# The Subtype Specificity of Genetic Loci Associated with Stroke in 16,664 cases and 32,792 controls

Running Head: Subtype Specificity of Stroke Genetic Loci

Matthew Traylor PhD[1,2], Christopher D. Anderson MD[3,4,5,6], Loes C.A. Rutten-Jacobs PhD[7], Guido J. Falcone MD[8], Mary E. Comeau MA[9], Hakan Ay MD[10,11], Cathie L.M. Sudlow FRCP(E)[12,13], Huichun Xu[14], Braxton D. Mitchell PhD[14,15], John Cole MD[14,16], Kathryn Rexrode MD[17], Jordi Jimenez-Conde MD[18,19], Reinhold Schmidt MD[20], Raji P. Grewal MD[21], Ralph Sacco MD[22], Marta Ribases MD[23,24,25], Tatjana Rundek MD[22], Jonathan Rosand MD[3,4,5,6], Martin Dichgans MD[26,27], Jin-Moo Lee MD[28], Carl D. Langefeld PhD[9], Steven J. Kittner MD[15,29], Hugh S. Markus DM[1], Daniel Woo MD[30], Rainer Malik PhD[26], on behalf of the NINDS Stroke Genetics Network (SiGN) and International Stroke Genetics Consortium (ISGC)

[1] Stroke Research Group, Department of Clinical Neurosciences, University of Cambridge, UK

[2] William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK.

[3] Center for Genomic Medicine, Massachusetts General Hospital (MGH), Boston, MA, USA.

[4] J. Philip Kistler Stroke Research Center, Department of Neurology, MGH, Boston, MA, USA.

[5] Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, USA.

[6] Division of Neurocritical Care and Emergency Neurology, Department of Neurology, MGH, Boston, MA, USA

[7] German Center for Neurodegenerative Diseases, Population Health Sciences, Bonn, Germany

[8] Division of Neurocritical Care and Emergency Neurology, Department of Neurology, Yale School of Medicine, New Haven, CT, USA

[9] Department of Biostatistical Sciences, Division of Public Health Sciences, School of Medicine, Wake Forest University, Winston-Salem, North Carolina, USA.

[10] Stroke Service, Massachusetts General Hospital, Boston, MA, USA.

[11] A.A. Martinos Center for Biomedical Imaging, Department of Radiology, Massachusetts General Hospital, Boston, MA, USA.

[12] Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh, UK

[13] Usher Institute of Population Health Sciences and Informatics, Nine Bioquarter, Edinburgh, UK

[14] Division of Endocrinology, Diabetes and Nutrition, Department of Medicine, University of Maryland School of Medicine, Baltimore, MD, USA

[15] Geriatrics Research and Education Clinical Center, Baltimore Veterans Administration Medical Center, Baltimore, MD, USA.

[16] Department of Neurology, Veterans Affairs Medical Center, Baltimore, MD, USA.

[17] Channing Division of Network Medicine and Division of Women's Health, Department of Medicine, Brigham and Women's Hospital, Boston, MA

[18] Neurovascular Research Unit, Department of Neurology, Institut Municipal d'Investigacio´ Medica-Hospital del Mar, Universitat Autonoma de Barcelona, Barcelona, Spain.

[19] Program in Inflammation and Cardiovascular Disorders, Institut Municipal d'Investigacio´ Medica-Hospital del Mar, Universitat Autonoma de Barcelona, Barcelona, Spain.

[20] Department of Neurology, Medical University of Graz, Austria.

[21] Neuroscience Institute, Saint Francis Medical Center, School of Health and Medical Sciences, Seton Hall University, South Orange, NJ.

[22] Department of Neurology, Miller School of Medicine, University of Miami, Miami, FL.

[23] Psychiatric Genetics Unit, Group of Psychiatry, Mental Health and Addiction, Vall d'Hebron Research Institute (VHIR), Universitat Autònoma de Barcelona, Barcelona, Spain.

[24] Department of Psychiatry, Hospital Universitari Vall d'Hebron, Barcelona, Spain.

[25] Biomedical Network Research Centre on Mental Health (CIBERSAM), Instituto de Salud Carlos III, Barcelona, Spain.

[26] Institute for Stroke and Dementia Research, Klinikum der Universität München, Ludwig-Maximilians-Universität München, Munich, Germany.

[27] Munich Cluster for Systems Neurology (SyNergy), Munich, Germany

[28] Department of Neurology, Radiology, and Biomedical Engineering, Washington University School of Medicine, St. Louis, MO, USA.

[29] Department of Neurology, University of Maryland School of Medicine, Baltimore, MD, USA.

[30] Department of Neurology and Rehabilitation Medicine and Comprehensive Stroke Center, University of Cincinnati, OH, USA

Corresponding Author:

Dr Matthew Traylor

Stroke Research Group

R3, Box 83

Department of Clinical Neurosciences

University of Cambridge

CB2 0QQ,

United Kingdom

+44 01273 217697

mt628@medschl.cam.ac.uk

Journal Subject Terms: Genetic, Association Studies; Intracranial Hemorrhage; Ischemic Stroke; Cardiovascular Disease

**Abstract**

Background: Genome-wide association studies have identified multiple loci associated with stroke. However, the specific stroke subtypes affected, and whether loci influence both ischaemic and haemorrhagic stroke, remains unknown. For loci associated with stroke, we aimed to infer the combination of stroke subtypes likely to be affected, and in doing so assess the extent to which such loci have homogeneous effects across stroke subtypes.

Methods: We performed Bayesian multinomial regression in 16,664 stroke cases and 32,792 controls of European ancestry to determine the most likely combination of stroke subtypes affected for loci with published genome-wide stroke associations, using model selection. Cases were subtyped under two commonly used stroke classification systems, Trial of Org 10172 Acute Stroke Treatment (TOAST) and Causative Classification of Stroke (CCS). All individuals had genotypes imputed to the Haplotype Reference Consortium 1.1 Panel.

Results: Sixteen loci were considered for analysis. Seven loci influenced both haemorrhagic and ischaemic stroke, three of which influenced ischaemic and haemorrhagic subtypes under both TOAST and CCS. Under CCS, 4 loci influenced both small vessel stroke and intracerebral haemorrhage. An *EDNRA* locus demonstrated opposing effects on ischaemic and haemorrhagic stroke. No loci were predicted to influence all stroke subtypes in the same direction and only one locus (12q24) was predicted to influence all ischaemic stroke subtypes.

Conclusions: Heterogeneity in the influence of stroke-associated loci on stroke subtypes is pervasive, reflecting differing causal pathways. However, overlap exists between haemorrhagic and ischaemic stroke, which may reflect shared pathobiology predisposing to small vessel arteriopathy. Stroke is a complex, heterogeneous disorder requiring tailored analytic strategies to decipher genetic mechanisms.

Keywords: Stroke, Multinomial, EDNRA, Genetics, intracerebral haemorrhage

## Introduction

The burden of stroke on global healthcare and society is substantial; it is consistently one of the leading causes of death and disability worldwide, [1] and a major cause of cognitive impairment and dementia. However, there exist significant gaps in our understanding of the pathological processes that underlie the disease. In recent years genome-wide association studies (GWAS) have made considerable advances in identifying genetic components underlying complex traits, in many cases identifying novel disease pathways and treatments.[2]

Characterizing the genetic component to stroke has been challenging, in part due to clinical heterogeneity, with at least three distinct major pathological processes (cardioembolism, large artery atherosclerosis, small vessel disease) underlying the majority of ischaemic strokes; and two processes underlying the majority of intracerebral haemorrhagic stroke (small vessel disease and cerebral amyloid angiopathy). [3, 4] However, recent GWAS have made considerable advances; 32 independent genome-wide significant loci were identified in the MEGASTROKE project. [5] The majority of these loci were identified as being associated with inclusive 'all stroke' or 'ischaemic stroke' categories, rather than specific stroke subtypes. This is in part due to study design, with much larger samples for these broader categories and only a fraction of stroke cases having detailed phenotyping. Indeed, this finding is in contrast to earlier studies that identified loci such as *HDAC9, PITX2* as being associated with specific subtypes. [6, 7] In order to interpret genetic risk associations in the context of biological mechanisms, a pertinent question is whether the newly identified stroke-associated loci truly confer risk across all stroke subtypes, or whether isolated or combinations of subtypes are affected. At least one of the novel variants (on chromosome 1q22) shows association with both ischaemic and haemorrhagic stroke, which might point to some shared mechanisms underlying these clinically distinct entities, which have thus far been separated in genetic studies.

Conventional approaches to GWAS, which employ within study analysis and subsequent meta-analysis across groups, do not enable detailed model comparison across different subgroups. In this analysis, we used multinomial logistic regression on well-characterized subjects with individual-level data to investigate the association of all identified genetic GWAS loci to date with all stroke subtypes (cardioembolic (CES), large artery stroke (LAS), small vessel stroke (SVS) and intracerebral haemorrhage (ICH)), determining the most likely combination of stroke subtypes affected at each locus. We performed our analysis using two established subtyping approaches: the Trial of Org 10172 in Acute Stroke Treatment (TOAST), [8] and Causative Classification of Stroke (CCS) system,[9] to provide a comprehensive account of these loci across available classification systems. Our overall aim was to evaluate genetic loci identified in previous studies using stroke datasets with well-defined phenotyping to determine if subtype specificity or cross-subtype associations could be identified.

## Methods

In order to minimize the possibility of unintentionally sharing information that can be used to re-identify private information, a subset of the data generated for this study are available at dbGAP and can be accessed at https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000615.v1.p1.

All contributing studies were approved by institutional review committees; subjects gave informed consent.

Full methods are provided in the Data Supplement.

## Results

After QC, there were up to 16,664 cases and 32,792 controls remaining for analysis (Table 1). In the merged dataset, a binomial genome-wide analysis of all cases against controls had a genomic inflation lambda=1.09, while the LDSCORE intercept value was 1.04, [10] suggesting that the majority of inflation was due to polygenicity and that any bias introduced by merging the datasets was minimal.

Sixteen loci contained SNPs with log(Bayes factors) of at least 4 in analyses of alternative stroke classification systems: Trial of Org 10172 Acute Stroke Treatment Classification System (TOAST) or Causative Classification of Stroke System (CCS) (causative system). We took these sixteen loci forward for further model selection. Plots for all loci under each classification system are provided in Supplementary Figures 1-16. For each of the sixteen loci, we identified the most likely combination of associated phenotypes at each locus (Figure 1, Table 2) based on model selection. A comparison of odds ratios for analysed loci from MEGASTROKE and the most recent ICH publication with those from our analysis showed high consistency ($r^2$=0.95, Supplementary Figure 17) despite slightly differing samples. LD values between our lead and previously published SNPs for the 16 loci in this analysis are provided in Supplementary Table 1.

For seven loci, the combination of phenotypes most likely to be influenced by the lead genetic variant at the loci included both ischaemic and haemorrhagic stroke subtypes. Four of these are shown in Figure 2. At these four loci: *EDNRA*, *1q22*, *MMP12*, *SH3PXD2A*, the ischaemic subtype included SVS, highlighting shared mechanisms underlying ICH and SVS, likely through predisposition to cerebral small vessel disease. At the *EDNRA* locus, the direction of association for ICH was opposite to that for LAS and SVS, pointing to contrasting influence on ischaemic and haemorrhagic stroke risk. We explored whether ICH-associated loci were specific to deep or lobar ICH. As in previous reports, [11, 12] associations at 1q22 and *COL4A2*

appear to be specific to deep ICH, with no effect in lobar ICH. For other regions, the evidence for specificity was more equivocal (Supplementary Table 2).

For four loci: *HDAC9*, *PITX2*, *ZFHX3*, *ANK2,* only one phenotype was affected by the lead variant (Figure 1, Supplementary Figures 10, 13, 16, 5) in the most likely configuration across all classification systems. Several other loci: 9p21, 12q24, 16q24, *FOXF2* were associated with only one phenotype under particular classification systems, but did not show consistency across TOAST and CCS (Supplementary Figures 2, 3, 4, 9). For *TSPAN2*, which was previously identified as being associated with LAS, [13] the best-fit model also included CES under CCS, albeit with a much weaker effect than LAS (rs17479660; CES, OR=1.08; LAS, OR=1.19 under CCS). Echoing previous results, the locus showed much stronger significance under CCS classifications than under TOAST (Supplementary Figure 15).

For *COL4A2*, the strongest association found under TOAST was for rs9515201. The most likely model contained ICH (OR=1.14) and SVS (OR=1.13), consistent with findings from previous analyses. [12] However, under CCS an alternate SNP, rs1927349, was the strongest associated. No association with SVS was observed, and a weak association with CES was observed instead. Reasons for this discrepancy between CCS and TOAST are not immediately clear, but non-overlapping samples between the two classification systems are a likely factor.

The mean (SD) number of stroke subtypes affected at each locus were 1.88 (0.89) under TOAST and 1.69 (0.87) under CCS. Under CCS, the most common combination of affected subtypes was SVS and ICH (4 loci).

## Discussion

We performed a large-scale genetic analysis, characterising the effects of established stroke risk loci with ischaemic and haemorrhagic stroke subtypes in up to 16,664 cases and 32,792 controls. Our main findings are twofold. First, for the vast majority of loci studied, multiple but never all stroke subtypes were affected at the locus. Only one locus (12q24) was assumed to influence all ischaemic stroke subtypes. This indicates that although these loci were identified in analyses of inclusive stroke phenotypes, in the main their effects are specific to particular combinations of stroke subtypes. The mean number of subtypes affected was 1.88 for TOAST and 1.69 for CCS classification systems. Notable exceptions were the *PITX2* and *ZFHX3* loci, which were associated with cardioembolic stroke most likely through atrial fibrillation (for which they are well-established loci [14]), and *HDAC9* which is associated with large vessel stroke. Under TOAST, the *FOXF2* locus was associated solely with SVS. However, under CCS, LAS was also implicated. For CCS, the 9p21 locus was predicted to influence only LAS. However, under TOAST, SVS was also implicated. Our analyses suggest that *ANK2* confers risk of stroke predominantly through its influence on *ICH.* We were unable to identify any loci for which the most likely model included all stroke phenotypes in the same direction and only one (12q24) which for which the most likely model included all ischaemic stroke subtypes.

Secondly, we find evidence that several loci influence both haemorrhagic and ischaemic stroke. This was evident for seven loci in total (1q22, *COL4A2*, *EDNRA*, *LINC01492*, *MMP12*, *SH3PXD2A*, *CDK6*). Under CCS, 4 loci (*SH3PXD2A*, *MMP12*, *EDNRA*, 1q22) influenced both SVS and ICH, highlighting shared mechanisms underlying small vessel disease. Previous GWAS analyses have tended to separate ischaemic and haemorrhagic stroke on the basis of presumed differing etiologies. Our results suggest that including haemorrhagic alongside ischaemic stroke in multiphenotype analyses will provide further insights.

For one locus: Endothelin Receptor Type A (*EDNRA*), the association with ICH was in the opposite direction to the ischaemic stroke subtypes, suggesting opposing risk mechanisms. This locus has previously been associated with a variety of vascular phenotypes, including coronary artery disease, carotid plaques, and peripheral arterial disease (all in concordant direction with ischaemic stroke), as well as intracranial aneurysm (in concordant direction with intracerebral haemorrhage). [15-18] The locus has also been associated with migraine in candidate gene studies, [19] but this has not been validated in GWA studies and is likely a false positive. [20] *EDNRA* encodes the type A receptor ($ET_A$) for Endothelin-1 (*ET-1*), a potent vasoconstrictor with pro-inflammatory effects. $ET_A$ -specific antagonists increase Nitric Oxide (NO)-mediated endothelium-dependent relaxation, reduce *ET-1* levels and inhibit atherosclerosis in mice, [21] suggesting that higher levels of $ET_A$ are pro-atherogenic: consistent with the observation that higher $ET_A$ levels are observed in atherosclerotic plaques. [22] Based on this, one might expect the *EDNRA* risk variant (C allele of rs17612742 in this study) to lead to increased risk of ischaemic stroke through elevated $ET_A$ levels. Indeed, in GWA studies of intracranial aneurysm the susceptibility variant (in LD with the T allele of rs17612742 in our study) was shown to result in higher transcription factor binding affinity, likely resulting in repression of the transcriptional activity of *EDNRA*. [17] This suggests that carries of the C allele have lower levels of *EDNRA,* which consequently higher *ET-1* levels and greater susceptibility to atherosclerosis. The reason why for carriers of T allele lower levels of $ET_A$ might promote intracranial aneurysm and intracerebral haemorrhage is not immediately obvious, but several mechanisms are possible. Levels of *ET-1* have been linked to vascular remodelling, an important process underlying ICH and IA; [23, 24] subtle changes in this process induced by altered availability of $ET_A$ is one such mechanism. Deep ICH and ischaemic SVS arise due to the same arteriopathy that arises in the deep perforating arteries of the brain. The *EDNRA* variant in this study points to a mechanism that influences whether the resulting pathology is ischaemic or haemorrhagic, and as such warrants further detailed investigation.

Some loci were notably more significant when phenotyped using CCS; *SH3XPD2A, MMP12*, *TSPAN2, FOXF2, EDNRA,* which might point to CCS having greater accuracy and therefore utility in stroke GWA studies. However, the opposite was also true for others: *16q24, HDAC9*. We note that some differences may be due to the fact that not all individuals were subtyped under both CCS and TOAST; the TOAST cohort was a least 20% larger. A detailed discussion of the relative merits of TOAST and CCS is beyond the scope of this article, but our results highlight that the importance of collecting individual phenotypic qualities that make up the etiologic subtypes in genetic studies of stroke so that associated loci can be more systematically examined.

Our study has several strengths. The dataset was a large stroke population including intracerebral haemorrhage and ischaemic stroke cases, the majority of which were subtyped under both TOAST and CCS. We had full access to genotype-level data enabling us full control over all analyses. The implementation of a multinomial regression approach enabled us to systematically assess which stroke subtypes were likely to be affected at each locus, which would not be formally possible under standard binomial regression approaches which analyse each stroke subtype separately. Ultimately, mechanistic studies will be required to determine the influence of associated genetic variants, but analyses such as this have utility in directing the focus and model systems suitable for such follow up studies.

Similarly there are limitations. We present results for the most likely combination of stroke phenotypes affected at each locus: the 'best-fitting' model. We had limited statistical power to determine with statistical certainty that this was the correct model; significantly larger samples would be required to achieve this. One consequence of this is that there remains the potential that some associations are due to random variation rather than true biological differences. It

would therefore be prudent to treat some of the findings here as preliminary until confirmed in larger samples. Due to the challenges of performing these analyses across different ancestry populations, and as we only had a small number of non-European ancestry ICH cases available which could lead to overfitting, we performed analyses in European populations only. The results can therefore not be generalized to all populations. Repeating these analyses once sufficient data from other ancestral groups are available should be highly prioritized to ensure advancements in the field are made for all ancestral groups. In all analyses we assume there is a single causal variant at the locus, which may not be true in all cases. Our analyses are based on use of a default prior, which has been used in many genetic studies. An alternative is to derive an empirical prior from associated genetic loci. As more loci are identified as being associated with stroke, this will become a more realistic possibility and should be explored in future analyses.

## Conclusions

Our findings suggest that although large scale genome-wide studies of broad 'all stroke' or 'all ischaemic stroke' phenotypes are able to identify multiple associations, it should not be assumed that such associations confer risk equally across stroke subtypes. Heterogeneity in the influence of genetic variants on different stroke subtypes is the norm, not the exception. The multinomial regression approach used here provided insights into the etiological stroke subtypes most prominently influenced by genetic variants at these loci – a prerequisite to decide on the most appropriate model systems to choose for further mechanistic studies. Stroke is a complex, heterogeneous disorder: our findings highlight the ongoing need for large, well phenotyped case collections and tailored analytic strategies to decipher the underlying genetic mechanisms.

## Sources of Funding

This work was supported by a British Heart Foundation Programme Grant (RG/16/4/32218). The National Institute of Neurological Disorders and Stroke – Stroke Genetics Network (NINDS-SIGN) study was funded by the US National Institute of Neurological Disorders and Stroke, National Institutes of Health (U01 NS069208 and R01 NS100178). Collection of the UK Young Lacunar Stroke DNA Study (DNA Lacunar) was primarily supported by the Wellcome Trust (WT072952) with additional support from the Stroke Association (TSA 2010/01). Genotyping of the DNA Lacunar samples was supported by a Stroke Association Grant (TSA 2013/01). The principal funding for the WTCCC2 stroke study was provided by the Wellcome Trust, as part of the Wellcome Trust Case Control Consortium 2 project (085475/B/08/Z and 085475/Z/08/Z and WT084724MA). Hugh Markus is supported by a National Institute for Health Research (NIHR) Senior Investigator award, and his work is supported by the Cambridge Universities NIHR Comprehensive Biomedical Research Centre. Dr. Anderson is supported by NIH R01NS103924 and K23NS086873.

## Author's Contributions

MT and RM designed the experiments. MT and MC performed the imputations. MT performed the statistical analyses. MT, CDA, LCARJ, HSM, DW, and RM wrote the first draft of the manuscript. All authors read and approved the final manuscript.

## Ethics approval and consent to participate

All research participants contributing clinical and genetic samples for analysis in this study provided written informed consent.

## Availability of data and materials

Data from the NINDS-SIGN Stroke study are available to researchers through dbGAP: https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000615.v1.p1.

Trinculo v0.96 is available from: https://sourceforge.net/projects/trinculo/files/.

MEGASTROKE data is available from http://megastroke.org.

**Competing interests**

Dr. Anderson has consulted for ApoPharma, Inc.

## References


1. Global, regional, and national disability-adjusted life-years (DALYs) for 333 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet*. 2017;390:1260-1344.
2. Visscher PM, et al. 10 Years of GWAS Discovery: Biology, Function, and Translation. *Am J Hum Genet*. 2017;101:5-22.
3. Martini SR, et al. Risk factors for intracerebral hemorrhage differ according to hemorrhage location. *Neurology*. 2012;79:2275-82.
4. Hankey GJ. Stroke. *Lancet*. 2017;389:641-654.
5. Malik R, et al. Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nat Genet*. 2018;50:524-537.
6. Bellenguez C, et al. Genome-wide association study identifies a variant in HDAC9 associated with large vessel ischemic stroke. *Nat Genet*. 2012;44:328-33.
7. Traylor M, et al. Genetic risk factors for ischaemic stroke and its subtypes (the METASTROKE collaboration): a meta-analysis of genome-wide association studies. *Lancet Neurol*. 2012;11:951-62.
8. Adams HP, Jr., 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.
9. Ay H, et al. A computerized algorithm for etiologic classification of ischemic stroke: the Causative Classification of Stroke System. *Stroke*. 2007;38:2979-84.
10. Bulik-Sullivan BK, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet*. 2015;47:291-5.
11. Woo D, et al. Meta-Analysis of Genome-Wide Association Studies Identifies 1q22 as a Susceptibility Locus for Intracerebral Hemorrhage. *Am J Hum Genet*. 2014;94:511-21.

12. Rannikmae K, et al. COL4A2 is associated with lacunar ischemic stroke and deep ICH: Meta-analyses among 21,500 cases and 40,600 controls. *Neurology*. 2017;89:1829-1839.
13. Loci associated with ischaemic stroke and its subtypes (SiGN): a genome-wide association study. *Lancet Neurol*. 2016;15:174-184.
14. Roselli C, et al. Multi-ethnic genome-wide association study for atrial fibrillation. *Nat Genet*. 2018;50:1225-1233.
15. Bis JC, et al. Meta-analysis of genome-wide association studies from the CHARGE consortium identifies common variants associated with carotid intima media thickness and plaque. *Nat Genet*. 2011;43:940-7.
16. Howson JMM, et al. Fifteen new risk loci for coronary artery disease highlight arterial-wall-specific mechanisms. *Nat Genet*. 2017;49:1113-1119.
17. Low SK, et al. Genome-wide association study for intracranial aneurysm in the Japanese population identifies three candidate susceptible loci and a functional genetic variant at EDNRA. *Hum Mol Genet*. 2012;21:2102-10.
18. Matsukura M, et al. Genome-Wide Association Study of Peripheral Arterial Disease in a Japanese Population. *PLoS One*. 2015;10:e0139262.
19. Tzourio C, et al. Association between migraine and endothelin type A receptor (ETA - 231 A/G) gene polymorphism. *Neurology*. 2001;56:1273-7.
20. Gormley P, et al. Meta-analysis of 375,000 individuals identifies 38 susceptibility loci for migraine. *Nat Genet*. 2016;48:856-66.
21. Barton M, et al. Endothelin ETA receptor blockade restores NO-mediated endothelial function and inhibits atherosclerosis in apolipoprotein E-deficient mice. *Proc Natl Acad Sci U S A*. 1998;95:14367-72.
22. Ramzy D, et al. Elevated endothelin-1 levels impair nitric oxide homeostasis through a PKC-dependent pathway. *Circulation*. 2006;114:I319-26.

388 23. Penn DL, et al. The role of vascular remodeling and inflammation in the pathogenesis
389 of intracranial aneurysms. *Journal of clinical neuroscience : official journal of the*
390 *Neurosurgical Society of Australasia*. 2014;21:28-32.
391 24. Amiri F, et al. Endothelium-restricted overexpression of human endothelin-1 causes
392 vascular remodeling and endothelial dysfunction. *Circulation*. 2004;110:2233-40.

393

394

## Tables and Figures

**Table 1.** Sample Sizes

| | TOAST | | | CCS | | |
|---|---|---|---|---|---|---|
| | N | Age (mean(SD)) | Male (%) | N | Age (mean(SD)) | Male (%) |
| CES | 3847 | 72(14) | 49 | 2826 | 75(12) | 44 |
| LAS | 2803 | 68(12) | 65 | 2204 | 67(12) | 62 |
| SVS | 3976 | 64(13) | 62 | 3093 | 63(13) | 62 |
| UND | 4085 | 65(16) | 54 | 4013 | 65(15) | 53 |
| ICH | 1953 | 71(13) | 53 | 1953 | 71(13) | 53 |
| Controls | 32792 | 62(17) | 46 | 28052 | 62(17) | 48 |

CES, cardioembolic Stroke; LAS, large artery atherosclerotic stroke; SVS, small artery occlusion stroke; UND, stroke of undetermined etiology; ICH, intracerebral haemorrhage; TOAST, Trial of Org 10172 Acute Stroke Treatment Classification System; CCS, Causative Classification of Stroke System (causative system). Age available not available for controls from WTCCC2 studies.

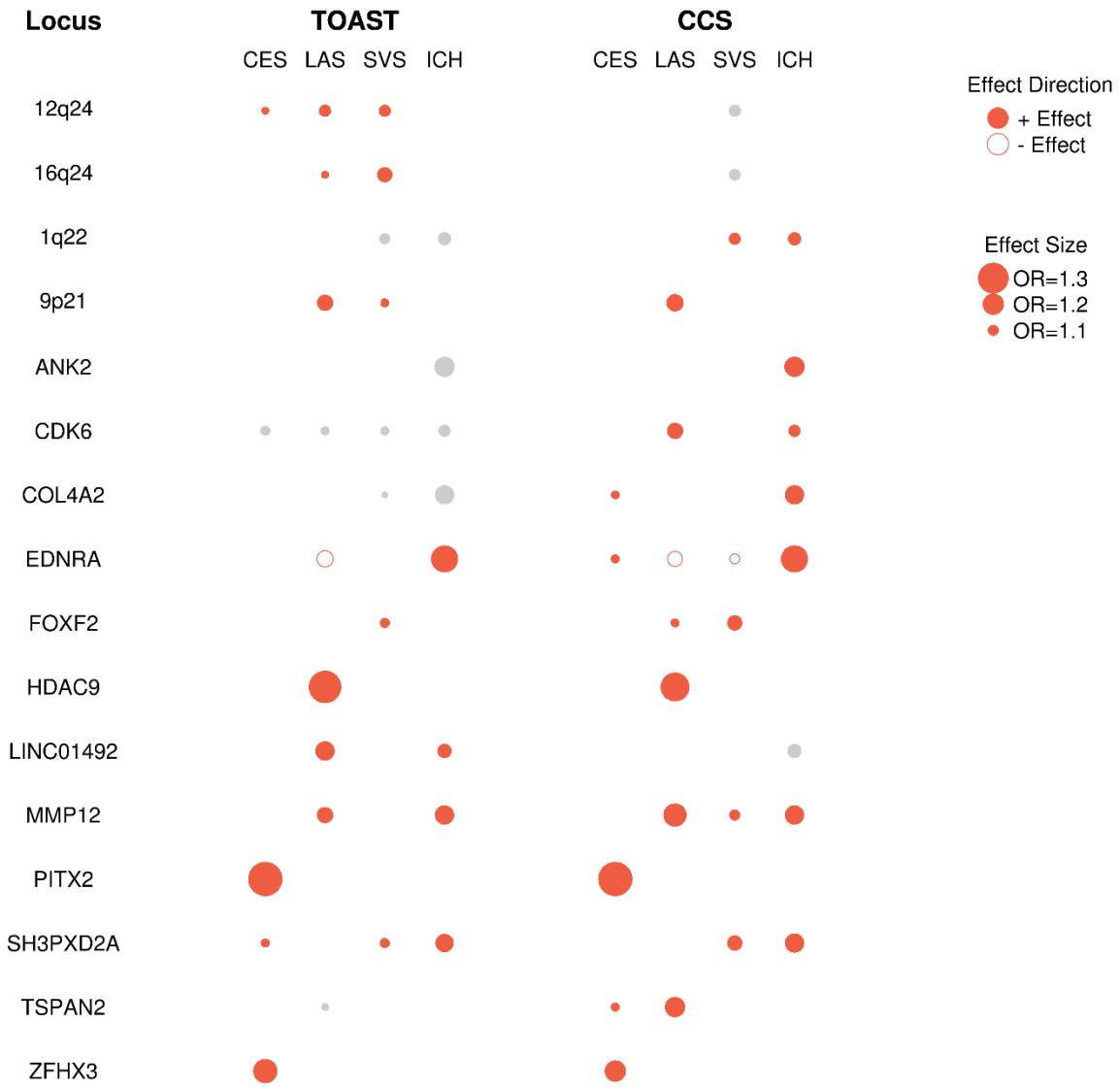
**Table 2 – Lead SNPs, Association Statistics, and Affected Stroke Subtypes for Each Locus**

| Locus | Lead SNP [Best Model] | log OR (SE) | log BF | Subtypes in Best Fitting Model |
|---|---|---|---|---|
| 1q22 | rs2758603 [CCS] | 0.10 (0.03) SVS<br>0.11 (0.05) ICH<br>0.02 (0.03) CES<br>0.07 (0.03) UNK<br>0.07 (0.03) LAS | 4.0 | SVS, ICH |
| 9p21 | rs1412830 [TOAST] | 0.08 (0.03) SVS<br>0.07 (0.04) ICH<br>-0.01 (0.03) CES<br>0.03 (0.03) UNK<br>0.14 (0.03) LAS | 5.7 | LAS, SVS |
| 12q24 | rs10774624 [TOAST] | 0.10 (0.03) SVS<br>-0.03 (0.05) ICH<br>0.07 (0.03) CES<br>0.07 (0.03) UNK<br>0.10 (0.03) LAS | 5.8 | CE, LAS, SVS |
| 16q24 | rs12445022 [TOAST] | 0.13 (0.03) SVS<br>0.05 (0.05) ICH<br>-0.01 (0.03) CES<br>0.07 (0.03) UNK<br>0.07 (0.03) LAS | 5.8 | LAS, SVS |
| ANK2 | rs149538932 [CCS] | 0.07 (0.03) SVS<br>0.18 (0.05) ICH<br>0.04 (0.03) CES<br>0.08 (0.03) UNK<br>0.02 (0.03) LAS | 6.4 | ICH |
| CDK6 | rs4272 [CCS] | 0.05 (0.04) SVS<br>0.10 (0.05) ICH<br>0.07 (0.03) CES<br>0.12 (0.03) UNK<br>0.14 (0.04) LAS | 8.5 | LAS, ICH |
| COL4A2 | rs1927349 [CCS] | -0.02 (0.03) SVS<br>0.16 (0.05) ICH<br>0.08 (0.03) CES<br>0.04 (0.03) UNK<br>0.02 (0.03) LAS | 5.0 | CES, ICH |
| EDNRA | rs17612742 [CCS] | 0.09 (0.04) SVS<br>-0.23 (0.06) ICH<br>-0.08 (0.04) CES<br>-0.00 (0.04) UNK<br>0.13 (0.04) LAS | 10.5 | CES, LAS, SVS, ICH |

| | | | | |
|---|---|---|---|---|
| FOXF2 | rs11242678 [CCS] | 0.13 (0.03) SVS<br>-0.05 (0.05) ICH<br>0.07 (0.03) CES<br>0.09 (0.03) UNK<br>0.09 (0.04) LAS | 7.4 | LAS, SVS |
| HDAC9 | rs2107595 [TOAST] | 0.04 (0.04) SVS<br>-0.08 (0.06) ICH<br>0.05 (0.04) CES<br>0.06 (0.03) UNK<br>0.27 (0.04) LAS | 19.2 | LAS |
| LINC01492 | rs10990643 [TOAST] | -0.02 (0.04) SVS<br>0.12 (0.06) ICH<br>0.03 (0.04) CES<br>0.01 (0.03) UNK<br>0.17 (0.04) LAS | 4.1 | LAS, ICH |
| MMP12 | rs470234 [CCS] | 0.09 (0.04) SVS<br>0.17 (0.06) ICH<br>0.04 (0.04) CES<br>0.03 (0.04) UNK<br>0.20 (0.04) LAS | 8.7 | LAS, SVS, ICH |
| PITX2 | rs2723334 [TOAST] | 0.0 (0.04) SVS<br>0.08 (0.06) ICH<br>0.29 (0.04) CES<br>0.03 (0.03) UNK<br>-0.03 (0.04) LAS | 48.0 | CES |
| SH3PXD2A | rs10883922 [CCS] | 0.13 (0.03) SVS<br>0.16 (0.05) ICH<br>0.02 (0.03) CES<br>0.02 (0.03) UNK<br>0.04 (0.03) LAS | 6.0 | SVS, ICH |
| TSPAN2 | rs7537796 [CCS] | -0.05 (0.03) SVS<br>-0.06 (0.05) ICH<br>0.06 (0.03) CES<br>-0.02 (0.03) UNK<br>0.14 (0.03) LAS | 6.8 | CES, LAS |
| ZFHX3 | rs67329386 [TOAST] | -0.02 (0.03) SVS<br>-0.05 (0.05) ICH<br>0.20 (0.03) CES<br>0.02 (0.03) UNK<br>0.00 (0.03) LAS | 13.8 | CES |

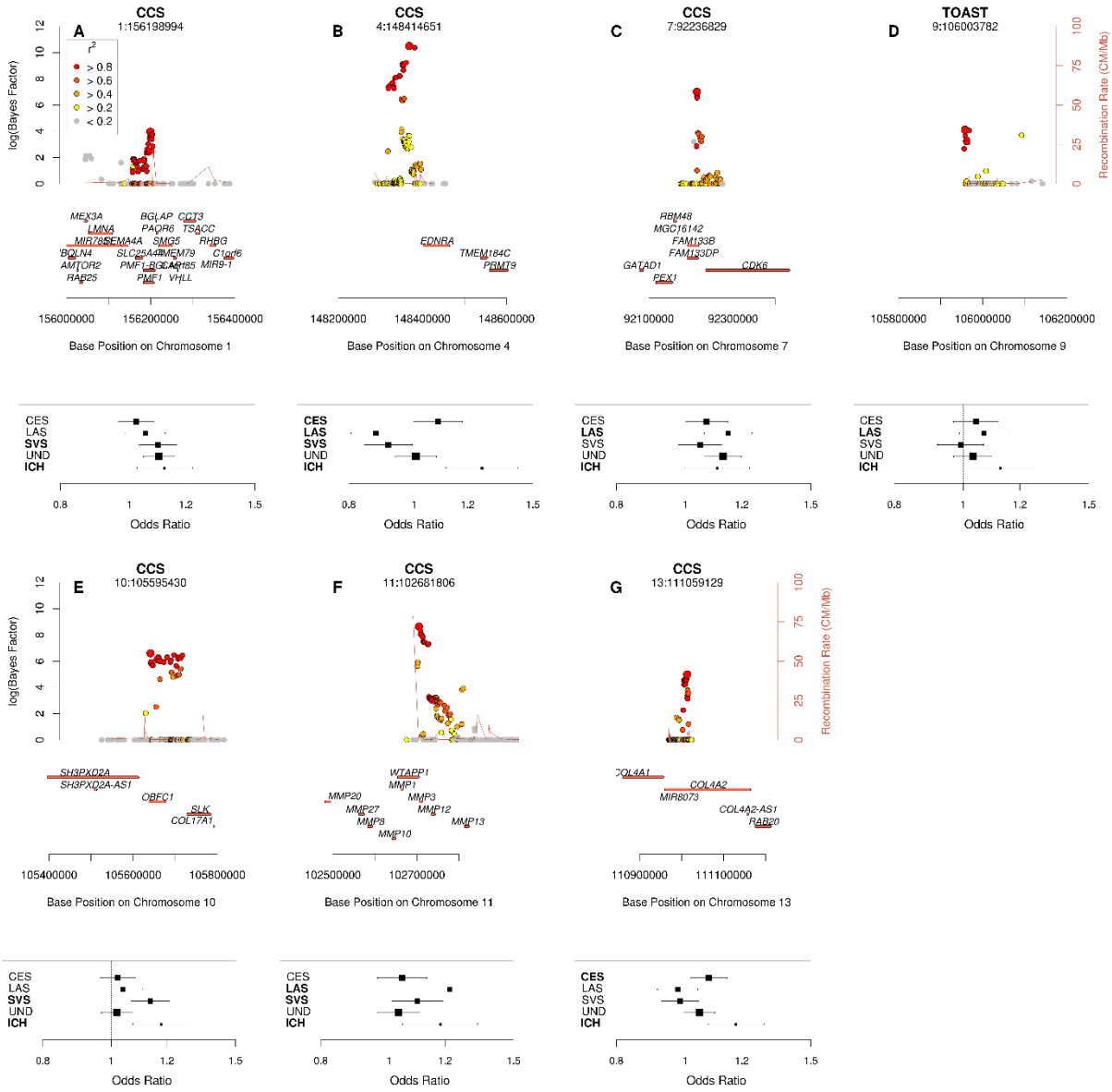
406 CES, Cardioembolic Stroke; LAS, Large artery Stroke; SVS, Small Vessel Stroke; ICH,
407 Intracerebral Haemorrhage; log BF, log transform of Bayes Factor; log OR, log transform of
408 Odds Ratio; SE, standard error; CCS, causative classification system of stroke; TOAST, Trial
409 of Org 10172 Acute Stroke Treatment Classification System

**Figure 1.** Stroke Subtypes in Best Fitting Model at Each Locus, for CCSc, CCSp, and TOAST classification Systems, with Size Weighted by Association Odds Ratio




CES, Cardioembolic Stroke; LAS, Large artery Stroke; SVS, Small Vessel Stroke; ICH, Intracerebral Haemorrhage. Results are presented for the 16 loci showing log(Bayes Factor)>4 in CCS or TOAST analyses. Classification/Locus combinations in grey indicate that the locus did not reach log(Bayes Factor)>4 in that analysis.

**Figure 2.** Local Plots showing Associations with Regions Conferring Risk of Ischaemic and Haemorrhagic Stroke and Odds Ratios for all stroke Subtypes




A, 1q22 region; B, EDNRA region; C, CDK6 region; D, LINC01492 region; E, SH3PXD2A region; F, MMP12 region; G, COL4A2 region; CE, cardioembolic stroke; LAS, large artery atherosclerotic stroke; SVS, small vessel stroke; ICH, intracerebral haemorrhage. Results are presented for the classification system in which the locus showed strongest significance. Stroke subtypes in bold indicate those included in the best fitting model and therefore predicted to be influenced by the lead genetic variant, based on Bayesian model selection.