD, PhD; Tara M. Stanne, PhD; Liisa Tomppo, MD, PhD; Vida Abedi, PhD; Philippe Amouyel, MD, PhD; Nicole D. Armstrong, PhD, MS; John Attia, PhD, FRCPC, FRACP, Clin Dip Pall Med; Steven Bell, PhD; Oscar R. Benavente, MD; Giorgio B. Boncoraglio, MD; Adam Butterworth, PhD; the Cervical Artery Dissections and Ischemic Stroke Patients (CADISP) Consortium; Jara Cárcel-Márquez, MSc; Zhengming Chen, DPhil; Michael Chong, MSc; Carlos Cruchaga, PhD; Mary Cushman, MD, MSc; John Danesh, FMedSci, DPhil; Stephanie Durette, MD, PhD; David J. Duggan, PhD; Jon Peter Durda, PhD; Gunnar Engstrom, MD, PhD; Christian Enzinger, MD; Jessica D. Faul, PhD, MPH; Natalie S. Fecteau, MPH; Israel Fernández-Cadenas, PhD; Christian Geiger, PhD; Anne-Katrin Giese, MD; Raji P. Grewal, MD; Ulrike Grittner, PhD; Aki S. Havulinna, DSc; Laura Heitsch, MD; Marc C. Hochberg, MD; Elizabeth Holliday, BSc, MSc, PhD; Jie Hu, MD, PhD; Andreea Ilinca, MD, PhD; the International Stroke Genetics Consortium (ISGC); the INVENT Consortium; Marguerite R. Irvin, PhD; Rebecca D. Jackson, MD; Mina A. Jacob, MD; Raquel Rabionet Janssen, PhD; Jordi Jimenez-Conde, MD, PhD; Julie A. Johnson, PharmD; Yoichiro Kamatani, MD, PhD; Sharon L.R. Kardia, PhD; Masaru Koido, PhD; Michiaki Kubo, MD, PhD; Leslie Lange, PhD; Jin-Moo Lee, MD, PhD; Robin Lemmens, MD, PhD; Christopher Levi, MBBS, BMed Sci, FRACP, FAHMS; Jiang Li, PhD; Liming Li, MPH; Kuang Lin, PhD; Haley Lopez, BS; Sothear Luke, MPH; Jane Maguire, PhD BNurs(Hons), BA, RN; Patrick F. McArdle, PhD; Caitrin W. McDonough, PhD; James F. Meschia, MD; Tiina Metso, MD, PhD; Martina Müller-Nurasyid, PhD; Timothy O'Connor, PhD; Martin O'Donnell, MB, PhD; Leema Reddy Peddareddygar, MBBS, MD; Joanna Pera, MD, PhD; James A. Perry, PhD; Annette Peters, PhD; Jukka Putaala, MD, PhD; Debashree Ray, BSc, MStat, PhD; Kathryn Rexrode, MD, MPH; Marta Ribases, PhD, BSc; Jonathan Rosand, MD, MSc; Peter M. Rothwell, MD, PhD, FMedSci; Tatjana Rundek, MD, PhD, FANA; Kathleen A. Ryan, MPH; Ralph L. Sacco, MD, MS, FAHA, FAAN, FANA; Veikko Salomaa, MD, PhD; Cristina Sanchez-Mora, PhD, BSc; Reinhold Schmidt, MD; Pankaj Sharma, MD, PhD, FRCP; Agnieszka Slowik, MD, PhD; Jennifer A. Smith, PhD, MPH; Nicholas L. Smith, PhD, MPH; Sylvia Wassertheil-Smoller, PhD; Martin Soderholm, MD, PhD; O. Colin Stine, PhD; Daniel Strbian, MD, PhD, MSc; Cathie LM Sudlow, DPhil, FRCP(E); Turgut Tatlisumak, MD, PhD; Chikashi Terao, MD, PhD; Vincent Thijs, MD, PhD; Nuria P. Torres-Aguila, PhD; David-Alexandre Trégouët, PhD; Anil Man Tuladhar, MD, PhD; Jan H. Veldink, MD, PhD; Robin G. Walters, PhD; David R. Weir, PhD; Daniel Woo, MD, MS; Bradford B. Worrall, MD, MSc; Charles C Hong, MD, PhD; Owen Ross, PhD; Ramin Zand, MD, MPH; F-E de Leeuw, MD, PhD; Arne G. Lindgren, MD, PhD; Guillaume Pare, MD, M.Sc., FRCPc; Christopher D. Anderson, MD, MMSc; Hugh S. Markus, MD; Christina Jern, MD, PhD; Rainer Malik, PhD; Martin Dichgans, MD; Braxton D. Mitchell, PhD, MPH; Steven J. Kittner, MD, MPH; International Stroke Genetics Consortium; Early Onset Stroke Genetics Consortium.

Thomas Jaworek, Division of Endocrinology, Diabetes and Nutrition, University of Maryland School of Medicine, Baltimore, MD, USA

Huichun Xu, Division of Endocrinology, Diabetes and Nutrition, Department of Medicine, University of Maryland School of Medicine, Baltimore, MD USA

Brady J. Gaynor, Division of Endocrinology, Diabetes and Nutrition, University of Maryland School of Medicine, Baltimore, MD, USA

John W. Cole, VA Maryland Health Care System, Baltimore, MD, USA  
Department of Neurology, University of Maryland School of Medicine, Baltimore, MD, USA

Braxton D. Mitchell, Division of Endocrinology, Diabetes and Nutrition, University of Maryland School of Medicine, Baltimore, MD, USA; Geriatric Research and Education Clinical Center, Veterans Administration Medical Center, Baltimore, MD, USA

Kristiina Rannikmäe, Centre for Medical Informatics, Usher Institute, University of Edinburgh, Edinburgh, UK

Tara M. Stanne, Institute of Biomedicine, Department of Laboratory Medicine, Sahlgrenska Academy, University of Gothenburg, Sweden

Liisa Tomppo, Department of Neurology, Helsinki University Hospital and University of Helsinki, Helsinki, Finland.

Vida Abedi, Department of Molecular and Functional Genomics, Geisinger Health System, Danville, PA, USA

Philippe Amouyel, LabEx DISTALZ-U1167, RID-AGE-Risk Factors and Molecular Determinants of Aging-Related Diseases, University of Lille, Lille, France; Inserm U1167, Lille, France; Centre Hospitalier Universitaire Lille, Lille, France; Institut Pasteur de Lille, Lille, France

Nicole D. Armstrong, Department of Epidemiology, University of Alabama at Birmingham, Birmingham, AL, USA

John Attia, School of Medicine and Public Health, University of Newcastle and Hunter Medical Research Institute, Newcastle, Australia

Steven Bell, Stroke Research Group, Department of Clinical Neurosciences, University of Cambridge, Cambridge, UK

Oscar R. Benavente, Department of Neurology, University of British Columbia, Vancouver, British Columbia, Canada

Giorgio B. Boncoraglio, Department of Cerebrovascular Diseases, Fondazione IRCCS Istituto Neurologico 'Carlo Besta', Milan, Italy

Adam Butterworth, British Heart Foundation Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK; British Heart Foundation Centre of Research Excellence, University of Cambridge, Cambridge, UK; National Institute for Health Research Blood and Transplant Research Unit in Donor Health and Genomics, University of Cambridge, Cambridge, UK; Health Data Research UK Cambridge, Wellcome Genome Campus and University of Cambridge, Cambridge, UK

Jara Cárcel-Márquez, Stroke Pharmacogenomics and Genetics group, Biomedical Research Institute Sant Pau (IIB Sant Pau), Barcelona, Spain

Zhengming Chen, Nuffield Department of Population Health  
University of Oxford, UK; MRC Population Health Research Unit, University of Oxford, UK

Michael Chong, DBCVS Research Institute, Population Health Research Institute, McMaster University, Hamilton, Ontario, Canada; Thrombosis & Atherosclerosis Research Institute (TaARI)

Carlos Cruchaga, Department of Psychiatry  
Washington University School of Medicine, St. Louis, MO, USA

Mary Cushman, Department of Hematology and Oncology, University of Vermont, Medical Center, Colchester, USA

John Danesh, British Heart Foundation Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK; British Heart Foundation Centre of Research Excellence, University of Cambridge, Cambridge, UK; National Institute for Health Research Blood and Transplant Research Unit in Donor Health and Genomics, University of Cambridge, Cambridge, UK; Health Data Research UK Cambridge, Wellcome Genome Campus and University of Cambridge, Cambridge, UK; Department of Human Genetics, Wellcome Sanger Institute, Hinxton, UK

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

David J. Duggan, Quantitative Medicine and Systems Biology Division, Translational Genomics Research Institute, An Affiliate of City of Hope, Phoenix, AZ, USA

Jon Peter Durda, Laboratory for Clinical Biochemistry Research, Department of Pathology and Laboratory Medicine, Larner College of Medicine, University of Vermont, Colchester, VT, USA

Gunnar Engstrom, Department of Clinical Sciences, Malmö, Lund University, Sweden

Christian Enzinger, Department of Neurology, Medical University Graz, Austria

Jessica D. Faul, Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI, USA

Natalie S. Fecteau, Department of Neurology, University of Maryland School of Medicine, Baltimore, MD, USA

Israel Fernández-Cadenas, Stroke Pharmacogenomics and Genetics group, Biomedical Research Institute Sant Pau (IIB Sant Pau), Barcelona, Spain. Stroke Pharmacogenomics and Genetics, Fundacio Docència i Recerca MutuaTerrassa, Terrassa, Spain.

Christian Geiger, Institute of Epidemiology, Helmholtz Zentrum München German Research Center for Environmental Health, Neuherberg, Germany.  
Research Unit of Molecular Epidemiology, Institute of Epidemiology, Helmholtz Zentrum München German Research Center for Environmental Health, Ingolstaedter Landstraße 1, 85764, Neuherberg, Germany.

Anne-Katrin Giese, Klinik und Poliklinik für Neurologie, Kopf- und Neurozentrum, Universitätsklinikum Hamburg-Eppendorf, Germany

Raji P. Grewal, Neuroscience Institute, Saint Francis Medical Center, Trenton, New Jersey, USA

Ulrike Grittner, Department for Biostatistics and Clinical Epidemiology. Charité-University Medical Centre, Berlin, Germany

Aki S. Havulinna, National Institute for Health and Welfare, Helsinki, Finland.

Laura Heitsch, Departments of Emergency Medicine and Neurology, Washington University School of Medicine, St. Louis, MO, USA

Marc C. Hochberg, Division of Rheumatology and Clinical Immunology, Department of Medicine, and Department of Epidemiology and Public Health, University of Maryland School of Medicine, Baltimore, MD, USA

Elizabeth Holliday, School of Medicine and Public Health, University of Newcastle, Australia

Jie Hu, Division of Women's Health, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA; Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA

Andreea Ilinca, Department of Clinical Sciences Lund, Neurology, Lund University, Department of Neurology and Rehabilitation Medicine, Skane University Hospital, Lund, Sweden

Marguerite R. Irvin, Epidemiology, School of Public Health, University of Alabama at Birmingham, Birmingham, AL, USA

Rebecca D. Jackson, Division of Endocrinology, Diabetes and Metabolism, Department of Internal Medicine and the Center for Clinical and Translational Science, The Ohio State University, Columbus, OH, USA.

Mina A. Jacob, Department of Neurology, Radboud University Medical Center; Donders Medical Center for Neuroscience; Nijmegen, The Netherlands

Raquel Rabionet Janssen, Department of Genetics, Microbiology and Statistics, Institute of Biomedicine (IBUB), University of Barcelona, Barcelona, Spain; Institut de Recerca Sant Joan de Déu, Esplugues de Llobregat, Spain; Centro de investigación biomédica en red (CIBERER), Spain

Jordi Jimenez-Conde, Neurovascular Research Group (NEUVAS), Neurology Department, Institut Hospital del Mar d'Investigació Mèdica, Universitat Autònoma de Barcelona, Barcelona, Spain

Julie A. Johnson, Department of Pharmacotherapy and Translational Research and Center for Pharmacogenomics, University of Florida, College of Pharmacy, Gainesville, FL, USA  
Division of Cardiovascular Medicine, College of Medicine, University of Florida, Gainesville, FL, USA

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

Sharon L.R. Kardia, Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI USA

Masaru Koido, Department of Cancer Biology, Institute of Medical Science, The University of Tokyo, Tokyo, Japan.

Michiaki Kubo, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan

Leslie Lange, Department of Medicine, University of Colorado Denver, Anschutz Medical Campus, Aurora, CO, USA.

Jin-Moo Lee, Department of Neurology, Washington University School of Medicine, St. Louis, MO, USA

Robin Lemmens, KU Leuven – University of Leuven, Department of Neurosciences, Experimental Neurology; VIB Center; For Brain & Disease Research; University Hospitals Leuven, Department of Neurology, Leuven, Belgium.

Christopher Levi, John Hunter Hospital, Hunter Medical Research Institute and University of Newcastle, Newcastle, Australia and Priority Research Centre for Stroke & Brain Injury, University of Newcastle, NSW, Australia

Jiang Li, Department of Molecular and Functional Genomics, Geisinger Health System, Danville, PA, USA

Liming Li, Peking University Health Science Center, Department of Epidemiology and Biostatistics, Peking University, Beijing, China

Kuang Lin, Nuffield Department of Population Health, University of Oxford, Oxford, UK

Haley Lopez, Department of Neurology, University of Maryland School of Medicine, Baltimore, MD, USA

Sothear Luke, Department of Neurology, Mayo Clinic, Jacksonville, FL, USA.

Jane Maguire, Faculty of Health, School of Nursing and Midwifery, University of Technology Sydney, NSW, Australia

Patrick F. McArdle, Division of Endocrinology, Diabetes and Nutrition, Department of Medicine, University of Maryland School of Medicine, Baltimore, MD USA

Caitrin W. McDonough, Department of Pharmacotherapy and Translational Research and Center for Pharmacogenomics, University of Florida, College of Pharmacy, Gainesville, FL, USA

James F. Meschia, Department of Neurology, Mayo Clinic, Jacksonville, FL, USA.

Tiina Metso, Department of Neurology, Helsinki University Central Hospital, Helsinki, Finland

Martina Müller-Nurasyid, Institute of Genetic Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany Department of Medicine I, Ludwig-Maximilians-University Munich, Munich, Germany

Timothy O'Connor, Division of Endocrinology, Diabetes and Nutrition, Department of Medicine, University of Maryland School of Medicine, Baltimore, MD USA; Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, MD, USA

Martin O'Donnell, Health Research Board Clinical Research Facility, Geata an Eolais, National University of Ireland, Galway, Ireland

Leema Reddy Peddareddygari, Neuroscience Institute, Saint Francis Medical Center, Trenton, New Jersey, USA

Joanna Pera, Department of Neurology, Jagiellonian University, Botaniczna 3, 31-503 Krakow, Poland

James A. Perry, Division of Endocrinology, Diabetes and Nutrition, Department of Medicine, University of Maryland School of Medicine, Baltimore, MD USA

Annette Peters, Institute of Epidemiology, Helmholtz Zentrum München German Research Center for Environmental Health, Neuherberg, Germany. Institute for Medical Information Processing, Biometry and Epidemiology, Ludwig-Maximilians-University Munich, Munich, Germany.

Jukka Putaala, Department of Neurology, Helsinki University Hospital and University of Helsinki, Helsinki, Finland.

Debashree Ray, Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA

Kathryn Rexrode, Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA

Marta Ribases, Psychiatric Genetics Unit, Group of Psychiatry, Mental Health and Addiction, Vall d'Hebron Research Institute (VHIR), Universitat Autònoma de Barcelona, Barcelona, Spain. Department of Psychiatry, Hospital Universitari Vall d'Hebron, Barcelona, Spain. Biomedical Network Research Centre on Mental Health (CIBERSAM), Instituto de Salud Carlos III, Madrid, Spain. Department of Genetics, Microbiology, and Statistics, Faculty of Biology, Universitat de Barcelona, Barcelona, Spain.

Jonathan Rosand, McCance Center for Brain Health, Massachusetts General Hospital, Boston, MA, USA. Center for Genomic Medicine, MGH, Boston, MA, USA. Department of Neurology, MGH, Boston, MA, USA. Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, USA

Peter M. Rothwell, Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, UK

Tatjana Rundek, Department of Neurology and Evelin F. McKnight Brain Institute, Miller School of Medicine, University of Miami, Miami, FL, USA

Kathleen A. Ryan, Division of Endocrinology, Diabetes and Nutrition, Department of Medicine, University of Maryland School of Medicine, Baltimore, MD USA

Ralph L. Sacco, Department of Neurology and Evelin F. McKnight Brain Institute, Miller School of Medicine, University of Miami, Miami, FL, USA

Cristina Sanchez-Mora, Psychiatric Genetics Unit, Group of Psychiatry, Mental Health and Addictions, Vall d'Hebron Research Institute (VHIR), Universitat Autònoma de Barcelona, Barcelona, Spain. Department of Psychiatry, Hospital Universitari Vall d'Hebron, Barcelona, Spain; Biomedical Network Research Centre on Mental Health (CIBERSAM), Barcelona, Spain.

Reinhold Schmidt, Department of Neurology, Medical University of Graz, Graz, Austria

Pankaj Sharma, Institute of Cardiovascular Research, Royal Holloway University of London, London, UK, and Ashford and St. Peters Hospital, Surrey, UK

Agnieszka Slowik, Department of Neurology, Jagiellonian University, Krakow, Poland

Jennifer A. Smith, Department of Epidemiology, School of Public Health, University of Michigan Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI, USA

Nicholas L. Smith, Group Health Research Institute, Group Health Cooperative, Seattle WA 98101, USA. Department of Epidemiology, University of Washington, Seattle WA 98195, USA. Seattle Epidemiologic Research and Information Center, VA Office of Research and Development, Seattle WA 98108, USA.

Sylvia Wassertheil-Smoller, Department of Epidemiology and Population Health, Albert Einstein College of Medicine, New York, NY, USA

Martin Soderholm, Department of Clinical Sciences, Malmö, Lund University, and Section of Neurology, Skåne University Hospital, Malmö, Sweden

O. Colin Stine, Department of Epidemiology and Public Health, University of Maryland School of Medicine, Baltimore, MD, USA

Daniel Strbian, Department of Neurology, Helsinki University Hospital and University of Helsinki, Helsinki, Finland.

Cathie LM Sudlow, Centre for Medical Informatics, University of Edinburgh, Edinburgh, UK; BHF Data Science Centre, Health Data Research UK, London, UK

Turgut Tatlisumak, Department of Neurology, Helsinki University Hospital and University of Helsinki, Helsinki, Finland. Department of Neurology, Sahlgrenska University Hospital, Gothenburg, Sweden. Department of Clinical Neuroscience, Institute of Neurosciences and Physiology, Sahlgrenska Academy at University of Gothenburg, Gothenburg, Sweden

Chikashi Terao, Laboratory for Statistical and Translational Genetics, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan.

Vincent Thijs, Stroke Theme, Florey Institute of Neuroscience and Mental Health, University of Melbourne, Heidelberg, Victoria, Australia; Department of Neurology, Austin Health, Heidelberg, Victoria, Australia

Nuria P. Torres-Aguila, Stroke Pharmacogenomics and Genetics group, Biomedical Research Institute Sant Pau (IIB Sant Pau), Barcelona, Spain

David-Alexandre Tréguët, University of Bordeaux, Bordeaux Population Health Research Center, INSERM UMR1219, 146 rue Léo Saignat, 33076 Bordeaux, France

Anil Man Tuladhar, Department of Neurology, Donders Institute for Brain, Cognition and Behaviour, Radboud University Center, Nijmegen, The Netherlands

Jan H. Veldink, Department of Neurology, University Medical Center Utrecht Brain Center, Utrecht University, Utrecht, The Netherlands

Robin G. Walters, Nuffield Department of Population Health, University of Oxford, Oxford, UK. MRC Population Health Research Unit, University of Oxford, UK

David R. Weir, Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI, USA

Daniel Woo, Department of Neurology and Rehabilitation Medicine; University of Cincinnati College of Medicine; Cincinnati, OH, USA

Bradford B. Worrall, Departments of Neurology and Public Health Sciences, University of Virginia School of Medicine, Charlottesville, VA, USA

Charles C. Hong, Department of Medicine, University of Maryland School of Medicine, Baltimore, MD, USA

Owen Ross, Department of Neurology, Mayo Clinic, Jacksonville, FL, USA.

Ramin Zand, Geisinger Neuroscience Institute, Geisinger Health System, Danville, PA USA,

F-E de Leeuw, Department of Neurology, Radboud University Medical Center; Donders Medical Center for Neuroscience; Nijmegen, The Netherlands

Arne G. Lindgren, Department of Clinical Sciences Lund, Neurology, Lund University; Section of Neurology, Skåne University Hospital, Lund, Sweden.

Guillaume Pare, DBCVS Research Institute, Population Health Research Institute, McMaster University, Hamilton, Ontario, Canada; Thrombosis & Atherosclerosis Research Institute (TaARI)

Christopher D. Anderson, Department of Neurology, Brigham and Women's Hospital, Boston, MA, USA  
McCance Center for Brain Health, Massachusetts General Hospital, Boston, MA, USA  
Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA

Hugh S. Markus, Stroke Research Group, Department of Clinical Neurosciences, University of Cambridge, Cambridge, UK

Christina Jern, Institute of Biomedicine, Department of Laboratory Medicine, Sahlgrenska Academy, University of Gothenburg, and Region Vastra Gotaland, Sahlgrenska University Hospital, Department of Clinical Genetics and Genomics, Gothenburg, Sweden

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

Martin Dichgans, Institute for Stroke and Dementia Research (ISD), University Hospital, LMU Munich, Munich, Germany. Munich Cluster for Systems Neurology (SyNergy), Munich, Germany.  
German Center for Neurodegenerative Diseases (DZNE), Munich, Germany.

Braxton D. Mitchell, Division of Endocrinology, Diabetes and Nutrition, University of Maryland School of Medicine, Baltimore, MD, USA Geriatric Research and Education Clinical Center, Veterans Administration Medical Center, Baltimore, MD, USA

Steven J. Kittner, Department of Neurology and Geriatric Research and Education Clinical Center, VA Maryland Health Care System, Baltimore, MD  
Department of Neurology, University of Maryland School of Medicine, Baltimore, MD, USA



Search Terms: Cerebral infarction; Stroke in young adults; Association studies in genetics

Submission Type: Manuscript

Title Character Count: 24

Number of Tables: 2

Number of Figures: 3

Word Count of Abstract: 242

Word Count of Paper: 4,888

Corresponding Author:

Steven J. Kittner

655 W. Baltimore St.

Baltimore, MD 21201

Phone: 410-706-0414

skittner@som.umaryland.edu

## Abstract

**Objective:** To determine the contribution of common genetic variants to risk of early onset ischemic stroke (IS).

**Methods:** We performed a meta-analysis of genome-wide association studies of early onset IS, ages 18-59, using individual level data or summary statistics in 16,927 cases and 576,353 non-stroke controls from 48 different studies across North America, Europe, and Asia. We further compared effect sizes at our most genome-wide significant loci between early and late onset IS and compared polygenic risk scores for venous thromboembolism between early versus later onset IS.

**Results:** We observed an association between early onset IS and *ABO*, a known stroke locus. The effect size of the peak *ABO* SNP, rs8176685, was significantly larger in early compared to late onset IS (OR 1.17 (95% C.I.: 1.11-1.22) vs 1.05 (0.99-1.12); p for interaction = 0.008). Analysis of genetically determined *ABO* blood groups revealed that early onset IS cases were more likely to have blood group A and less likely to have blood group O compared to both non-stroke controls and to late onset IS cases. Using polygenic risk scores, we observed that greater genetic risk for venous thromboembolism, another prothrombotic condition, was more strongly associated with early, compared to late, onset IS (p=0.008).

**Conclusion:** The *ABO* locus, genetically predicted blood group A, and higher genetic propensity for venous thrombosis are more strongly associated with early onset IS, compared with late onset IS, supporting a stronger role of prothrombotic factors in early onset IS.

## Introduction

Substantial advances have been made in recent years towards identifying common genetic variation associated with risk of ischemic stroke (IS) <sup>1,2</sup>. This progress has been largely based on meta-analysis of genome-wide association study (GWAS) results derived from predominantly late onset cases. Given that a higher heritability of early onset IS observed in multiple studies<sup>3-6</sup>, there is a strong need for genetics studies focusing on early onset stroke. A pressing question is whether the genetic contribution to early onset stroke includes mechanisms that may be novel or specific to early onset ischemic stroke but may still have translational importance across the whole age spectrum as has been found from studies of early onset cases in other complex diseases <sup>7-9</sup>.

Because atherosclerosis is a less common cause of stroke in young adults, we hypothesized that non-atherosclerotic, prothrombotic mechanisms may be more important and discernable in studies of early onset IS <sup>10,11</sup>. This concept is supported by associations reported between early onset IS and multiple prothrombotic candidate genes <sup>10-14</sup>. In this report, we present findings from the Genetics of Early Onset Ischemic Stroke Consortium, contrast the effect sizes of known stroke loci in early versus later onset stroke, and evaluate differing contributions of prothrombotic loci to early and late onset IS.

## Methods

The Early Onset Stroke Consortium (EOSC) is a collaboration of investigators representing 48 different studies across North America, Europe, Japan, Pakistan, and Australia who have pooled their data for a GWAS meta-analysis of early onset IS in cases aged 18-59 years. Collectively, these studies contributed 17,077 cases (16,927 cases included for analysis) and 576,353 non-stroke controls (eMethods and eTable 1). All patients had brain imaging to exclude diagnoses other than IS. Additional screening was performed in some, but not all, studies to exclude cases believed to be due to a known monogenic cause (e.g., sickle cell disease) or to a known non-genetic cause (e.g., drug use, complications of procedures). Ischemic stroke subtyping was performed using the TOAST criteria <sup>15</sup> by most, but not all, sites.

The EOSC includes cases from two different sources: early onset stroke cases who previously participated in the Stroke Genetics Network (SiGN) <sup>16</sup> (n = 7,619), and early onset cases from additional non-SiGN study sites (n = 9,598). eTable 1 lists the 48 sites contributing early onset stroke cases and sources of controls. With one exception (Group 7 including cases from Barcelona and BASICMAR), controls from each study were of the same age or older than cases. Analysis groups were assigned as previously described to combine cases and controls of similar genetic ancestry groups and genotyped on arrays of similar density <sup>16</sup>. Clinical characteristics of the stroke cases are shown in eTable 2. Genotypes for all studies except Helsinki were imputed using the TOPMed reference panel on the University of Michigan Imputation Server <sup>17</sup>. The Helsinki Study imputed genotypes using a Finnish population-specific reference panel and the BEAGLE software. All genotype data were based on genome build hg38. Genotyping platforms and cohort-specific quality control and analysis parameters are provided in eTable 3.

GWAS of stroke cases and controls were conducted within sites or within groupings of sites, and then meta-analyzed. Prior to analysis, we removed stroke cases from analysis if there were fewer than 40 in the analysis strata, if they could not be assigned to a genetic ancestry group, or if there was not an adequate number of controls. Filtering on these criteria left 16,927 (of 17,077) early onset cases for analysis. The overall analytic design is depicted in Figure 1. Our primary analysis was a transethnic

meta-analysis. In parallel, we performed a European only meta-analysis. Logistic regression was performed to test for association between stroke occurrence and single variants. Covariates included sex and up to 10 principal components to adjust for population stratification. Power calculations indicated that our study provided 80% power to detect odds ratios ranging from 1.09 to 1.20 for common genetic variants with minor allele frequencies (MAF) > 5% at the genome-wide threshold for significance, i.e.,  $5 \times 10^{-8}$ . For comparison, the previously detected odds ratios for IS from MEGASTROKE GWAS (67,162 cases) ranged from 1.05 – 1.09<sup>1</sup>. We also performed TOAST-defined stroke subtype analyses for those sites providing subtype classification.

We compared effect sizes between early and late onset stroke cases of European ancestry at 40 loci previously associated with IS in MEGASTROKE,<sup>1</sup> MEGASTROKE and UK Biobank (UKB) combined,<sup>2</sup> or in a previously published meta-analysis of small vessel stroke<sup>18</sup>. A late onset stroke case cohort (age onset  $\geq 60$  years) from the SiGN Consortium consisting of 9,272 late onset cases and 25,124 controls of European ancestry was created for this comparison. Effect sizes between early and late onset stroke were compared using a Wald test.

To follow-up our peak association at the *ABO* locus, we assessed the association of early onset stroke with *ABO* blood groups and then compared these associations with late onset stroke. The *ABO* locus encodes two glycosyltransferases, A and B, that define the serologic blood groups, A, B, AB, and O. To explore associations of *ABO* blood group with stroke, we first tested whether histo-blood group *ABO* was associated with IS by comparing the distribution of these blood groups among early and late onset cases and controls. We assigned the diploid *ABO* blood group using genotypes at two SNPs (rs8176719 and rs8176746), as described by Groot et al.<sup>19</sup> (see eMethods).

At the haploid level, the *ABO* blood groups and subgroups are defined by 5 common haplotypes, each of which is tagged by a single SNP<sup>20</sup> (see eMethods). The 5 *ABO* haplotypes are A1 (the ancestral haplotype, tagged by rs2519093-T), O1 (tagged by a frameshift deletion, rs8176719-delG), A2 (tagged by rs1053878-A), B (tagged by rs8176743-T), and O2 (tagged by rs41302905-T)<sup>20</sup>. Goumidi and colleagues have recently shown that relative to haplotype O1, the genetically defined haplotypes A1 and B are strongly associated with venous thrombosis risk, and haplotype A2 is associated with a modest increase in risk<sup>20</sup>.

We hypothesized that the strong association of early onset stroke with the *ABO* locus was related to the prothrombotic properties of the *ABO* blood group. We evaluated this hypothesis first by testing whether the two lead *ABO* variants were associated with venous thromboembolism (VTE), another prothrombotic condition, in the UKB and if the association was more prominent in early compared to later onset VTE. Using summary level association results (VTE results from the INVENT Consortium<sup>21</sup>), we then estimated pairwise genetic correlations among early onset stroke, late onset stroke, and VTE using LD Score Regression analysis (LDSC)<sup>22</sup>. We then tested if genetic predisposition to VTE, as measured by a polygenic risk score (PRS), was more prominently associated with early compared to later onset stroke. For this purpose, we generated a VTE PRS for individuals from the EOSC and late onset stroke subset from SiGN based on a large prior GWAS of VTE<sup>23</sup> using PRSice software<sup>24</sup>. The VTE PRS included 255 SNPs using a GWAS p-value threshold of  $1 \times 10^{-5}$  (see eMethods). We tested the association between the VTE PRS score with stroke in the European ancestry sample using logistic regression with 10 principal components for ancestry and sex included as covariates. Effect sizes between early and late onset stroke were compared using a Wald test (See eMethods).

We identified several disorders and plasma biomarkers (late onset stroke, VTE, and plasma levels of von Willebrand factor (VWF) and Factor VIII) for which associations at the *ABO* locus have previously

been reported and then used the coloc software<sup>25</sup> to assess evidence that the same causal SNPs associated with EOS also drove associations with the second trait. Briefly, coloc utilizes a Bayesian approach and summary level association level results for two traits to calculate the posterior probabilities of five competing hypotheses (H0-H4) that assess whether the associations are due to the same (corresponding to H4) or a different (corresponding to H3) causal variant (see eMethods).

**Standard Protocol Approvals, Registrations, and Patient Consents.** All participating sites obtained IRB or Ethics Board approval, and informed consent was obtained from all participants or their legally authorized representative.

**Data Availability.** Summary results will be made available upon application to the contact authors and consortium approval of the request. Individual level data from a subset of sites will be made available on dbGaP.

## Results

**Genome-wide analysis results.** The trans-ethnic meta-analysis included a total of 16,927 early onset IS cases, representing 1,073 of African ancestry, 10,549 of European ancestry, 189 of Hispanic ancestry, 230 of Afro-Caribbean ancestry and 4,204 of pan-Asian ancestry. There was little evidence for inflation of p-values across sites, as indicated by genomic control ranging from 0.86 to 1.14 (eTable 3). We identified two loci associated with early onset IS at genome-wide significance. The most strongly associated variant at the first locus was rs8176685 with an odds ratio (OR) of 1.17 (95% confidence interval [95% CI]: 1.11-1.22;  $p = 8.25 \times 10^{-11}$ ), marking a 12 base pair insertion/deletion polymorphism at the *ABO* locus, a locus previously associated with IS<sup>1,26</sup> (Figure 1A and eTable 4), and in near perfect linkage disequilibrium with rs635634, the lead *ABO* SNP associated with all IS in MEGASTROKE<sup>1</sup>.

The second locus significantly associated with early onset IS was *MLXIP* (MLX Interacting Protein) (rs377424471, OR=0.90, 95% CI: 0.86-0.94;  $p=3.96 \times 10^{-8}$ ) (eTable 4). To our knowledge, this locus has not previously been associated with stroke. This SNP showed no evidence for association with late onset stroke in SIGN (OR = 1.00, 95% CI: 0.97-1.04,  $p = 0.99$ ), further highlighting the need for this result to be replicated.

The European ancestry only meta-analysis, based on 10,549 early onset stroke cases, also indicated a strong association with the *ABO* locus, albeit with a different lead SNP, rs529565 (previously rs912805253). *ABO* rs529565 T allele had an OR for association of 0.88 (95% CI: 0.85-0.91;  $p = 3.17 \times 10^{-13}$ ) (Figure 1B and eTable 5). Stratified analyses revealed no difference in effect sizes of either *ABO* rs8176685 or rs529565 between men and women.

In addition to the two genome-wide significant loci, we observed 23 “suggestive” loci with sub-threshold levels of significance (i.e.,  $p < 1 \times 10^{-6}$ ) in the trans-ethnic analysis (eTable 4) and 15 loci in the European only analysis (eTable 5) (9 of which were EUR only). Among these loci were *SH2B3* (SH2B Adaptor Protein 3, rs3184504), which has previously been associated with IS<sup>1</sup>, and SNPs at several compelling candidate loci, including *F11* (Plasma Thromboplastin Antecedent) encoding factor XI in the coagulation cascade, and *MC4R* (Melanocortin Receptor 4), an obesity-associated gene. Twelve of the 32 unique loci associated with early onset IS in either the Transethnic or European only analysis at sub-genome thresholds (i.e.,  $p < 1 \times 10^{-6}$ ) were nominally associated with late onset IS at  $p < 0.05$  (eTables 4 and 5), and the odds ratios for all 12 were approximately similar between early and late onset stroke.

Analysis of IS subtypes revealed 11 SNPs associated with stroke subtypes at genome-wide thresholds of significance (eTable 6), although the sample sizes were relatively small for each subtype (ranging from 886-5149 for transethnic analysis and 376-1502 for European only analysis) and the frequencies of the associated SNPs were low (8 with EUR MAF < 0.02 and the 3 remaining MAF < 0.09).

There was no evidence for replication of the *HABP2* rs11196288 variant, which was previously associated with all early onset IS in our earlier phase 1 transethnic meta-analysis from the EOSC<sup>12</sup>, although its minor allele frequency is only ~3% in European ancestry populations, as indicated in gnomAD). In the expanded meta-analysis presented in this report (16,927 currently vs 4,505 cases previously), there was no evidence for association of this SNP with early onset stroke in any of the new sites, including those of non-European ancestry (eFigure 1).

**Associations of index *ABO* variants with all IS and IS subtypes in early and late onset IS.** As described below, the rs8176685 deletion and the rs529565 T allele at the *ABO* locus tag blood groups A1 and O1, respectively. The associations we observed for these SNPs with early onset stroke are substantially higher than the peak associations previously reported at the *ABO* locus in predominantly older stroke populations (e.g., OR = 1.08; 95% CI: 1.05-1.11 in MEGASTROKE<sup>1</sup>). Both SNPs also had significantly larger effect sizes for all IS in early compared to late onset IS (*ABO* rs8176685: OR = 1.17 [95% CI: 1.11-1.22] vs 1.05 [95% CI: 0.99-1.12], and *ABO* rs529565: OR = 0.88 [95% CI: 0.85-0.92] vs 0.96 [95% CI: 0.91-1.00], p-values for interaction = 0.008 and 0.0006, respectively).

In subtype-specific analyses, the effect sizes of the A1-defining SNP rs8176685 did not differ significantly between early and late onset IS except for undetermined strokes, for which the A1-defining allele had a stronger effect in early compared to late onset stroke (p-value for homogeneity = 0.011). In contrast, the O1-defining allele at rs529565 was more protective in early versus late onset IS for large artery stroke, cardioembolic stroke, and undetermined stroke (p-values for homogeneity = 0.040, 0.040 and 0.002 respectively) (eFigures 2a and 2b).

We assessed the associations of *ABO* SNPs rs2519093 (in high LD with rs8176685, which was not present in UKB, and encoding blood group A1) and rs529565 (encoding blood group O1) with early and late onset IS in the UKB as a quasi-replication, 'quasi' because early onset UKB cases were included as part of the primary EOSC analyses, but late onset cases were not. The analysis was limited to IS cases and based on ICD codes, as described in the eMethods. Similar to our analysis in EOSC and SiGN, we observed stronger associations of both SNPs in early than in late onset stroke. The odds ratios of *ABO* rs2519093 (A1-defining) were 1.10 (95% CI: 1.01-1.19; p = 0.03) for early and 1.05 (95% CI: 1.00-1.02; p = 0.07) for late onset stroke, and the odds ratios of *ABO* rs529565 (O1-defining) were 0.93 (95% CI: 0.86-0.99; p = 0.02) for early and 0.95 (95% CI: 0.90-0.99; p = 0.02) for late onset stroke.

We performed a joint analysis of *ABO* SNPs rs8176685 and rs529565 to assess their independent associations with early onset IS. The frequencies of the rs8176685 deletion and the rs529565 T allele at the *ABO* locus in European ancestry populations are 0.20 and 0.64, respectively, leading to a modest correlation between them ( $r^2=0.44$ ), although the rs8176685 deletion appears only on the background of the more common rs529565 T allele, resulting in a  $D'$  of 1 (eFigure 3). Because of the high  $D'$  between these 2 SNPs, we used a stratified approach to assess the contribution of rs529565 to stroke risk in the absence of the rs8176685 deletion; that is, we estimated the association of rs529565 with stroke in individuals without the rs8176685 T allele. In this analysis, which was restricted to those strata for whom we had individual level data, rs529565 remained associated with all IS and at approximately the same effect size (OR = 0.90, 95% CI: 0.85-0.96; p=0.003), implying an association of this SNP that was independent of the rs8176685 indel SNP.

**Associations of ABO blood group-defining SNPs with early and late onset stroke.** We tested associations of these blood group-defining haplotypes with early and late onset stroke in the European ancestry subgroup. As indicated previously, the top ABO SNPs identified in our GWAS, rs8176685 and rs529565, are in high linkage disequilibrium with the tagging SNPs for haplotypes A1 and O1, which were not included in our dataset. We also used rs1137827 as a high LD tag for rs8176743 (blood group B) (Table 1). Consistent with the VTE findings from Goumidi et al, our analyses indicate haplotype A1 (rs8176685) to be strongly associated with early onset stroke (OR = 1.18, 95% CI: 1.11-1.25;  $p = 1.95 \times 10^{-8}$  and blood group O1 (rs529565) to be protective (OR = 0.88, 95% CI: 0.85-0.91;  $p = 3.17 \times 10^{-13}$ ) (Table 1). Unlike for VTE, blood group B (rs1137827) showed little evidence for association with stroke (OR = 1.04, 95% CI: 0.94-1.14;  $p = 0.463$ ). These trends persisted for late onset stroke, although the strength of associations was markedly reduced; that is, blood group A1 was only modestly associated with late onset stroke (OR = 1.05, 95% CI: 1.00-1.10;  $p = 0.044$ ) and blood group O1 was only modestly associated with protection against late onset stroke (OR = 0.96, 95% CI: 0.92-1.00;  $p = 0.036$ ).

To evaluate if rs8176685 and rs529565 accounted for all of the genetic effects at the ABO locus, we performed a conditional analysis in Europeans only of the ABO locus to test for association of all SNPs at the ABO locus ( $\pm 50$  kb from the gene) with early onset IS after including rs8176685 and rs529565 in the model as covariates. These analyses revealed 7 SNPs, falling within 4 LD groups, to be associated with all IS at a  $p$ -value  $< 0.01$ . Of these, rs5598407 (MAF = 0.045) was the most strongly associated with IS (OR = 1.21,  $p = 3.13 \times 10^{-4}$ ). This SNP was in LD with all of the blood group-defining SNPs ( $r^2 < 0.03$  but  $D' = 1$  for all), and with rs176694 ( $r^2=0.347$ ), which is strongly associated with E-selectin levels ( $p < 10^{-406}$ ) in the GWAS catalogue.<sup>27</sup> None of the tag SNPs for the 3 other blood group haplotypes (rs1053878-A2, rs1137827-B, and rs41302905-O2) showed evidence for association after conditioning on rs8176685 and rs529565 ( $p > 0.40$  for all), which tag blood groups A1 and O1, respectively.

**Association of haplotype-defined ABO serologic blood group with early and late onset stroke.** We assessed the relation of serologic blood groups at the ABO locus (i.e., A, B, AB, O blood groups) to stroke by comparing the distribution of blood types between early onset IS cases, late onset IS cases, and non-stroke controls, defining the blood groups A, B, and O genetically using a combination of SNPs<sup>19</sup>. These comparisons revealed that the distribution of blood groups differed significantly between early and late onset stroke cases ( $p = 0.0001$ ), between early onset stroke cases and controls ( $p < 0.00001$ ), and between late onset stroke cases and controls ( $p = 0.0009$ ) (Table 2). Pairwise comparisons then revealed the frequency of blood group A to be more common in early onset compared to late onset stroke (48.4% vs 45.2%;  $p = 0.0001$ ) and least common in non-stroke controls (44.4%;  $p < 0.00001$  for early onset stroke vs controls and  $p = 0.20$  for late onset stroke vs controls). Conversely, blood group O was less common in early compared to late onset stroke cases (35.5% vs 39.1%,  $p < 0.0001$ ), and most frequent in non-stroke controls (41.1%,  $p < 0.0001$  compared to both early and late onset stroke cases). The frequency of blood group B was slightly higher in both early (11.3%) and late (11.0%) onset stroke cases compared to non-stroke controls (10.1%) ( $p = 0.004$  and  $0.012$  respectively). We considered the possibility that this association could be due to a survival bias (i.e., a preferential survival of those with blood group O by assessing the association of age with blood group). Among non-stroke controls, age group (age  $< 60$  vs  $\geq 60$  years) was not associated with blood group, implying that the association of early onset stroke with blood group was not driven by differential survival between subjects with different blood groups.

**Associations of other established stroke loci with early onset stroke.** We compared the effect sizes of 40 loci previously found to associate with IS<sup>1, 2, 16</sup> between early (age at first stroke  $< 60$ ) and

late ( $\geq$  age 60) onset stroke. As indicated in eTable 7, the odds ratios associated with EOS were generally consistent with those estimated for later onset stroke with two exceptions, *RGS7* rs146390073 and *TM4SF4* rs7610618, although the minor allele frequencies were relatively rare for both SNPs and there were no statistically significant differences between early and late onset IS.

**Genetically defined ABO and risk of early and late onset VTE.** Because the *ABO* locus has been previously associated with VTE and other prothrombotic states<sup>28</sup>, we assessed whether there was a similar graded age-at-onset association between *ABO* SNPs rs2519093 (encoding blood group A1) and rs529565 (encoding blood group O1) and VTE in the UKB. The *ABO* rs2519093-A1 allele was more strongly associated with early onset VTE (age < 60 years; n = 3,514 cases) (OR = 1.64, 95% CI: 1.54-1.74;  $p = 1.42 \times 10^{-54}$ ) than with late onset VTE ( $\geq$  age 60 years; n = 5,043 cases) (OR = 1.34, 95% CI: 1.27-1.41;  $p = 2.89 \times 10^{-29}$ ); p-value for homogeneity of ORs =  $4.01 \times 10^{-7}$ . Similarly, the *ABO* rs529565-O1 allele was more strongly associated with early onset VTE (OR = 0.66, 95% C.I.: 0.62-0.69;  $p = 2.95 \times 10^{-58}$ ) than with late onset VTE (OR = 0.77, 95% C.I.: 0.74-0.81;  $p = 6.63 \times 10^{-28}$ ); p-value for homogeneity of OR =  $2.15 \times 10^{-6}$ ) (details in eMethods).

**Genetic risk of VTE on risk of early and late onset stroke.** We also used 2 approaches to evaluate whether genetic risk of VTE confers risk for stroke, and especially for earlier onset stroke. First, according to LDSC analysis on our current GWAS and the GWAS of VTE from INVENT, the genetic correlation between VTE and early onset stroke was  $0.376 \pm 0.153$ ,  $p = 0.014$  and between VTE and late onset stroke was  $0.032 \pm 0.198$ ,  $p = 0.871$ . The genetic correlation between early and late onset stroke was  $0.455 \pm 0.128$ ,  $p = 0.0004$ ).

Second, genetic risk of VTE was measured as a 255-SNP-based polygenic risk score (PRS),<sup>23</sup> as described in the eMethods. A 1-SD unit increase in VTE PRS was associated with a 1.13-fold increase in risk of early onset stroke (95% C.I.: 1.10 -1.16,  $p < 5.21 \times 10^{-16}$ ) and a 1.04-fold increase in risk of late onset stroke (95% C.I.: 1.01 -1.08,  $p = 0.010$ ); p-value for homogeneity of OR = 0.0002). These results were essentially unchanged when the analysis was repeated after removing 7 SNPs at the *ABO* locus from the PRS.

**Co-location of associations at the ABO locus between EOS and other thrombotic-related disorders and biomarkers.** Figure 3 provides a visualization of the SNPs most strongly associated with early onset stroke and their corresponding associations with VTE and two prothrombotic biomarkers. Each panel shows the zoom plot at the *ABO* locus depicting the association of SNPs at the *ABO* locus ( $\pm 50$  kb) with early onset IS. The heat map in each panel and the color coding of the variants depict the association with late onset stroke (Panel 3A), VTE (Panel 3B), plasma levels of von Willebrand factor (VWF, Panel 3C), and plasma levels of Factor VIII (F8, Panel 3D). The SNPs most strongly associated with EOS tended also to be the ones most strongly associated with VTE, VWF, and F8. This trend was less apparent for late onset stroke. Consistent with these observations, there was strong evidence using formal colocalization analyses to support the hypothesis of colocalization of at least 1 shared causal SNP between EOS and VTE, EOS and VWF, and EOS and FVIII (posterior probability supporting H4 > 99% for all pairs). Consistent with the weak and disperse set of associations with late onset stroke at this locus, there was insufficient evidence to strongly support either colocalization or absence of colocalization of shared causal SNPs between EOS and late onset stroke (posterior probability supporting H4 = 39%; posterior probability supporting H3 = 61%).



## Discussion

Our analyses revealed two variants at the *ABO* locus that were highly associated with early onset IS. These variants tag two of the *ABO* blood groups, blood group A1 and O1, showing a strong deleterious and protective association with IS, respectively. Non-O blood groups have been associated previously with risk of IS,<sup>29-31</sup> but the novel contributions of our analysis are in showing a significantly stronger association of these blood groups with early compared to late onset stroke and in linking risk predominantly to blood group haplotype A1. In particular, our analyses suggest that the haplo-blood group A1 and O1-tagging variants rs8176685 and rs529565 are sufficient for capturing nearly all of the *ABO*-mediated genetic association with early (and perhaps late) onset stroke. Stratified analyses indicate that both SNPs are independently associated with stroke and further association analyses at the *ABO* locus that condition on the effects of these two SNPs reveal only modest additional signal at this locus.

It should be highlighted that the stronger association of *ABO* in early compared to late onset stroke is seen on the relative scale. That is, the associations observed in our study were estimated using logistic regression in case-control study designs. Since the absolute risk of stroke is smaller at younger ages, it is unclear if the larger differential effect is seen also on the additive scale. Future prospective studies that can estimate risk directly are needed to investigate the absolute risks of IS incurred by *ABO*.

The association of *ABO* in early onset stroke suggests that the associations of *ABO* with IS may be mediated via thrombotic mechanisms and not atherosclerosis. Supporting this speculation, prior studies have shown *ABO* rs687289, which is in high LD with rs529565 and tags blood group O1, to be strongly associated with plasma levels of coagulation factor VIII and its carrier protein, von Willebrand factor<sup>32</sup>. These factors regulate hemostasis and thrombosis, and higher levels are associated with risk of arterial and venous thrombosis<sup>33</sup>. We have further shown in our analyses of UKB data that *ABO* rs529565 is more strongly associated with early compared to late onset VTE. Genetic correlations and PRS further support that genetic risk of VTE, a well-recognized prothrombotic-related disorder, is also more strongly associated with early compared to late onset stroke.

Although our study had limited power to examine IS subtypes, it is notable that the peak *ABO* SNP in the trans-ancestry meta-analysis, rs8176685, was also significantly associated with the subtypes of large artery atherosclerosis, cardioembolism, and undetermined. This SNP was not associated with the small artery occlusion or the other determined subtypes, although the latter subtype had a very small samples size. Perhaps because of the weaker association, in MEGASTROKE, the *ABO* locus was associated only with all IS but not with any subtype.

Differential associations of the *ABO* locus with myocardial infarction and coronary artery calcification further support a prothrombotic mechanism. The *ABO* locus has been associated with susceptibility to myocardial infarction (ICD-10 code 411.2) in CARDIoGRAM<sup>34</sup> and in the UKB ( $p = 2.1E-5$ ; from <https://pheweb.org/UKB-TOPMed/variant/9:133274084-C-T>, Accessed 4/5/2021). However, there is no evidence for a strong association of this locus with coronary artery calcification, a marker for subclinical atherosclerosis<sup>35,36</sup>, implying that this association may be mediated by a mechanism other than atherosclerosis (e.g., hemostasis). Taken together, our analyses with these additional observations support an increased genetic role of prothrombotic mechanisms in early compared to later onset stroke. This speculation is further supported by our colocalization analyses supporting the presence of a shared causal variant at the *ABO* locus between EOS and VTE, EOS and VWF, and EOS and FVIII. Evidence supporting a shared causal variant at the *ABO* locus between EOS and late onset stroke was

equivocal, perhaps because of the relatively low strength of association between late onset stroke and the *ABO* locus and consequently low power of the colocalization test.

This leads to the question, what are the clinical implications of an enrichment of prothrombotic mechanisms in early onset stroke? Clinical translation will require a better understanding of the prothrombotic mechanisms in early-onset stroke and, likely, a personalized secondary prevention strategy. The effect sizes of the stroke-associated common variants at the *ABO* locus are too small per se to have immediate clinical implications, but gene-gene and gene-environment interaction deserve future study<sup>37</sup>. One path to translation would be to identify gene-drug interactions (e.g. oral contraceptives and genetic risk for thrombosis) and determine whether the joint effect has implications for primary prevention. Additional research implications are that rare variant studies should target prothrombotic and related pathways, which could identify variants of larger effect size.

In addition to *ABO*, we detected genome-wide evidence for association of early onset stroke with *MLXIP* rs377424471. This locus has not previously been associated with stroke to our knowledge, nor did rs377424471 show any evidence for association with late onset stroke in SiGN. *MLXIP* does, however, play a role in glucose-responsive gene regulation, making it a plausible candidate gene for stroke.<sup>38</sup> The association of this SNP with early onset stroke requires future replication in an independent early onset stroke cohort.

Our study is not without limitations. Although 35% of subjects in the EOSC are of non-European ancestry, the diversity of the current EOSC cohort is still somewhat limited, reducing power to detect variants whose frequencies might be high in non-European populations yet low in Europeans. A second major limitation is that the sample size for all IS is still small by GWAS standards and power to detect subtype-specific variants is even more limited.

In summary, our genome-wide analysis indicates a stronger association of *ABO* risk variants tagging blood groups A1 and O1 with earlier compared to later onset stroke, and stronger associations of the same *ABO* variants with early compared to late onset VTE, another prothrombotic condition. Similarly, we observed genetic risk for VTE to be more strongly correlated with early compared to late onset stroke. Our findings are consistent with an increased role for prothrombotic mechanisms in early compared to later onset stroke.

## Study funding

This study was supported by the National Institutes of Health grants R01 NS105150, R01 NS100178 and R01NS114045 and Department of Veterans Affairs grant BX004672-01A1. The project utilized the resources and facilities at the VA Maryland Health Care System, Baltimore, Maryland.

The contents do not represent the views of the U.S. Department of Veterans Affairs or the United States Government.

Tom Jaworek was supported by a T32 AG000262 Epidemiology of Aging grant. Huichun Xu was supported by AHA Career Development Award (19CDA34760258).

Please see eAppendix 1 for the funding and acknowledgements for each contributing study.

## Disclosures

A. Butterworth reports grants outside of this work from AstraZeneca, Biogen, BioMarin, Bioverativ, Merck, Novartis, Pfizer and Sanofi and personal fees from Novartis.

J. Cárcel-Márquez is supported by an AGAUR Contract (agència de gestió d'ajuts universitaris i de recerca; FI\_DGR 2019, grant number 2019\_FI\_B 00853) co-financed with Fons Social Europeu (FSE).

J. Danesh reports grants, personal fees and non-financial support from Merck Sharp & Dohme (MSD), grants, personal fees and non-financial support from Novartis, grants from Pfizer and grants from AstraZeneca outside the submitted work. J. Danesh sits on the International Cardiovascular and Metabolic Advisory Board for Novartis (since 2010); the Steering Committee of UK Biobank (since 2011); the MRC International Advisory Group (ING) member, London (since 2013); the MRC High Throughput Science 'Omics Panel Member, London (since 2013); the Scientific Advisory Committee for Sanofi (since 2013); the International Cardiovascular and Metabolism Research and Development Portfolio Committee for Novartis; and the Astra Zeneca Genomics Advisory Board (2018).

A.M. Tuladhar is a junior staff member of the Dutch Heart Foundation (grant number 2016T044).

F.E. de Leeuw received a grant from the Dutch Heart Foundation (grant 2014 T060 to FE.d.L) and Bike4Brains. AMT received a grant from the Dutch Heart Foundation (grant 2016 T044 to A.M.T) and from the Netherlands CardioVascular Research Initiative: the Dutch Heart Foundation (CVON 2018-28 & 2012-06 Heart Brain Connection to A.M.T).

A.G. Lindgren reports personal fees from Bayer, Astra Zeneca, BMS Pfizer, and Portola.

C.D. Anderson receives sponsored research support from the National Institutes of Health of the United States (R01NS103924, U01NS069763), the American Heart Association, the Bugher Foundation, Massachusetts General Hospital, and Bayer AG, and has consulted for ApoPharma and Invitae.

All other authors report no disclosures.

## Figure Legends

Figure 1: Overview of the study design.

Figure 2. Manhattan plot of early onset IS meta-analysis in the EOSC, random effects model, in TRANS ancestry (panel A) and European ancestry (panel B).

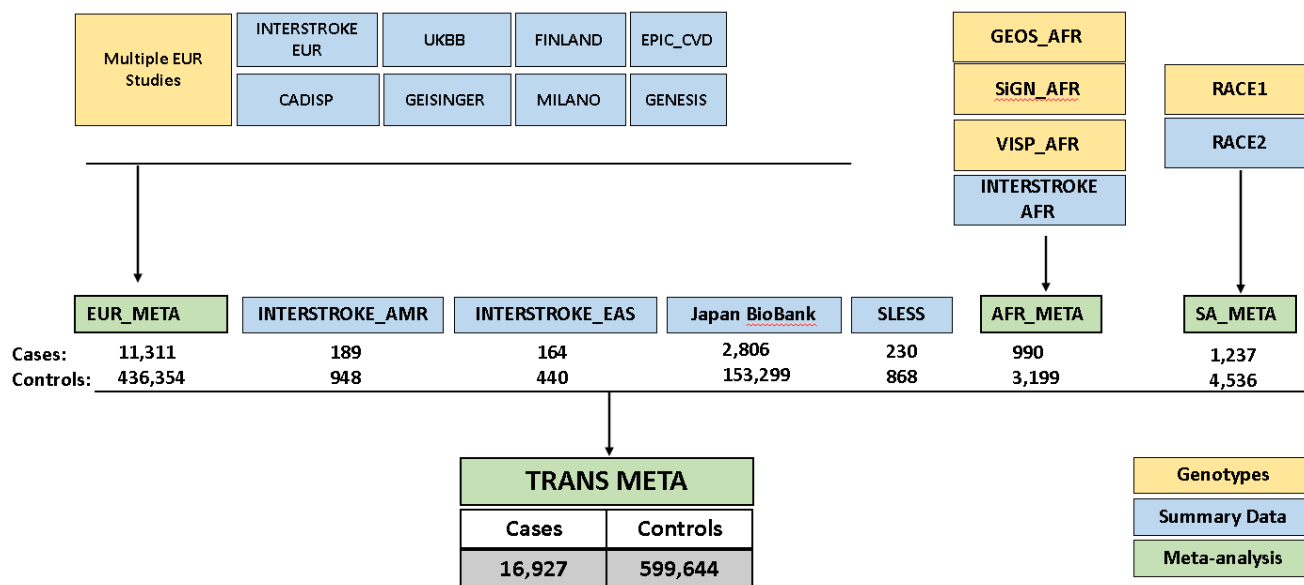
Figure 3: Association plots showing SNPs associated with early onset stroke at the *ABO* locus ( $\pm 50$  kb) and the corresponding associations of these SNPs with late onset stroke (Panel 3A), VTE (Panel 3B), plasma levels of von Willebrand factor (Panel 3C), and plasma levels of Factor VIII (Panel 3D).

## References

1. Malik R, Chauhan G, Traylor M, et al. Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nat Genet* 2018.
2. Malik R, Rannikmae K, Traylor M, et al. Genome-wide meta-analysis identifies 3 novel loci associated with stroke. *Ann Neurol* 2018;84:934-939.
3. MacClellan LR, Mitchell BD, Cole JW, et al. Familial aggregation of ischemic stroke in young women: the Stroke Prevention in Young Women Study. *Genet Epidemiol* 2006;30:602-608.
4. Seshadri S, Beiser A, Pikula A, et al. Parental occurrence of stroke and risk of stroke in their children: the Framingham study. *Circulation* 2010;121:1304-1312.
5. Bluher A, Devan WJ, Holliday EG, et al. Heritability of young- and old-onset ischaemic stroke. *Eur J Neurol* 2015;22:1488-1491.
6. Traylor M, Makela KM, Kilarski LL, et al. A novel MMP12 locus is associated with large artery atherosclerotic stroke using a genome-wide age-at-onset informed approach. *PLoS Genet* 2014;10:e1004469.
7. Fajans SS, Bell GI, Polonsky KS. Molecular mechanisms and clinical pathophysiology of maturity-onset diabetes of the young. *N Engl J Med* 2001;345:971-980.
8. Couch FJ, Nathanson KL, Offit K. Two decades after BRCA: setting paradigms in personalized cancer care and prevention. *Science* 2014;343:1466-1470.
9. Ference BA, Robinson JG, Brook RD, et al. Variation in PCSK9 and HMGCR and risk of cardiovascular disease and diabetes. *N Engl J Med* 2016;375:2144-2153.
10. Hanson E, Nilsson S, Jood K, et al. Genetic variants of coagulation factor XI show association with ischemic stroke up to 70 years of age. *PLoS ONE* 2013;8:e75286.
11. Cole JW, Xu H, Ryan K, et al. Genetics of the thrombomodulin-endothelial cell protein C receptor system and the risk of early-onset ischemic stroke. *PLoS ONE* 2018;13:e0206554.
12. Cheng YC, Stanne TM, Giese AK, et al. Genome-wide association analysis of young-onset stroke identifies a locus on chromosome 10q25 near HBP2. *Stroke* 2016;47:307-316.
13. Hamedani AG, Cole JW, Cheng Y, et al. Factor V Leiden and ischemic stroke risk: The Genetics of Early Onset Stroke (GEOS) Study. *J Stroke Cerebrovasc Dis* 2013;22:419-423.
14. Jiang B, Ryan KA, Hamedani A, et al. Prothrombin G20210A mutation is associated with young-onset stroke: The Genetics of Early-Onset Stroke Study and meta-analysis. *Stroke* 2014;45:961-967.
15. Adams HP, Jr., Bendixen BH, Kappelle LJ, et al. Classification of subtype of acute ischemic stroke. Definitions for use in a multicenter clinical trial. TOAST. Trial of Org 10172 in Acute Stroke Treatment. *Stroke* 1993;24:35-41.
16. NINDS Stroke Genetics Network, International Stroke Genetics Consortium. Loci associated with ischaemic stroke and its subtypes (SiGN): a genome-wide association study. *Lancet Neurol* 2016;15:174-184.
17. Traylor M, Persyn E, Tomppo L, et al. Genetic basis of lacunar stroke: a pooled analysis of individual patient data and genome-wide association studies. *Lancet Neurol* 2021;20:351-361.
18. Groot HE, Villegas Sierra LE, Said MA, Lipsic E, Karper JC, van der Harst P. Genetically determined ABO blood group and its associations with health and disease. *Arterioscler Thromb Vasc Biol* 2020;40:830-838.
19. Goumidi L, Thibord F, Wiggins KL, et al. Association between ABO haplotypes and the risk of venous thrombosis: impact on disease risk estimation. *Blood* 2021;137:2394-2402.
20. Germain M, Chasman DI, de Haan H, et al. Meta-analysis of 65,734 individuals identifies TSPAN15 and SLC44A2 as two susceptibility loci for venous thromboembolism. *Am J Hum Genet* 2015;96:532-542.
21. Bulik-Sullivan B, Finucane HK, Anttila V, et al. An atlas of genetic correlations across human diseases and traits. *Nat Genet* 2015;47:1236-1241.

22. Klarin D, Busenkell E, Judy R, et al. Genome-wide association analysis of venous thromboembolism identifies new risk loci and genetic overlap with arterial vascular disease. *Nat Genet* 2019;51:1574-1579.
23. Choi SW, O'Reilly PF. PRSice-2: Polygenic Risk Score software for biobank-scale data. *GigaScience* 2019;8.
24. Giambartolomei C, Vukcevic D, Schadt EE, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet* 2014;10:e1004383.
25. Williams FM, Carter AM, Hysi PG, et al. Ischemic stroke is associated with the ABO locus: The EuroCLOT study. *Ann Neurol* 2013;73:16-31.
26. Machiela MJ, Chanock SJ. LDassoc: an online tool for interactively exploring genome-wide association study results and prioritizing variants for functional investigation. *Bioinformatics* 2018;34:887-889.
27. Morange PE, Tregouet DA. Current knowledge on the genetics of incident venous thrombosis. *J Thromb Haemost* 2013;11 Suppl 1:111-121.
28. Sabino Ade P, Ribeiro DD, Dominghetti CP, et al. ABO blood group polymorphisms and risk for ischemic stroke and peripheral arterial disease. *Mol Biol Rep* 2014;41:1771-1777.
29. Vasan SK, Rostgaard K, Majeed A, et al. ABO blood group and risk of thromboembolic and arterial disease: A study of 1.5 million blood donors. *Circulation* 2016;133:1449-1457; discussion 1457.
30. Zakai NA, Judd SE, Alexander K, et al. ABO blood type and stroke risk: the REasons for Geographic And Racial Differences in Stroke Study. *J Thromb Haemost* 2014;12:564-570.
31. Sabater-Lleal M, Huffman JE, de Vries PS, et al. Genome-Wide Association Transethnic Meta-Analyses Identifies Novel Associations Regulating Coagulation Factor VIII and von Willebrand Factor Plasma Levels. *Circulation* 2019;139:620-635.
32. Spiel AO, Gilbert JC, Jilma B. von Willebrand factor in cardiovascular disease: focus on acute coronary syndromes. *Circulation* 2008;117:1449-1459.
33. Schunkert H, König IR, Kathiresan S, et al. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat Genet* 2011;43:333-338.
34. O'Donnell CJ, Kavousi M, Smith AV, et al. Genome-wide association study for coronary artery calcification with follow-up in myocardial infarction. *Circulation* 2011;124:2855-2864.
35. van Setten J, Isgum I, Smolonska J, et al. Genome-wide association study of coronary and aortic calcification implicates risk loci for coronary artery disease and myocardial infarction. *Atherosclerosis* 2013;228:400-405.
36. Ohira T, Cushman M, Tsai MY, et al. ABO blood group, other risk factors and incidence of venous thromboembolism: the Longitudinal Investigation of Thromboembolism Etiology (LITE). *J Thromb Haemost* 2007;5:1455-1461.
37. Yamashita H, Takenoshita M, Sakurai M, et al. A glucose-responsive transcription factor that regulates carbohydrate metabolism in the liver. *Proc Natl Acad Sci U S A* 2001;98:9116-9121.

Figure 1. Overview of Study Design



\*Case/Control number equals the number included in the final analysis.

\*\*See eTable 1 for list of the studies with genotype data used in the European meta-analysis

Figure 2. Manhattan plot of early onset stroke meta-analysis in the EOSC, random effects model, in TRANS ancestry (panel A) and European ancestry (panel B) showing association with *ABO*. The orange line marks the GWAS significant threshold of  $5 \times 10^{-8}$  and the black line marks the GWAS suggestive threshold of  $1 \times 10^{-6}$ .

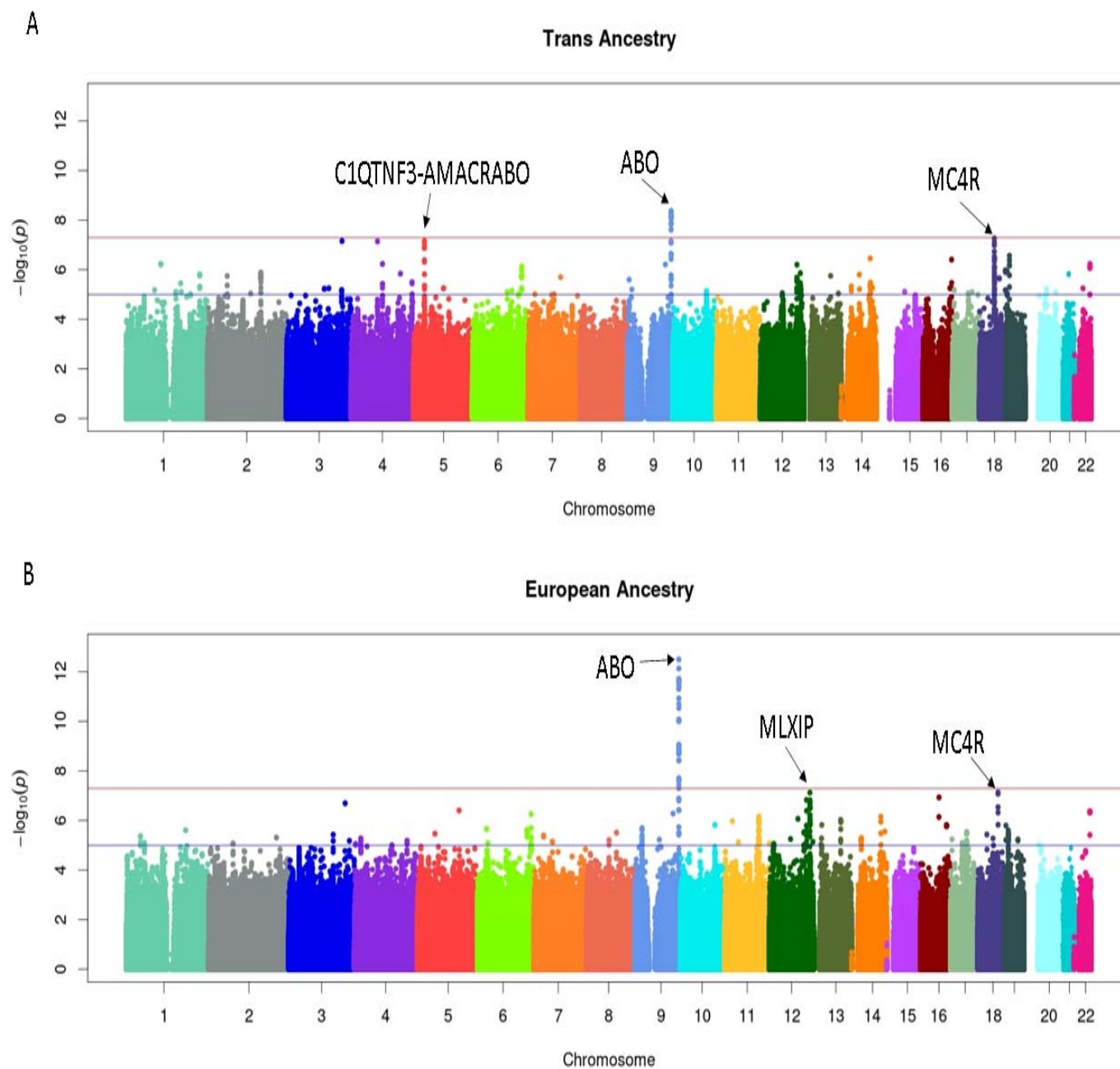




Figure 3: Association plots showing SNPs associated with early onset stroke at the *ABO* locus ( $\pm 50$  kb) and the corresponding associations of these SNPs with late onset stroke (Panel 3A), VTE (Panel 3B), plasma levels of von Willebrand factor (Panel 3C), and plasma levels of Factor VIII (Panel 3D).

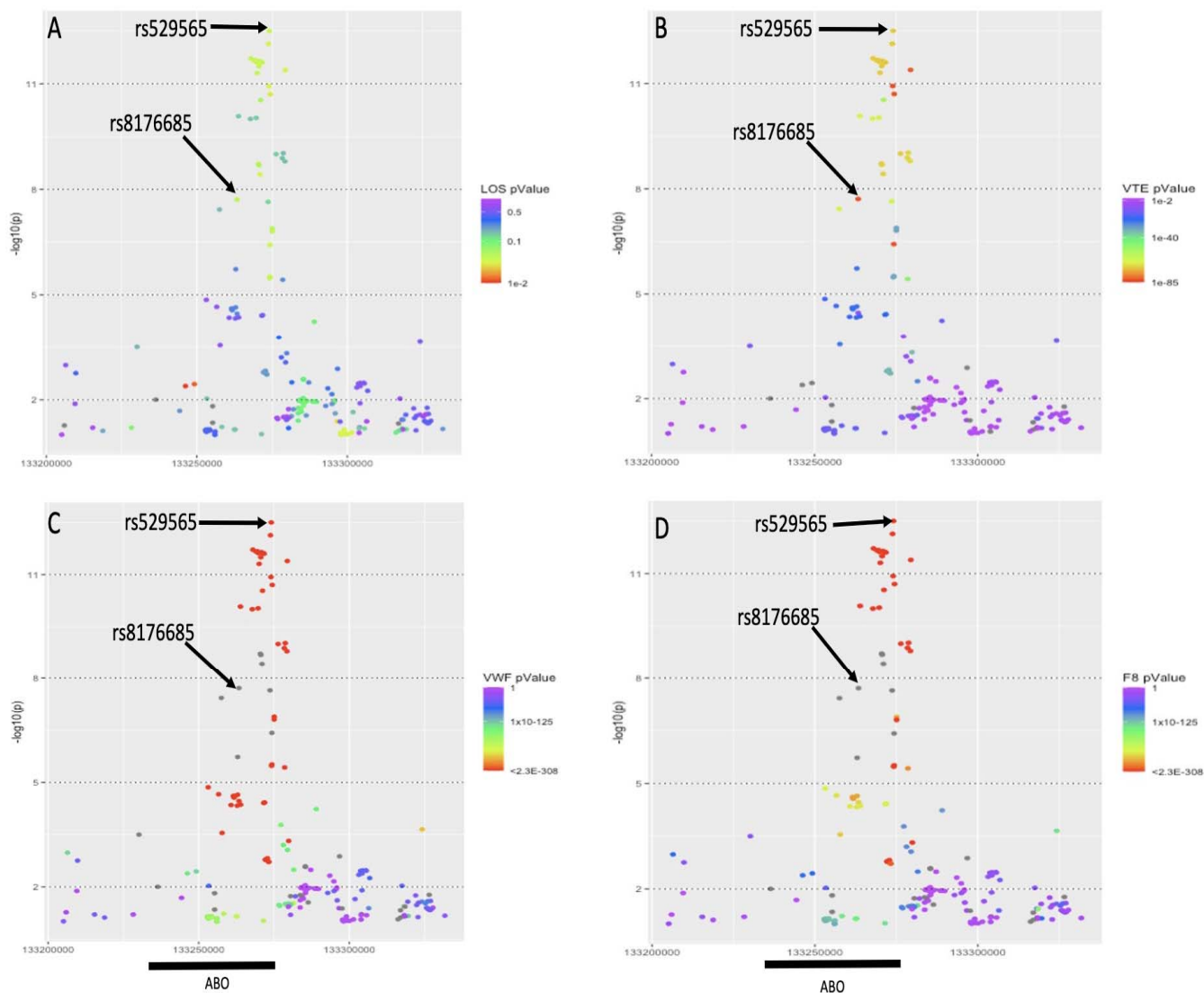


Table 1. Association of the ABO blood group-defining SNPs with early and late onset stroke, EUR only

SNP		Blood group	Alleles	Blood group-determining allele	Freq of blood group-determining allele***	Early onset stroke		Late onset stroke		P-value for interaction (early vs late)
						OR	P-value	OR	P-value	
rs8176685	high LD proxy for rs2519093 ( $r^2 = 0.97$ , $D' = 1.00$ ****)	A1**	CGCCCACCACT ACGCC/CGCC	deletion	0.20	1.18	1.95E-08	1.05	0.044	0.002
rs529565	high LD proxy for rs8176719 ( $r^2 = 0.94$ , $D' = 0.98$ ****)	O1	T/C	T	0.64	0.88	3.17E-13	0.96	0.036	6.37E-04
rs1053878		A2*	G/A	A	0.08	1.00	0.963	1.00	0.977	0.96
rs1137827	high LD proxy for rs8176743 ( $r^2 = 1.00$ , $D' = 1.00$ ****)	B**	C/T	T	0.09	1.04	0.463	1.02	0.591	0.76
rs41302905		O2	C/T	T	0.02	0.95	0.377	0.90	0.111	0.58

\* blood group-associated allele associated with moderate VTE risk; \*\* blood group-associated allele associated with high VTE risk; \*\*\*Frequency values obtained from gnomAD-Genomes European samples; \*\*\*\*LD values based on 1000 Genome CEU population.

Table 2. Distribution of ABO blood types in early onset stroke cases, late onset stroke cases, and in non-stroke controls\*

Blood group	Early onset stroke cases		Late onset stroke cases		Controls		p-values for omnibus test for difference in distribution of blood groups (3 df)			p-values for pairwise test for difference in distribution of blood groups (1 df)		
	n	proportion	n	proportion	n	proportion	early vs late	early vs controls	late vs controls	early vs late	early vs controls	late vs controls
A	2817	0.484	4187	0.452	13023	0.444	0.0001	< 0.00001	0.0009	0.0001	< 0.00001	0.20
AB	282	0.048	439	0.047	1270	0.043				0.77	0.10	0.08
B	661	0.113	1021	0.110	2964	0.101				0.53	0.004	0.012
O	2065	0.355	3622	0.391	12063	0.411				< 0.00001	< 0.00001	0.0004
TOTAL	5825		9269		29320							

\* European ancestry subjects only