17q25 locus is associated with white matter hyperintensity volume in ischemic stroke, but not with lacunar stroke status

Poneh Adib-Samii MBBS[1*], Natalia Rost MD[2,3*], Matthew Traylor MSc[1], William Devan BS[2] , Alessandro Biffi MD[2], Silvia Lanfranconi MD[1], Kaitlin Fitzpatrick BSc[2], Steve Bevan PhD[1], Allison Kanakis BS[2], Valerie Valant BA[2], Andreas Gschwendtner MD[4], Rainer Malik PhD[4], Alexa Richie MPH[5], Dale Gamble MHSc[5], Helen Segal PhD[6], Eugenio A Parati MD[7], Emilio Ciusani PhD[8], Elizabeth G Holliday PhD[9], Jane Maguire PhD[10], Joanna Wardlaw MD[11], Bradford Worrall MD[12], Unnur Thorsteinsdottir PhD[13,14], Kari Stefansson PhD[13,14], Gudmar Thorleifsson PhD[13], Joshua Bis PhD[15], Kerri Wiggins RD[15], Will Longstreth MD[15], Steve J Kittner MD[16,17], Yu-Ching Cheng PhD[16], Thomas Mosley PhD[18], Guido J Falcone MD[2], Karen L Furie MD[19], Carlos Leiva-Salinas MD[20], Benison C Lau BS[20], Muhammed Saleem Khan MSc[21], Australian Stroke Genetics Collaborative[22], Wellcome Trust Case-Control Consortium-2(WTCCC2)[23], METASTROKE, Pankaj Sharma PhD[21], Myriam Fornage PhD[24], Braxton D Mitchell PhD[16], Bruce M Psaty PhD[15,25], Solveig Gretarsdottir PhD[13],Cathie Sudlow DPhil[11], Christopher Levi MD[9], Giorgio B. Boncoraglio MD[7], Peter M Rothwell FMedSci[6], James Meschia MD[5], Martin Dichgans MD[4] , Jonathan Rosand MD[2,26+] Hugh S Markus DM[1+] on behalf of the International Stroke Genetics Consortium.

*These authors contributed equally and share authorship.

+These authors contributed equally

Affiliations:

[1]Stroke and Dementia Research Centre, St George's University of London, UK

[2]Center for Human Genetic Research and Department of Neurology, Massachusetts General Hospital, Boston, USA

[3]Department of Neurology, Boston University School of Medicine, Boston, MA, USA

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

[5]Department of Neurology, Mayo Clinic, Jacksonville, USA

[6]Stroke Prevention Research Unit, Nuffield Department of Neuroscience, University of Oxford, UK

[7]Department of Cerebrovascular Diseases, Fondazione IRCCS Istituto Neurologico "Carlo Besta", Milano, Italy

[8]Laboratory of Clinical Pathology and Medical Genetics, Fondazione IRCCS Istituto Neurologico "Carlo Besta", Milano, Italy

[9]Centre for Clinical Epidemiology and Biostatistics, Hunter Medical Research Institute and School of Medicine and Public Health, University of Newcastle, NSW, Australia

[10]Priority Research Centre for Translational Neuroscience and Mental Health and Faculty of Health, School of Nursing and Midwifery, University of Newcastle, NSW, Australia

[11]Division of Clinical Neurosciences, Neuroimaging Sciences and Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK

[12]University of Virginia, Charlottesville, USA

[13]deCODE Genetics, Reykjavik, Iceland

[14]Faculty of Medicine, University of Iceland, Reykjavik, Iceland

[15]Cardiovascular Health Research Unit, University of Washington, Seattle, USA

[16]Department of Neurology, University of Maryland School of Medicine, USA

[17]Veterans Affairs Medical Center, Baltimore, Maryland, USA

[18]University of Mississippi Medical Center, Jackson, USA

[19]Department of Neurology, Brown University

[20]Department of Radiology, University of Virginia, Charlottesville, USA

[21]Imperial College Cerebrovascular Research Unit (ICCRU), Imperial College London, UK

[22]Australian Stroke Genetics Collaborative and [23]Wellcome Trust Case Control Consortium-2 (WTCCC2) memberships are listed in Supplementary Material.

[24]University of Texas Health Science Center at Houston; Houston, TX, USA

[25]Group Health Research Institute, Group Health, Seattle, WA, USA.

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

Corresponding Author:

Hugh Markus

St Georges, University of London

SW17 0RE

email: hmarkus@sgul.ac.uk

Keywords: Stroke, Small Vessel Disease, Leukoaraiosis, Genetics, Genome-wide Association Study

Title:120 characters

Abstract:232 words

Manuscript:4490

3 Figures, 3 Tables

ABSTRACT

BACKGROUND: Recently, a novel locus at 17q25 was associated with white matter hyperintensities (WMH) on magnetic resonance imaging (MRI) in stroke-free individuals. We aimed to replicate the association with WMH volume (WMHV) in patients with ischemic stroke. If the association acts by promoting a small vessel arteriopathy it might be expected to also associate with lacunar stroke.

METHODS: We quantified WMH on MRI in the stroke-free hemisphere of 2588 ischemic stroke cases. Association between WMHV and six single nucleotide polymorphisms (SNPs) at chromosome 17q25 was assessed by linear regression. These SNPs were also investigated for association with lacunar stroke, in 1854 cases and 51939 stroke-free controls from METASTROKE. Meta-analyses with previous reports and a genetic risk score approach were applied to identify other novel WMHV risk variants and uncover shared genetic contributions to WMHV in community-participants without stroke and ischemic stroke.

RESULTS: SNPs at 17q25 were associated with WMHV in ischemic stroke, the most significant being rs9894383 (p=0.0006). In contrast there was no association between any SNP and lacunar stroke. A genetic risk score analysis revealed further genetic components to WMHV shared between community-participants without stroke and ischemic stroke.

CONCLUSION: This study provides support for an association between the 17q25 locus and WMH. In contrast, it is not associated with lacunar stroke suggesting that the association does not act by promoting small vessel arteriopathy or the same arteriopathy responsible for lacunar infarction.

INTRODUCTION

A recent report by the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium identified a novel genetic locus at chromosome 17q25 associated with white matter hyperintensities (WMH) on MRI in stroke-free community individuals.[1,2] WMH are common in stroke-free adults and increase with age. Prospective studies show WMH predict increased risks of cognitive decline, stroke and death.[3]

One might expect risk factors for WMH in community populations to also confer increased risk of WMH in stroke patients. However, the underlying pathology of WMH is heterogeneous; small punctate lesions are associated with mixed etiologies, while more confluent areas correspond primarily to small vessel disease (SVD).[4] In stroke patients, WMH are more frequent and extensive compared with healthy age-matched individuals and are usually associated with SVD.[4]

We hypothesized that risk factors for WMH in community populations without stroke also confer increased risk of WMH in stroke patients. Therefore, we assessed whether the 17q25 locus is associated with WMH as a quantitative trait in ischemic stroke. If the 17q25 locus acted to increase WMH through promotion of the small vessel arteriopathy, one might expect it to also increase risk of lacunar stroke, another SVD phenotype. To test this hypothesis, its association with lacunar stroke status was also examined in a cases-control analysis.

METHODS & MATERIALS

A. Association of 17q25 locus with WMHV

Subjects

Stroke cases were recruited from one community- and 8 hospital-based cohorts of ischemic strokes with genome-wide association study (GWAS) data available as well as MRI scans for WMH volume (WMHV) measurement. Details of populations are in Table 1. Inclusion criteria were: age>18 years, self-reported European ancestry, and a diagnosis of ischemic stroke of any subtype. Exclusion criteria were CADASIL, vasculitis, demyelinating and mitochondrial disorders.

Clinical characteristics

Hypertension was defined as antihypertensive prescription prior to stroke or systolic blood pressure >140 or diastolic >90mmHg more than one week post-stroke. Hypercholesterolemia was defined as lipid-lowering treatment prior to stroke, and/or elevated serum cholesterol (>5.2mmol/l) on stroke admission. Ever-smoker was defined as current and ex-smokers. Diabetes was defined as a previous diagnosis of diabetes mellitus. Ischemic stroke cases were subtyped according to the TOAST criteria[5] or similar (Supplementary Table S1).

MRI analysis

MRI was acquired as part of routine clinical practice in stroke evaluation using different scanners at individual centers. Fluid attenuated inversion recovery (FLAIR) sequences were primarily used for WMH analysis; however, in their absence, T2 sequences were used (Supplementary Table S2).

WMHV was measured in the hemisphere contralateral to the acute infarction to avoid confounding by T2 hyperintense signal due to acute stroke. All supratentorial white matter and deep gray matter lesions were included in WMHV with the exception of WMH corresponding to lacunar infarcts. MRI scans with excessive movement artefact, incomplete brain coverage, or bihemispheric infarcts (other than lacunar) were excluded. To account for normal inter-individual variability in head size, an estimate of total intracranial volume (TICV) was derived, using site-specific volumetric methodology.

Scans were analyzed anonymously, blinded to genetic information. MRIs from the Massachusetts General Hospital (MGH), Ischemic Stroke Genetics Study (ISGS), and Australian Stroke Genetics Collaborative (ASGC) studies were analyzed in Boston. FLAIR sequences were analyzed using a MRIcro (www.mricro.com) semi-automated method as previously described.[6] Using operator-mediated quality assurances, overlapping regions of interest (ROIs) corresponding to WMH produced the final maps for WMHV calculation.[6] To adjust for head size, intracranial area (ICA) was used as a validated marker of TICV.[7] The ICA constituted the average of two mid-sagittal slices traced using anatomical landmarks on T1 sequences.[7] SWISS scans were analyzed in the same way at the University of Virginia by the Boston-trained rater.

The Wellcome Trust Case Control Consortium-2 (WTCCC2) and Milan cohorts were analyzed in London using DISPunc semi-automated lesion drawing software[8]. A 'seed' at the lesion border was manually marked, following which the program automatically outlined the lesion based on the signal intensity gradient. Each ROI was visually inspected and manually corrected as required. To estimate TICV, T2 and, in its absence, FLAIR sequences were analyzed using an

automated segmentation program, SIENAX[9], and summing the cerebrospinal fluid, gray and white matter volumes.

WMH quantification agreement across the two main reading centers was performed for 50 randomly selected scans; agreement was very good (intraclass correlation coefficient 0.95, CI 0.91-0.97 n=50).

Genotyping

WTCCC2-UK and German cases were genotyped on the Illumina Human660W-Quad. Milan cases were genotyped using Illumina Human610-Quad or Human660W-Quad. ASGC samples were genotyped using the Human610-Quad. Both ISGS and SWISS samples were genotyped using the Illumina650K-Quad. MGH were genotyped on the Affymetrix 6.0, Illumina Human610-Quad or Illumina OmniExpress beadchips.

For all cohorts, individuals were removed if their inferred sex was discordant with the recorded sex or if >5% missing genotype data. Autosomal single nucleotide polymorphisms (SNPs) were excluded for minor allele frequency <1%, >5% missing data, or Hardy-Weinberg equilibrium $p<1x10^{-6}$. Each center performed checks for relatedness and population stratification. After quality-control procedures there was an effective sample size of 2588.

Imputation was performed in all centers using IMPUTE2.[10] All centers were imputed to HapMap3 and 1000 Genomes Project Phase pilot (June 2010) with the exception of MGH Omni, which was imputed to 1000 Genomes Integrated Release (June 2011). Post imputation, poorly

imputed SNPs (r^2<0.3) were removed resulting in 2706548 SNPs common to all centers.

Genome-wide Analyses and Meta-Analyses

To account for differences in MRI image acquisitions and population characteristics, we used a joint-analysis strategy; each cohort was analyzed independently and then meta-analyzed. Stroke cases with T2 images were analyzed separately from the FLAIR images. MGH samples were also sub-grouped based on genotyping platform (Supplementary Table S2).

Single hemisphere WMHV was doubled to obtain whole brain values and adjusted for normal inter-individual variation in head size by multiplying by the ratio of mean TICV to individual TICV. Values were natural log-transformed to a normal distribution. Within each group, rank-transformed residuals were derived from a linear regression model predicting WMHV with age, sex, and the first two ancestry principle components as covariates in GenABEL.[11] Thus the phenotype was adjusted for age because WMHV is highly age-dependent.[3,4] Principle components, derived using EIGENSTRAT (http://genepath.med.harvard.edu/~reich/Software.htm), were included to correct for potential population stratification. GWA analysis was undertaken in PLINK (http://pngu.mgh.harvard.edu/~purcell/plink/) using pseudo-dosages, a fractional count of 0 to 1 alleles for each genotype weighted by imputation probability, within a linear regression (additive) model. A meta-analysis of all the groups was performed under a fixed-effects inverse-variance model within METAL.[12] SNPs showing heterogeneity ($p<0.05$) were removed.

Two further analyses were performed, in which WMHV was additionally adjusted for

hypertension alone or combined cardiovascular risk factors (hypertension, hypercholesterolemia, ever-smoker and diabetes). Individuals with missing cardiovascular risk factors were removed from the latter analysis resulting in an effective sample size of 1932.

The CHARGE consortium reported 62 SNPs significant at $p<1x10^{-5}$ in their discovery meta-analysis; 61 of these were typed or imputed in at least nine out of thirteen groups. A Z-score meta-analysis was performed of the association of these SNPs in CHARGE and the WMH cohorts. P-values were weighted by sample size consistent with the CHARGE report.

SNPs at locus 17q25

Six SNPs in high linkage disequilibrium at locus 17q25 were examined (Figure 1). All SNPs were typed or imputed to high quality ($r^2 \geq 0.9$) in all cohorts with the exception of MGH-Affymetrix in which rs936393 was typed and only rs11869977 was well imputed (Supplementary Table S3).

Allelic dosage for the most significant SNP was extracted and used in a conditional analysis of SNPs within a 100kb window using ProbABEL[13], followed by meta-analysis within METAL. Also a meta-analysis was performed of the association results of these SNPs in discovery CHARGE report[1], Rotterdam replication report[2] and our WMH cohorts.

Genetic Risk Score

A genetic risk score (GRS) is a composite metric derived from a number of informative genetic variants associated with a phenotype of interest. We used SNPs highly associated with WMHV

($p<1x10^{-5}$) in stroke-free populations[1] to construct scores for ischemic strokes and test association with WMHV. SNPs were divided into 13 unique loci in relative linkage equilibrium ($r^2<0.2$) as determined by SNP Annotation and Proxy Search(www.broadinstitute.org/mpg/snap/). One SNP per locus was chosen based on the largest effect size across WMH cohorts, and the lowest p-value in the original CHARGE report. SNPs had to be typed or well imputed ($r^2>0.8$) in all WMH cohorts. Twelve SNPs were ultimately included in the GRS as rs10012573 was not adequately imputed in three centers and there were no other associated SNPs at this locus (Supplementary Table S3). Individuals with missing genotype(s) at these SNPs were excluded resulting in an effective sample size of 2564. The GRS was calculated by summing the risk allele doses and was not weighted because overall effect sizes are not available from the z-score meta-analysis employed in the CHARGE report.

Statistical Analysis

Statistical analyses were performed in R(http://R-project.org) and $p<0.05$ was considered significant. Within each group log-transformed WMHV normalized for TICV was adjusted for sex and age by obtaining standardized residuals from a linear regression model. The GRS both including, and excluding, the 17q25 locus was assessed as a predictor of adjusted WMHV by linear regression. The analyses were repeated with WMH additionally adjusted for hypertension, hypercholesterolemia, smoking and diabetes. Individuals with missing cardiovascular risk factors were removed from these analyses. The analyses were performed per center and meta-analyzed using inverse-variance method.

B. Association of 17q25 locus with lacunar stroke

Association between the 17q25 SNPs and lacunar stroke status was tested within METASTROKE, a consortium of ischemic stroke case-control GWA-studies.[14] 1854 lacunar strokes and 51939 controls free of symptomatic stroke from twelve cohorts were included (Table 2). Details of genotyping, quality-control, and imputation are available in the online supplement. All cases had brain imaging (CT and/or MRI) except in Atherosclerosis Risk in Communities Study (ARIC, 98% of cases), Australian Stroke Genetics Collaboration (ASGC, 97.4%), Cardiovascular Health Studies (CHS, 94.5%) and Heart and Vascular Health study (HVH, 96.5%).

Lacunar strokes were designated based on compatible clinical and neuroimaging findings using the TOAST criteria or similar.[5] TOAST requires normal neuroimaging or relevant brainstem/subcortical hemispheric lesions (<1.5cm), whereas HVH/CHS required normal CT with a typical lacunar syndrome or subcortical infarction (≤2cm). ARIC defined strokes as lacunar if the infarct was in a typical location (basal ganglia, brainstem, thalamus, internal capsule, or cerebral white matter) and of unstated size or ≤2cm. All cohorts excluded cases with recognized sources of emboli or large vessel atherosclerosis. Samples were not entirely independent of the WMH cohorts, with 16% of cases included in WMHV analysis.

For Genetics of Early Onset Stroke Study, rs3744028 and rs11869977 were not imputed. For Genes Affecting Stroke Risk and Outcome Study only rs936393 and rs1055129 were typed and rs3744017 was poorly imputed (r^2=0.64) and therefore not included in this SNP's meta-analysis. All six SNPs were typed or adequately imputed (r^2≥0.8) in the remaining cohorts (Supplementary Table S5). Association analyses were performed in each center by logistic

regression assuming an additive model. Results were adjusted by genomic control factor, followed by meta-analysis under an inverse-variance weighted model.

RESULTS

A. Association of 17q25 locus with WMHV

All SNPs at the 17q25 locus were significantly associated with WMHV as a quantitative trait in a direction and magnitude of effect consistent with previous reports. Most associations remained significant after Bonferroni adjustment for multiple comparisons ($p<0.008$) except for rs1055129 ($p=0.015$) (Table 3). The correction may be over-conservative given these SNPs are highly correlated ($r^2>0.85$), with the exception of rs1055129, which is in moderate correlated with the others ($r^2\approx0.5$)(Figure 1). The most significant SNP was rs9894383 (B=0.13 SE=0.04 p=0.0006) the risk allele was associated with a 13% (CI 6-22%) increase in the geometric mean of WMHV adjusted for age and sex (Figure 2). There was no evidence of between-study heterogeneity. The results remained significant when adjusting for hypertension alone or cardiovascular risk factors (Supplementary Table S6).

Conditioning on rs9894383 did not reveal any other significant variants within a 100kb window. Meta-analysis of the six 17q25 SNPs with the original CHARGE and Rotterdam replication reports revealed rs9894383 as most significantly associated with WMHV ($p=1.0 \times 10^{-11}$). Meta-analysis of moderately significant SNPs ($p<1\times10^{-5}$) reported by CHARGE only revealed a further genome-wide significant SNP at locus 17q25 (Supplementary Table S7).

The GRS was a significant predictor of WMHV adjusted for age and sex (p=0.001,B=0.031,CI 0.012-0.050) and after additional adjustment for hypertension (p=0.002,B=0.030,CI 0.011-0.049), and combined cardiovascular risk factors (p=0.003,B=0.031, CI 0.011-0.051). A GRS without the 17q25 locus was also significantly associated with WMHV (p=0.023,B=0.022,CI 0.003-0.042) and after adjustment for hypertension (p=0.020,B=0.023,CI 0.004-0.042) and cardiovascular risk factors (p=0.025,B=0.026,CI 0.003-0.048). Figure 3 shows an approximate linear relationship between mean WMHV residuals and quintiles of the GRS with and without 17q25 locus. Supplementary Table S8 gives the mean GRS for each quintile.

B. Association of 17q25 locus with lacunar stroke

There was no association between the six SNPs and lacunar stroke status in the Metastroke case-control analysis (Table 3). The 17q25 SNPs were not associated with cardioembolic or large artery stroke (data not shown).

DISCUSSION

This study provides further support for an association between the 17q25 locus and WMH, and replication in ischemic stroke supports the hypothesis of shared genetic contribution to WMHV in community participants without stroke and ischemic stroke cases. The most significant SNP here was rs9894383 and meta-analysis with two previous reports[1,2] revealed a combined p-value of 1.0×10^{-11}. Most SNPs within the 17q25 region were highly correlated and conditional analysis did not reveal more than one independent signal.

In contrast the 17q25 locus was not associated with another manifestation of SVD, lacunar

stroke. This may suggest that any causal variant linked to this locus does not act by directly promoting small vessel arteriopathy or the type of arteriopathy primarily underlying lacunar infarction. The study was powered to detect an association with an odds ratio between 1.1-1.15. Importantly, however, controls did not have brain imaging to exclude silent brain infarction, which is found in some 20% of healthy elderly[15] and this could have limited study power.

We performed a meta-analysis with the top SNPs reported by CHARGE; however, no novel loci reached genome-wide significance. To determine whether there were additional genetic variants shared by WMH occurring in community-participants without stroke and ischemic stroke, we calculated GRS excluding the 17q25 locus. This was a significant predictor of WMHV consistent with additional shared genetic variants.

There are several genes in the 17q25 region; however, cis-expression quantitative trait loci, primarily of HapMap lymphoblastoid cells, implicate *TRIM47*. TRIM47 could modulate brain responses to ischemic injury as its RING domain confers protein ubiquitination, which promotes proteolysis and cellular homestasis.[16] Imbalance in ubiquitin-proteasome pathways is integral to cerebral ischemic injury mechanisms[17] and also evident in WMH expression profiles.[18]

This study used volumetric MRI techniques, which have been demonstrated to be reliable and accurate[7-9] with good agreement across reading centers. A limitation is the variability in MR-imaging protocols, resulting from the use of clinical imaging in these GWAS databases. However, studies applying volumetric techniques have shown high reproducibility across

acquisition protocols[8] and scanner models.[19] To minimize effects of MRI heterogeneity, centers were analyzed separately and WMHV z-scores were derived prior to association testing. We measured whole brain rather than regional WMHV and therefore could not investigate genetic differences between periventricular and subcortical WMH, which are suggested to have differing pathological, risk factor and functional associations.[4] Another limitation is that genotyping was performed on multiple platforms. However, top SNPs were imputed to a high quality with consistent allele frequencies.

In conclusion, our data provide further support for an association between the 17q25 locus and WMHV in patients with ischemic stroke. Future studies are warranted to explore these genetic associations in order to understand biological underpinnings of this complex cerebrovascular phenotype. The lack of association with lacunar stroke may suggest that the 17q25 locus does not act via promoting small vessel arteriopathy.

Acknowledgments

See http://stroke.ahajournals.org

Funding

The principal funding for this study was provided by the Stroke Association. The authors are supported by MRC (Training Fellowship,PAS) and NINDS (K23NS064052,NSR).
Funding for collection, genotyping and analysis of stroke samples was provided by Wellcome Trust (WTCCC2), the Intramural Research Program (NIA)(MGH,ISGS), National Institute for Neurological Disorders and Stroke (SWISS, GASROS, ISGS, CHS, HVH, GEOS), Bugher

Foundation of the American Heart Association, MGH Deane Institute for Integrative Study of Atrial Fibrillation and Stroke(GASROS), National Institutes of Health Genes, Environment and Health Initiative, Medical Research Service of the Department of Veterans Affairs and Centers for Disease Control (GEOS), National Health & Medical Research Council (ASGC), Italian Ministry of Health (Milan), National Human Genome Research Institute(GASROS, ARIC), National Heart, Lung, and Blood Institute (ARIC, CHS, HVH), Henry Smith Charity and the British Council (BRAINS).

Disclosures: None

References

[1]Fornage M, Debette S, Bis JC, Schmidt H, Ikram MA, Dufouil C, et al. Genome-wide association studies of cerebral white matter lesion burden: the CHARGE consortium. *Ann Neurol.* 2011;69:928-39.

[2]Verhaaren BF, de Boer R, Vernooij MW, Rivadeneira F, Uitterlinden AG, Hofman A, et al. Replication study of chr17q25 with cerebral white matter lesion volume. *Stroke***.** 2011;42:3297-3299.

[3]Debette S, Markus HS. The clinical importance of white matter hyperintensities on brain magnetic resonance imaging: systematic review and meta-analysis. *BMJ.* 2010;341:c3666

[4]Schmidt R, Schmidt H, Haybaeck J, Loitfelder M, Weis S, Cavalieri M, et al. Heterogeneity in age-related white matter changes. *Acta Neuropathol*. 2011;122:171-185.

[5]Adams HP, Jr., Bendixen BH, Kappelle LJ, Biller J, Love BB, Gordon DL, Marsh EE. 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.

[6]Rost NS, Rahman RM, Biffi A, Smith EE, Kanakis A, Fitzpatrick K, et al. White matter hyperintensity volume is increased in small vessel stroke subtypes. *Neurology*. 2010;75:1670-1677.

[7]Nandigam RN, Chen YW, Gurol ME, Rosand J, Greenberg SM, Smith EE. Validation of intracranial area as a surrogate measure of intracranial volume when using clinical MRI**.** *J Neuroimaging.* 2007;17:74-77.

[8]Grimaud J, Lai M, Thorpe J, Adeleine P, Wang L, Barker GJ, et al. Quantification of MRI lesion load in multiple sclerosis: a comparison of three computer-assisted techniques. *Magn Reson Imaging*. 1996;14:495–505.

[9]Smith SM, Zhang Y, Jenkinson M, Chen J, Matthews PM, Federico A, De Stefano N. Accurate, robust, and automated longitudinal and cross-sectional brain change analysis. *Neuroimage.* 2002;17:479-489

[10]Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. 2009;5:e1000529.

[11]Aulchenko YS, Ripke S, Isaacs A, van Duijn CM. GenABEL: an R library for genome-wide association analysis. *Bioinformatics.* 2007;23:1294-1296.

[12]Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*. 2010;26:2190-2191

[13]Aulchenko YS, Struchalin MV, van Duijn CM. ProbABEL package for genome-wide association analysis of imputed data. *BMC Bioinformatics*. 2010;11:134-143

[14]Traylor M, Farrall M, Holliday EG, Sudlow C, Hopewell, JC, Cheng Y, et al. Genetic risk factors for ischaemic stroke and its subtypes (the METASTROKE collaboration): a meta-analysis of genome-wide association studies. *Lancet Neurology.* Online 4 Oct 2012.

[15]Vermeer SE, Longstreth WT Jr, Koudstaal PJ. Silent brain infarcts: a systematic review. *Lancet Neurol.* 2007;6:611-619.

[16]Meroni G, Diez-Roux G. TRIM/RBCC, a novel class of 'single protein RING finger' E3 ubiquitin ligases. *Bioessays.* 2005;27:1147–1157

[17]Di Napoli M, McLaughlin B. The proteasome ubiquitin system as a drug target in cerebrovascular disease: The therapeutic potential of proteasome inhibitors. *Curr Opin Investig Drugs.* 2005;6:686–699.

[18]Simpson JE, Hosny O, Wharton SB, Heath PR, Holden H, Fernando MS, et al. Microarray RNA expression analysis of cerebral white matter lesions reveals changes in multiple functional pathways. *Stroke.* 2009;40:369-375

[19]Jovicich J, Czanner S, Han X, Salat D, van der Kouwe A, Quinn B, et al. MRI-derived measurements of human subcortical, ventricular and intracranial brain volumes: Reliability effects of scan sessions, acquisition sequences, data analyses, scanner upgrade, scanner vendors and field strengths. *Neuroimage.* 2009;46:177-179

Figure 1. Linkage Disequilibrium Map of Six SNPs. Numbers indicate r-squared expressed as percentiles. Most SNPs are highly correlated (r^2 >0.85) and rs1055129 is moderately correlated ($r^2 \approx 0.5$).

Figure 2. Forest Plot showing inverse-variance weighted meta-analysis of the association of rs9894383 with adjusted WMH volume.

Figure 3. Mean adjusted WMHV residuals and 95% confidence intervals for quintiles of genetic risk scores. Analyses are shown both including (A), and excluding (B), the 17q25 locus.

Table 1. Clinical characteristics of ischemic stroke cohorts included in WMH analysis.

	Number	Mean(SD) age (years)	Male(%)	Hypertension (%)	Hypercholester-olaemia (%)	Ever Smoker (%)	Diabetes (%)
St George's*	381	70.6(13.5)	241(63.2)	300(78.7)	286(75.1)	263/380(69.2)	78(20.5)
Oxford*	170	67.0(13.7)	99(58.2)	114(67.1)	83(48.8)	91(53.5)	22(12.9)
Edinburgh*	72	68.6(13.9)	38(52.7)	34(47.2)	Unknown	51/68(75.0)	5(6.9)
Munich*	756	66.6(12.3)	467(61.7)	525(69.4)	346(45.7)	273(36.1)	166(21.9)
Milan	153	57.5(14.3)	92(60.1)	87(56.8)	94(61.4)	64(41.8)	21(13.7)
ISGS	209	67.9(13.8)	129(61.7)	127(60.7)	49/128(38.2)	51(24.4)	8/26(30.7)
ASGC	104	64.8(13.3)	59(56.7)	80(76.9)	52(50.0)	27/100(27.0)	18(17.3)
MGH	975	65.7(14.2)	606(62.2)	618(63.4)	408(41.8)	606(62.2)	199(20.4)
SWISS	115	66.3(11.4)	56(48.7)	85(73.9)	Unknown	Unknown	Unknown
Total	2935	66.4(14.3)	1734(60.9)	1911(65.1)	1318/2667(49.4)	1426/2811(50.7)	535/2628(20.4)

* Part of the Wellcome Trust Case Control Consortium 2 (WTCCC-2), ISGS=Ischemic Stroke Genetics Study, ASGC=Australian Stroke Genetics Collaborative, MGH=Massachusetts General Hospital, SWISS=Siblings with Ischemic Stroke Study.

Table 2. Demographics for lacunar stroke cases and stroke-free controls and effective sample size in METASTROKE

		Cases			Controls		
Continent	Study	n	Mean Age(SD)	% Males	n	Mean Age(SD)	% Male
Europe	WTCCC2-UK	474	75.4(12.5)	52.3%	5175	~52*	49.5%
	WTCCC2-Munich	106	65.1(12.9)	72.6%	797	62.7(10.9)	51.4%
	BRAINS	97	73.9(15.4)	52.5 %	407	≥65†	35.8%
	deCode	240	68.8(10.2)	56.5%	26970	57.3(21.4)	38.0%
	Milan	25	56.2(17.3)	56.7%	407	50.8(8.1)	87.7%
North	GEOS	54	44.3(4.1)	72.2%	498	39.5(6.7)	56.6%
America	GASROS	38	65.7(14.2)	60.3%	1202	47.5(8.5)	59.1%
	HVH	173	67.6(9.3)	32.9 %	1290	66.6(9.1)	47.7%
	ARIC	63	55.3(6.2)	58.7%	8803	54.1(5.7)	46.5%
	CHS	73	74.3(6.1)	34.2%	2817	85.8(5.6)	45.0%
	ISGS/SWISS	201	64.6(13.6)	60.3%	2329	64.8(12.6)	48.0%
Australia	ASGC	310	77.5(13.1)	57.4%	1244	70.2(12.1)	50.2%
Total		1854	70.2(14.0)	54.2%	51939	N/A	43.2%

CHS=Cardiovascular Health Study, ARIC=Atherosclerosis Risk in Communities, BRAIN=Bio-repository of DNA in stroke, GEOS=Genetics of Early Onset Stroke Study, GASROS=Genes Affecting Stroke Risk and Outcome Study, HVH=Heart and Vascular Health, ISGS= Ischemic Stroke Genetics Study.*Approximate age at genotyping of the 2738 controls from the 1958 Birth Cohort. Age was not available for the remaining controls.†All controls were aged 65 years or older at the time of genotyping. No further information was available

Table 3. Association statistics of SNPs at locus 17q25 with WMH volume and lacunar stroke

		WMH volume		Lacunar Stroke	
SNP	Risk Allele	Effect Size (SE)	P-value	Effect Size (SE)	P-value
rs3744028	C	0.12(0.04)	0.0030	-0.005(0.048)	0.92
rs9894383	G	0.13(0.04)	0.00064	0.005(0.046)	0.91
rs11869977	G	0.12(0.04)	0.00069	-0.001(0.047)	0.98
rs936393	G	0.11(0.04)	0.0012	0.014(0.045)	0.77
rs3744017	A	0.12(0.04)	0.0032	-0.001(0.046)	0.98
rs1055129	G	0.08(0.03)	0.015	0.031(0.039)	0.43

ONLINE SUPPLEMENT

The 17q25 locus is associated with white matter hyperintensity lesion volume in patients with ischemic stroke, but not with lacunar stroke.

The Supplementary Material has the following sections:

1. Supplementary Tables 1-8
2. Study Populations
3. Genotyping, quality controls, and imputation
4. Expression Quantitative Trait Loci
5. Funding
6. References
7. Acknowledgements

1. SUPPLEMENTARY TABLES

S1. Stroke Subtype (according to TOAST criteria) of ischemic stroke cases included in WMH analysis.

	Number	SVD	LAA	CE	Other determined	Tandem	Unknown
St George's*	381	108 (28.3%)	87 (22.8%)	81 (21.3%)	0	34 (9%)	71 (19%)
Oxford*	170	24 (14.1%)	25 (14.7%)	8 (4.7%)	8 (4.7%)	7 (4%)	98 (58%)
Edinburgh*	72	16 (22.2%)	7 (9.7%)	5 (6.9%)	0	11 (15%)	33 (45%)
Munich*	756	63 (8.3%)	223(29.5%)	179 (23.6%)	31 (41.0%)	11 (1%)	249 (33%)
Milan	153	9 (5.9%)	35 (22.9%)	29 (19.0%)	13 (8.5%)	9 (6%)	58 (38%)
ISGS	209	27(12.9%)	38(18.2%)	57(27.3%)	8(3.8%)	N/A	79
ASGC	104	5(4.8%)	20(19.2%)	46(44.2%)	0	N/A	27
MGH	975	100(10.2%)	189(19.4%)	312(32.0%)	189(19.4%)	N/A	47
SWISS	115	N/A	N/A	N/A	N/A	N/A	N/A
Total	2943	352/2820 (12.5%)	624/2820 (22.1%)	717/2820 (25.4%)	249/2820 (8.8%)	72/1532 (4.7%)	662/2820 (23.5%)

* Part of the Wellcome Trust Case Control Consortium 2, ISGS=Ischemic Stroke Genetics Study, ASGC=Australian Stroke Genetics Collaborative, MGH= Massachusetts General Hospital, SWISS= Siblings with Ischemic Stroke Study. SVD= small vessel disease, LAA= large artery atherosclerotic, CE= cardioembolic, N/A= not applicable.

S2. Post-quality control numbers, magnetic resonance imaging (MRI) sequences and MRI models for cohorts in WMH GWAS. Note sub-grouping based on availability of MRI sequences.

	Number	WMH Sequence	TICV/ICA Sequence	MRI scanner
St George's	307	Axial FLAIR	Axial T2	1.5T Philips, 1.5T GE Signa LX
Oxford FLAIR	78	Coronal FLAIR	Axial T2	1.5T GE Medical Signa, 1.5T Philips
Oxford T2	66	Axial T2	Axial T2	1.5T GE Medical Signa, 1.5T Philips
Edinburgh	67	Axial FLAIR	Axial FLAIR	1.5T GE Medical Signa, 1.5T Siemens
Munich FLAIR	468	Axial FLAIR	Axial FLAIR	1.5T Siemens Magnetom, 1T Siemens, 1.5T GE Medical Signa
Munich T2	215	Axial T2	Axial T2	1.5T Siemens Magnetom, 3T & 1.5T GE Medical Signa, 1T Siemens
Milan	152	Coronal FLAIR (n=146) Axial FLAIR (n=7)	Axial T2	1.5T Siemens, 0.5T Philips.
ISGS	201	Axial FLAIR	Sagittal T1	1.5T GE Medical Signa
ASGC	95	Axial FLAIR	Sagittal T1	1.5T Siemens Magnetom Avanto
MGH Affymetrix	502	Axial FLAIR	Sagittal T1	1.5T GE Medical Signa
MGH Illumina	242	Axial FLAIR	Sagittal T1	1.5T GE Medical Signa
MGH Omni	84	AXIAL FLAIR	Sagittal T1	1.5T GE Medical Signa
SWISS	111	Axial FLAIR	Sagittal T1	1.5T GE Medical Signa

* WMH= white matter hyperintensity, ICA= intracranial area, TICV= total intracranial volume

S3. Imputation Quality (r-squared) and minor allele frequency (MAF) of 17q25 single nucleotide polymorphisms (SNPs) in ischemic stroke cohorts included WMH analysis.

	rs3744028		rs11869977		rs9894383		rs936393		rs3744017		rs1055129	
	R^2	MAF	R^2	MAF	R^2	MAF	R^2	MAF	R^2	MAF	R^2	MAF
MGH Affymetrix	0.65	0.09	0.92	0.20	0.68	0.22	typed	0.21	0.61	0.22	0.27	0.55
MGH Illumina	0.93	0.17	0.96	0.19	0.97	0.19	0.97	0.18	typed	0.18	0.96	0.29
MGH Omni	0.99	0.19	typed	0.19	typed	0.19	typed	0.19	typed	0.19	0.98	0.30
ASGC	0.94	0.14	0.96	0.15	0.96	0.16	0.97	0.15	typed	0.14	0.96	0.31
ISGS/SWISS	0.91	0.19	0.94	0.20	0.94	0.21	0.93	0.20	typed	0.20	0.89	0.30
Milan	0.92	0.21	0.94	0.22	0.97	0.22	0.95	0.21	typed	0.21	0.96	0.33
WTCCC2-D	0.94	0.19	0.96	0.20	0.98	0.20	0.98	0.20	typed	0.20	0.96	0.31
WTCCC2-UK	0.93	0.16	0.96	0.18	0.98	0.18	0.97	0.18	typed	0.17	0.96	0.29

WTCCC2=Wellcome Trust Case Control Consortium 2, -D= Munich, -UK= St George's, Edinburgh, and Oxford

S4. Imputation Quality (r-squared) and case minor allele frequency (MAF) of SNPs in Genetic Risk Score in ischemic stroke cohorts included WMH analysis.

Chr	SNP	MGH Omni		MGH Affymetrix		MGH Illumina		ISGS & SWISS		ASGC		Milan		WTCCC2-Germany		WTCU
		R2	MAF	R2	MAF	R2	MAF	R2	MAF	R2	MAF	R2	MAF	R2	MAF	R2
1	rs2842873	0.99	0.36	0.98	0.36	0.98	0.33	0.98	0.35	0.98	0.37	0.95	0.29	0.99	0.33	0.99
1	rs12073947	0.99	0.32	1.00	0.33	0.98	0.35	0.98	0.31	0.97	0.31	0.97	0.37	0.98	0.31	0.98
3	rs2167089	1.00	0.32	0.89	0.31	0.90	0.31	0.93	0.30	0.88	0.28	0.92	0.31	0.94	0.31	0.93
4	rs11731436	0.99	0.36	1.00	0.35	1.00	0.35	1.00	0.34	1.00	0.35	1.00	0.34	1.00	0.35	1.00
5	rs16901064	typed	0.16	0.83	0.16	typed	0.16	typed	0.16	typed	0.16	typed	0.20	typed	0.16	typed
7	rs6945846	0.93	0.23	0.72	0.21	0.84	0.19	0.84	0.21	0.84	0.25	0.81	0.14	0.83	0.19	0.86
8	rs6992136	0.99	0.20	0.99	0.20	0.96	0.16	0.98	0.17	0.99	0.15	0.96	0.21	0.97	0.20	0.96
9	rs10814323	typed	0.24	1.00	0.23	typed	0.19	typed	0.23	typed	0.28	typed	0.17	typed	0.20	typed
10	rs17724534	0.99	0.13	0.78	0.14	0.79	0.21	0.84	0.13	0.81	0.12	0.83	0.09	0.83	0.10	0.84
10	rs11191612	0.95	0.38	0.91	0.40	0.91	0.35	0.91	0.37	0.89	0.37	0.98	0.25	0.99	0.23	0.99
14	rs11625623	0.99	0.25	0.97	0.23	0.99	0.23	0.98	0.26	0.93	0.21	0.98	0.25	0.99	0.23	0.99
17	rs11869977	1.00	0.19	0.92	0.20	0.96	0.19	0.94	0.20	0.96	0.15	0.94	0.22	0.96	0.20	0.96

Chr = Chromosome, SNP= single nucleotide polymorphism

S5. Imputation Quality (r-squared) and minor allele frequency (MAF) of 17q25 SNPs in twelve SVD stroke-subtype case-control analyses from Metastroke

Study	rs3744028		rs11869977		rs9894383		rs936393		rs3744017		rs1055129	
	R^2	MAF	R^2	MAF	R^2	MAF	R^2	MAF	R^2	MAF	R^2	MAF
WTCCC2-UK	0.99	0.18	0.97	0.19	0.97	0.19	0.97	0.19	typed	0.18	0.98	0.30
WTCCC2-D	0.99	0.19	0.98	0.19	0.98	0.19	0.98	0.19	typed	0.19	0.97	0.30
BRAIN	0.95	0.16	0.93	0.18	0.96	0.18	0.90	0.17	typed	0.17	0.92	0.25
DeCode	0.99	0.18	0.98	0.19	0.99	0.18	0.98	0.19	typed	0.18	0.97	0.31
Milan	0.97	0.23	0.95	0.25	0.95	0.25	0.95	0.25	typed	0.23	0.96	0.35
GEOS	N/A	N/A	N/A	N/A	typed	0.18	typed	0.17	typed	0.17	N/A	N/A
GASROS	N/A	N/A	N/A	N/A	N/A	N/A	typed	0.15	0.64*	0.06	typed	0.26
HVH	0.92	0.18	0.93	0.21	0.95	0.20	0.93	0.21	typed	0.19	0.93	0.31
ARIC	0.82	0.20	0.79	0.21	0.79	0.21	0.79	0.21	0.79	0.20	0.86	0.32
CHS	0.88	0.18	0.91	0.21	0.92	0.21	0.91	0.21	typed	0.19	0.89	0.31
ISGS/SWISS	0.84	0.19	0.95	0.20	0.95	0.20	0.92	0.20	typed	0.20	0.87	0.28
ASGC	0.98	0.18	0.97	0.18	0.98	0.18	0.97	0.18	typed	0.18	0.96	0.29

*excluded from analyses due to poor imputation and inconsistent MAF. WTCCC2= Wellcome Trust Case Control Consortium 2, BRAIN= Bio-repository of DNA in stroke, GEOS= Genetics of Early Onset Stroke Study, GASROS= Genes Affecting Stroke Risk and Outcome Study, HVH= Heart and Vascular Health, ARIC= Atherosclerosis Risk in Communities, CHS= Cardiovascular Health Study, ISGS = Ischemic Stroke Genetics Study, SWISS= Siblings with Ischemic Stroke Study ASGC=Australian Stroke Genetics Collaborative.

S6. Association results for SNPs at locus 17q25 with WMHV in Ischaemic Stroke.

SNP	Model 1		Model 2		Model 3	
	Effect (SE)	P-value	Effect (SE)	P-value	Effect (SE)	P-value
rs3744028	0.12 (0.04)	3.0×10^{-3}	0.12 (0.04)	4.2×10^{-3}	0.10 (0.05)	3.2×10^{-2}
rs9894383	0.13 (0.04)	6.4×10^{-4}	0.13 (0.04)	8.6×10^{-4}	0.10 (0.04)	1.4×10^{-2}
rs11869977	0.12 (0.04)	6.9×10^{-4}	0.13 (0.04)	8.7×10^{-4}	0.10 (0.04)	1.2×10^{-2}
rs936393	0.11 (0.04)	1.2×10^{-3}	0.13 (0.04)	6.2×10^{-4}	0.10 (0.04)	7.5×10^{-3}
rs3744017	0.12 (0.04)	3.2×10^{-3}	0.14 (0.04)	5.1×10^{-4}	0.11 (0.04)	9.3×10^{-3}
rs1055129	0.08 (0.03)	1.5×10^{-2}	0.08 (0.03)	1.6×10^{-2}	0.07 (0.04)	5.0×10^{-2}

Model 1 is WMH adjusted for age, gender only. Model 2 is WMH adjusted for age, gender and hypertension. Model 3 is adjusted for age, gender, hypertension, ever-smoker, diabetes and hypercholesterolemia. All models adjusted for first two ancestry principle components.

S7. Z-score based meta-analysis with CHARGE SNPs with p<1 x 10^{-5}.

Chr	Genes	SNP	Allele	CHARGE *P*	Study *P*	Meta-Analysis *P*
1	*PMF1;BGLAP;*	rs1052053	a	5.00E-06	0.16	6.05E-05
	SLC24A44	rs2842873	c	6.40E-06	0.027	7.51E-06
		rs2758605	g	9.40E-06	0.027	9.79E-05
1	*RP11-518D3.1*	rs1892525	g	7.20E-07	0.23	5.79E-05
		rs7521244	g	1.20E-06	0.22	6.68E-05
		rs10789247	a	1.20E-06	0.21	6.21E-05
		rs10889722	g	1.20E-06	0.21	6.28E-05
		rs11590313	g	1.60E-06	0.21	7.37E-05
		rs7521135	g	1.60E-06	0.20	6.25E-05
		rs7520899	c	1.70E-06	0.19	6.34E-05
		rs12025677	t	1.90E-06	0.18	6.09E-05
		rs11209171	a	2.90E-06	0.26	1.42E-04
		rs12073947	a	3.80E-06	0.08	2.33E-05
		rs1317272	a	3.80E-06	0.10	3.20E-05
3	*AC098970.2*	rs2167089	g	6.00E-06	0.38	6.38E-04
4	*AC097110.1*	rs11731436	c	3.30E-06	0.16	8.48E-02
4	*COL25A1*	rs10012573	a	6.00E-06	0.90	4.82E-03
5	*RNASEN*	rs16901064	c	7.80E-06	0.52	1.01E-03
		rs12189086	a	8.20E-06	0.49	9.07E-04
		rs7714912	c	9.60E-06	0.52	1.56E-03
7	*FOXP2*	rs6945846	c	7.90E-06	0.13	1.03E-04
8	*AC018437.10*	rs6992136	g	3.20E-06	0.87	6.70E-03
		rs9325770	c	3.90E-06	0.75	1.02E-02
		rs9650356	g	5.40E-06	0.86	3.56E-03
		rs6996022	g	5.50E-06	0.74	2.34E-03
		rs2720623	c	6.30E-06	0.93	7.11E-03
		rs17122137	g	6.50E-06	0.84	9.61E-03
		rs7826382	a	6.50E-06	0.91	7.79E-03
		rs9650357	c	6.50E-06	0.87	3.90E-03
		rs1968723	t	6.80E-06	0.88	8.67E-03
9	*C9orf62*	rs9410016	g	9.70E-06	0.57	1.65E-03

9		rs10814323	a	1.70E-06	0.33	1.42E-04
10	*RP11-753C18.7; CNNM2*	rs17724534	c	8.20E-06	1.00	7.51E-03
10	*NT5C2; RP11-332O19.4;PCGF6; PDCD11;CALHM1; RP11-225H22.4*	rs1163238	g	4.80E-06	0.45	8.27E-04
		rs11191612	a	8.10E-06	0.34	5.70E-04
		rs11598702	t	5.60E-06	0.27	2.51E-04
		rs7894407	t	6.10E-07	0.69	9.28E-04
14	*PTGDR*	rs11625623	g	7.70E-06	0.90	9.85E-03
	MTHFD1;C14orf50	rs11629135	g	8.60E-06	0.31	3.44E-04
17	*UNC13D; WBP2; TRIM47; TRIM65; MRPL38; FBF1*	rs3744028	c	4.00E-09	4.42E-03	6.39E-09
		rs9894383	g	5.30E-09	1.01E-03	8.71E-10
		rs11869977	g	5.70E-09	1.15E-03	1.10E-09
		rs936393	g	6.80E-09	2.25E-03	4.81E-09
		rs3744017	a	7.30E-09	6.30E-03	8.49E-09
		rs1055129	g	4.10E-08	2.45E-02	3.58E-08
		rs1463485	g	8.30E-08	1.41E-03	7.94E-09
		rs1551619	t	2.00E-07	2.07E-02	5.53E-07
		rs3785437*	t	2.90E-07	4.73E-03	***6.15E-08***
		rs2290771	g	4.50E-07	1.99E-02	1.02E-06
		rs3760128	g	8.00E-07	3.33E-02	3.13E-06
		rs9895947	c	2.80E-06	4.48E-02	1.01E-05
		rs9302994	a	3.00E-06	5.47E-02	1.44E-05
		rs1046446	t	4.20E-06	4.69E-02	1.37E-05
		rs8067076	c	5.20E-06	0.20	1.20E-04
		rs7221792	g	5.50E-06	0.21	1.99E-04
		rs1135889	a	5.60E-06	4.52E-02	1.52E-05
		rs7217432	g	5.80E-06	0.35	5.23E-04
		rs17581728	t	6.10E-06	0.10	4.97E-05
		rs2290769	c	8.10E-06	0.22	2.22E-04
		rs7223416	g	5.40E-06	0.22	2.04E-04
		rs2608882	c	7.30E-06	6.26E-02	2.93E-05

* novel genome-wide significant SNP ($p<1x10^{-7}$)

S8. Table showing the mean GRS and standard deviation per quintile used in Figure 3.

Quintiles	Number	Mean GRS (SD) with 17q25 locus	Mean GRS (SD) without 17q25 locus
0-20%	513	10.68 (1.18)	10.33 (1.05)
20-40%	513	12.67 (0.70)	12.25 (0.37)
40-60%	514	13.82 (0.70)	13.42 (0.40)
60-80%	513	14.87 (0.65)	14.51 (0.41)
80-100%	511	16.60 (1.07)	16.21 (0.91)

2. STUDY POPULATIONS

The Atherosclerosis Risk in Communities Study (ARIC)

The ARIC study is a prospective population-based study of atherosclerosis and clinical atherosclerotic diseases in 15,792 men and women, including 11,478 non-Hispanic white participants, drawn from 4 U.S. communities (Suburban Minneapolis, Minnesota; Washington County, Maryland; Forsyth County, North Carolina, and Jackson, Mississippi). Only self-identified whites were included in the analyses.[1]

Hospitalized strokes that occurred by December 31, 2007 (median follow-up, 18.7 years) were included in the analyses. This was determined by annual telephone interviews and obtaining hospital record for any hospitalizations. Moreover, all local hospitals annually provided lists of stroke discharges, which were surveyed for ARIC participant discharges. Details on quality assurance for ascertainment and classification of stroke are described elsewhere [2]. A stroke was classified as ischemic when a brain CT or MRI revealed acute infarction without hemorrhage. All definite ischemic strokes were further classified as lacunar, non-lacunar thrombotic, or cardioembolic on the basis of the recorded neuroimaging results. A stroke was classified as "lacunar" when 2 criteria were met: (1) typical location of the infarct (basal ganglia, brain stem, thalamus, internal capsule, or cerebral white matter); and (2) infarct size of $\leq$2 cm or unstated size.[3] Furthermore cases with a recognized source of emboli were excluded from this subtype.

Australian Stroke Genetics Collaborative (ASGC)

ASGC stroke cases comprised European-ancestry stroke patients admitted to four clinical centers across Australia (The Neurosciences Department at Gosford Hospital, Gosford, New South Wales (NSW); the Neurology Department at John Hunter Hospital, Newcastle, NSW; The Queen Elizabeth Hospital, Adelaide ; and the Royal Perth Hospital, Perth) between 2003 and 2008. Stroke was defined by WHO criteria as a sudden focal neurologic deficit of vascular origin,

lasting more than 24 hours and confirmed by imaging such as computerized tomography (CT) and/or magnetic resonance imaging (MRI) brain scan. Other investigative tests such as electrocardiogram, carotid Doppler and trans-esophageal echocardiogram were conducted to define stroke etiology as clinically appropriate. Cases were excluded if aged <18 years, diagnosed with haemorrhagic stroke or transient ischemic attack, or were unable to undergo baseline brain imaging. A total of 1230 ischemic stroke cases were included and subtypes were assigned using Trial of Org 10172 in Acute Stroke Treatment (TOAST) classification system.[4]

ASGC controls were participants in the Hunter Community Study (HCS), a population-based cohort of individuals aged 55-85 years, predominantly of European Caucasian ancestry.[5] Briefly, participants were randomly selected from the NSW State electoral roll and contacted by mail between 2004 and 2007. Consenting participants completed detailed self-report questionnaires and attended the data collection centre, at which time a series of clinical measures were obtained. A total of 1280 HCS participants were genotyped for the current study.

Bio-Repository of DNA in Stroke (BRAINS)

The Bio-Repository of DNA in Stroke (BRAINS) is an international study recruiting highly phenotyped patients with stroke.[6] For the purposes of the current work all patients were Caucasians. Diagnosis of stroke was confirmed using positive imaging (MRI or CT) and ischemic stroke subtypes were assigned using TOAST criteria, based on clinical, imaging and risk factor data. Controls were European-Ancestry, stroke-free participants from the shared WTCCC controls, a prospectively collected cohort of individuals born in 1958 (1958 Birth Cohort) (http://www.b58cgene.sgul.ac.uk/).

Cardiovascular Health Study (CHS)

The CHS is a population-based cohort study of risk factors for coronary heart disease (CHD) and stroke in adults ≥65 years conducted across four field centers in the United States. The original predominantly European Ancestry (self-reported as “white”) cohort of 5,201 persons (4,964 whites) was recruited in 1989-1990 from a random sample of people on Medicare eligibility lists.[7] The discovery cohort was limited to those of self-reported European ancestry to reduce the possibility of confounding by population stratification. Participants were examined annually from enrollment to 1999, and continue to be contacted twice a year to identify potential cardiovascular events, including stroke. In addition, all hospitalizations were screened for potential stroke events and, when available, CT and/or MRI scans or reports were reviewed centrally. Final adjudication of the occurrence of stroke, stroke types, and subtypes was undertaken by vascular neurologists at a consensus conference using all available information. Strokes were classified as ischemic if there was imaging (CT or MRI within 4 weeks), surgical or autopsy evidence excluding a hemorrhage.

deCode Genetics

Cases, irrespective of age, were identified from a registry of individuals diagnosed with ischemic stroke or TIA at Landspitali University Hospital in Reykjavik, the only tertiary referral centre in Iceland, during the years 1993 to 2006.[8] The ischemic stroke or TIA diagnoses were based on standard WHO criteria and imaging evidence (either CT or MRI), and were clinically confirmed by neurologists. Eligible patients who survived the stroke were invited to participate the genetic study, either by attending a recruitment centre for deCODE's genetic studies, or they were visited at their home by a study nurse. Control subjects were participants from a large variety of genetic programs at deCODE. Individuals with confirmed stroke (identified by cross-matching with hospital lists), who had participated in genetic studies other than those of cardiovascular diseases (CVD) (but not participated in CVD studies) were excluded as controls.

The Genetics of Early Onset Stroke (GEOS) Study, Baltimore, USA

GEOS is a population-based case-control study designed to identify genes associated with early-onset stroke in patients with first-ever ischemic stroke aged 15-49 years from the greater Baltimore-Washington area between 1992 and 2008. Only patients of European descent are included in this meta-analysis. Cases were identified through discharge surveillance from 59 participating hospitals and direct physician referral from a defined geographic region. Abstracted medical records were reviewed and adjudicated for ischemic stroke subtype by two neurologists, with discrepancies resolved by a third neurologist. Stroke was defined according to the WHO criteria and ischemic stroke was defined based on the criteria of the NINDS Stroke Data bank.[9] Cases had a head CT and/or brain MRI that was consistent with cerebral infarction. Visualization of the infarct was not required, only that no alternative etiology was identified. Ischemic stroke subtypes were assigned using TOAST criteria, based on clinical and imaging findings. Controls with no history of ischemic stroke were identified through random digit dialing and were frequency-matched to cases based on sex, age, geographic location and, during the later study periods, ethnicity.

The Heart and Vascular Health Study (HVH)

The setting for this study was Group Health (GH), a large integrated health care system in western Washington State. Data were utilized from an ongoing case-control study of incident myocardial infarction (MI) and stroke cases with a shared common control group. Methods for the study have been described previously[10-11] and are briefly summarized below. All study participants were GH members and aged 30-79 years. MI and stroke cases were identified from hospital discharge diagnosis codes and were validated by medical record review. Controls were a random sample of GH members frequency matched to MI cases on age (within decade), sex, treated hypertension, and calendar year of identification. The index date for controls was a computer-generated random date within the calendar year for which they had been selected. For stroke cases, the index date was the date of admission for the first acute stroke. Participants were excluded if they were recent enrollees at GH, had a history of prior stroke, or if the incident event was a complication of a procedure or surgery.

Trained medical record abstractors collected eligibility and risk factor information from a review of the GH medical record using only data available prior to the index date and through a telephone interview. Medication use was ascertained using computerized GH pharmacy records. A venous blood sample was collected from all consenting subjects, and DNA was extracted from white blood cells using standard procedures. Diagnostic criteria for ischemic stroke were adopted from the Cardiovascular Health Study. Ischemic stroke cases satisfied one or more of the following criteria: (a) Focal deficit, without evidence of blood on CT or MRI, (b) Focal deficit, with mottled appearance in the appropriate location on CT, or (c) surgery or autopsy evidence of infarction. Subtyping of lacunar stroke ("SVD") required either: (a) CT/MRI demonstrates a deep area of infarction (decreased density) less than 2 cm. across, or (b) A normal CT, but the clinical syndrome is typical of a lacunar infarction, that is: a pure motor stroke, a pure sensory stroke, hemiparesis plus ataxia, or dysarthria plus a clumsy hand (c) Exclude cases with a recognized source of emboli or large vessel atherosclerosis.

Siblings With Ischemic Stroke Study (SWISS)

Siblings with Ischemic Stroke Study (SWISS) was a prospective, multicenter affected sibling pair study of first-ever or recurrent ischemic stroke.[12] Probands were recruited from 70 clinical centers across the US and Canada. Ischemic stroke affected and unaffected siblings were recruited primarily using proband-initiated contact. All affected individuals had WHO-defined stroke confirmed by a study neurologist to be ischemic on the basis of head CT or brain MRI. All confirmed cases of ischemic stroke were further classified by a study neurologist using the TOAST criteria.[4] Peripheral blood DNA samples were collected between October 2000 and December 2009.

The Ischemic Stroke Genetics Study (ISGS)

Ischemic Stroke Genetics Study (ISGS) was a 5-center, prospective, case-control study of first-ever ischemic stroke cases and concurrently enrolled controls individually matched for age, sex and recruitment site.[13] All affected individuals had WHO-defined stroke confirmed by a study neurologist to be ischemic on the basis of head CT or brain MRI. All confirmed cases of ischemic stroke were further classified by a centralized phenotype committee consisting of study neurologists using the TOAST as well as other standardized systems. Peripheral blood DNA samples were collected between May 2003 and September 2008.

Genes Affecting Stroke Risk and Outcome Study (GASROS)

Cases were all consecutive patients aged ≥ 18 years presenting with ischemic stroke and admitted to the Massachusetts General Hospital (MGH) Stroke Unit through the Emergency Department, or evaluated in the MGH Neurology outpatient clinics, as well as on the inpatient Medical and Vascular Surgical services from January 2003 to July 2008.[14] Only patients of European ancestry (confirmed by principal component analysis using genome-wide SNP data) were included in the present analysis. Ischemic stroke was defined as either (1) a

radiographically proven (head CT or MRI) infarct associated with the appropriate clinical stroke syndrome, or (2) a fixed neurological deficit persisting more than 24 hours, consistent with a vascular pattern of involvement and without radiographic evidence of demyelinating disease, or other non-vascular structural disease. All subjects were evaluated by a neurologist upon presentation and clinical and laboratory data were collected during the admission for qualifying ischemic stroke event. All patients had brain imaging (CT and/or MRI) as well as ancillary diagnostic investigations where clinically relevant to qualify ischemic stroke event. Diagnostic work-up included: head CT (100%), brain MRI (90%), cervical and intracranial vessel imaging using CTA or MRA (75%), carotid and/or transcranial ultrasound (24%), echocardiography (86%), and Holter monitoring (16%). Controls were recruited among the stroke-free adults presenting to the MGH outpatient clinics and matched with the stroke cases on the basis of age, sex and ancestry information obtained from principal component analysis of GWAS data.

Milan

This study includes consecutive Italian patients referred to Besta Institute from 2000 to 2009 with stroke and included in the Besta Cerebrovascular Diseases Registry (CEDIR).[15] Ischemic stroke cases, first ever or recurrent, confirmed on brain imaging, were selected for this study. All cases were of self reported Caucasian ancestry and had clinically relevant diagnostic workup performed. All cases were phenotyped by an experienced stroke neurologist according to TOAST criteria, based on relevant clinical imaging and available information on cardiovascular risk factors. Controls are Italian individuals enrolled within the PROCARDIS Study, with no personal or sibling history of coronary heart disease before age 66 years.

Wellcome Trust Case-Control Consortium 2 (WTCCC2)

The WTCCC2 samples were genotyped as part of the WTCCC 2 ischemic stroke study. Stroke cases included samples recruited by investigators at from three centers in the UK (St. George's Oxford and Edinburgh) and one centre in Germany, University and Klinikum Großhadern, Ludwig-Maximilians-University, Munich.[15] The St George's Stroke Study consecutively recruited ischemic stroke patients attending cerebrovascular services between 1995 and 2008 (n=1224). The Oxford Vascular Study recruited patients with acute ischaemic stroke or transient ischaemic attack (TIA) with evidence of infarction on brain imaging between 2002 and 2008 as part of a population-based study (n=896). The Edinburgh Stroke Study prospectively recruited consecutive stroke inpatients and outpatients between 2002 and 2005. The Munich study recruited consecutively between 2002 and 2008, from a single Stroke Unit with a high rate of MR imaging (>80%) (n=1383). All subjects were over 18 years of age, of self-reported European ancestry and with a diagnosis of ischaemic stroke classified according to TOAST by an experienced neurologist or stroke physician. All patients had brain imaging (CT and/or MRI) as well as ancillary diagnostic investigations where clinically relevant. All cases were of self reported Caucasian ancestry.

Controls for the UK samples were drawn from shared WTCCC controls obtained from the 1958 Birth Cohort. This is a prospectively collected cohort of individuals born in 1958 (http://www.b58cgene.sgul.ac.uk/), and ascertained as part of the national child development study (http://www.cls.ioe.ac.uk/). Data from this cohort are available as a common control set for a number of genetic and epidemiological studies. For the German samples controls were Caucasians of German origin participating into the population KORAgen study (http://epi.gsf.de/kora-gen/index_e.php). This survey represents a gender- and age stratified random sample of all German residents of the Augsburg area and consists of individuals 25 to 74 years of age, with about 300 subjects for each 10-year increment. All controls were free of a history of stroke or transient ischemic attack.

3. GENOTYPING, IMPUTATION & QUALITY CONTROL IN METASTROKE COHORTS

The Atherosclerosis Risk in Communities Study (ARIC)

Genotyping was performed with the GeneChip SNP Array 6.0 (Affymetrix). Subject specific quality control filters included filters for call rate, heterozygosity, sex mismatch. SNP specific quality control filters included filters for call rate, minor allele frequency (MAF), Hardy-Weinberg equilibrium (HWE), and differential missingness by outcome or genotype. The set of genotyped input SNPs used for imputation was selected based on high quality GWA data. We used a call-rate >95%, HWE p-value>$5x10^{-6}$; MAF>0.01. A total of 704,588 SNPs passing quality control (QC) criteria were used for imputation, which was performed with the MaCH (http://www.sph.umich.edu/csg/yli/mach/index.html) v1.0.16 software.

Australian Stroke Genetics Collaborative (ASGC)

The ASGC sample was genotyped using the Illumina HumanHap610-Quad array. Quality control excluded SNPs with genotype call rate <0.95, deviation from HWE $P<1\times10^{-6}$ or MAF <0.01. At the individual level, samples were excluded if: (i) genotype call rate <95%; ii) genome-wide heterozygosity < 23.3% or > 27.2%; iii) disconcordant clinical and genotypic gender; iv) an inferred first- or second-degree relative in the sample based on pair-wise allele sharing estimates (estimated genome proportion shared identical by descent (IBD): pi-hat >0.1875). Principal components analysis (PCA) was performed, incorporating genotype data from Phase 3 HapMap populations (CEU, CHB, JPT, TSI, YRI) in order to identify and remove non-European ancestry individuals. Following quality control, 1162 cases and 1244 controls were available for association analyses at 551,514 SNPs.

Genotype imputation in the filtered sample was performed using MACH v1.0.16, based on HapMap Phase 2 (release 24) phased haplotypes for European-ancestry (CEU) samples. Subsequent quality control excluded imputed SNPs with MAF <0.01 or ratio of observed dosage variance to expected binomial variance of r^2 <0.3.

Bio-Repository of DNA in Stroke (BRAINS)

The BRAINS sample was genotyped using the Illumina HumanHap610-Quad array. Quality control excluded SNPs not genotyped on all case and control collections and SNPs with genotype call rate <0.95, deviation from Hardy-Weinberg equilibrium ($P<1\times10^{-6}$) or minor allele frequency <0.01. Individual samples were excluded due to low call rates (<95%), gender discrepancy, unexpected relatedness or evidence of non-European ancestry.

Genotype imputation was performed using MACH v1.0.16 based on HapMap Phase 2 CEU samples (release #22). Quality control removed imputed SNPs with MAF <0.01 or ratio of observed dosage variance to expected binomial variance of r^2 <0.3.

Cardiovascular Health Study (CHS)

In 2007-2008, genotyping was performed at the General Clinical Research Center's Phenotyping/Genotyping Laboratory at Cedars-Sinai using the Illumina 370CNV BeadChip system on 3980 CHS participants who were free of CVD at baseline, consented to genetic testing, and had DNA available for genotyping.

A total of 1908 persons were excluded from the GWAS study sample due to the presence at study baseline of coronary heart disease, congestive heart failure, peripheral vascular disease, valvular heart disease, stroke or transient ischemic attack or lack of available DNA. Because the other cohorts were predominantly of European Ancestry, the African American participants were excluded from this analysis to reduce the possibility of confounding by population structure. Participants were also excluded for sex mismatch (discordance with genotyping) or call rate $\leq$ 95%. To date, genotyping has been successful among 3,271 of 3,373 white participants on whom genotyping was attempted; the latter constitute the CHS sample for this study. In CHS, the following exclusions were applied to identify a final set of 306,655 autosomal SNPs: call rate < 97%, Hardy-Weinberg p-value < 10^{-5}, > 2 duplicate errors or Mendelian inconsistencies (for reference CEPH trios), heterozygote frequency = 0, SNP not found in HapMap. Imputation was performed using BIMBAM v0.99 (http://www.bcm.edu/cnrc/mcmcmc/bimbam/) with reference to HapMap CEU using release 22, build 36 using one round of imputations and the default expectation-maximization warm-ups and runs. For each imputed SNP a reliability of imputation was estimated as the ratio of the observed dosage variance to the expected binomial dosage variance.

deCODE Genetics

The Icelandic chip-genotyped samples were assayed with the Illumina Human Hap300, Hap CNV370, Hap 610, 1M or Omni-1 Quad bead chips at deCODE genetics. SNPs were excluded if (i) yield was lower than 95%, (ii) MAF < 0.01 (iii) significant deviation from HWE in controls p< 0.001, (iv) an excessive inheritance error rate (> 0.001) was produced or (v) there was a substantial difference in allele frequency between chip types (from just a single chip if that

resolved all differences, but from all chips, otherwise). All samples with a call rate below 97% were excluded from the analysis. Imputation was performed using IMPUTE (https://mathgen.stats.ox.ac.uk/impute/impute.html).

The Genetics of Early Onset Stroke (GEOS) Study, Baltimore, USA

Samples were genotyped at the Johns Hopkins Center for Inherited Disease Research (CIDR), and using the Illumina HumanOmni1-Quad_v1-0_B BeadChip. Case and control samples were balanced across the plates. Allele cluster definitions for each SNP were determined using Illumina BeadStudio Genotyping Module version 3.3.7, Gentrain version 1.0 and the combined intensity data from all released samples. Genotypes were not called if the quality threshold (GenCall score) was below 0.15.

All samples had a genotype call rate > 98%. Genotyping concordance rate was 99.996% based on study duplicates. Samples were excluded due to unexpected duplicates, gender discrepancy, unexpected relatedness or evidence of non-European ancestry based on principal components analysis. Individual SNPs were excluded from analysis if they had excessive deviation from HWE in controls $p< 1.0x10^{-6}$, genotype call rates <97.5% or MAF < 0.01.

The Heart and Vascular Health Study (HVH)

Genotyping was performed at Cedars-Sinai using the Illumina 370CNV BeadChip and called using the Illumina BeadStudio software. Samples were excluded from analysis for sex mismatch or call rate < 95%. The following exclusions were applied to identify a final set of 301,321 autosomal SNPs: call rate < 97%, HWE $P < 10^{-5}$, > 2 duplicate errors or Mendelian inconsistencies (for reference CEPH trios), heterozygote frequency = 0, SNP not found in HapMap, inconsistencies across genotyping batches. Imputation was performed using BIMBAM with reference to HapMap CEU using release 22 (build 36). SNPs were excluded from analysis for variance on the allele dosage ≤0.01.

Logistic regression was used to investigate the association of each SNP with the risk of stroke adjusting for the matching factors of age, sex, hypertension status and index year. Linear additive models were used with robust standard errors and estimated risk for each additional copy of the variant allele, using R.

The Ischemic Stroke Genetics Study (ISGS)/ Siblings With Ischemic Stroke Study (SWISS)

SWISS and ISGS cases were genotyped using Illumina 650K Quad arrays at the Department of Molecular Neuroscience and Reta Lilla Weston Laboratories, Institute of Neurology, University College London. Controls utilized in this study are participants of the Baltimore Longitudinal Study of Aging (BLSA). No known first degree relatives with stroke or other neurological disease are included in the controls for the present analysis. Controls were genotyped using

Illumina 550Kv1 or 550Kv3 arrays at the Laboratory of Neurogenetics, National Institute on Aging, NIH (Bethesda, MD).

Genes Affecting Stroke Risk and Outcome Study (GASROS)

Cases and controls were genotyped using the Affy 6.0 array. Quality control procedures excluded SNPs with >5% missingness, minor allele frequency <0.01, or Hardy-Weinberg p-value < 10^{-7}. Individual samples were excluded if they exhibited genotype missingness >5%, cryptic relatedness (one of each pair demonstrating IBD pi_hat > 0.15 was removed), or non-European ancestry based on multi-dimensional scaling analysis using HapMap Phase 3 populations. Analyses were performed using PLINK v 1.6. Imputation was performed using MaCH v 1.0.16 and the HapMap 3 CEU+TSI training set.

Milan

Italian cases were genotyped using Illumina Human610-Quad v1_B or Human660W-Quad v1_A chips. Italian controls were genotyped with the Illumina HumanHap610-Quad chip. PCA with HapMap 3 on the Italian cases showed that Italian PROCARDIS controls had similar ancestry to the cases. All samples had a genotype call rate > 95%. Samples were excluded due to unexpected duplicates or evidence of non-European ancestry based on principal components analysis. Quality control procedures excluded SNPs with MAF <0.01 or Hardy-Weinberg P-value <5×10-6 in either the case or control collections.

Wellcome Trust Case-Control Consortium 2 (WTCCC2)

All WTCCC2 cases were genotyped as part of the WTCCC2 Ischemic Stroke study using the Illumina Human660W-Quad array. British controls were genotyped using the Illumina Human1.2M-Duo. German controls were genotyped on the Illumina Human 550k platform. Quality control procedures in the WTCCC2 excluded SNPs not genotyped on all case and control collections and SNPs with Fisher information measure <0.98, genotype call rate <0.95, MAF <0.01 or Hardy-Weinberg P-value <1×10-20 in either the case or control collections. Samples were excluded if identified as outliers on call rate, heterozygosity, ancestry and average probe intensity based on a Bayesian clustering algorithm. Samples were also removed if they exhibited discrepancies between inferred and recorded gender and cryptic relatedness with other WTCCC2 samples (pairwise identity-by-descent >0.05). Autosomal genotype imputation was performed using MaCH based on HapMap Phase 2 European (CEU) reference data.

4. EXPRESSION QUANTITAIVE TRAIT LOCI (eQTL)

We searched for cis and trans effects of the six SNPs at locus 17q25 using several searchable databases: 1) mRNA SNP Browser v 1.0.1 (http://www.sph.umich.edu/csg/liang/asthma/), a database of eQTLs from asthma studies; 2) GTEx (http://www.ncbi.nlm.nih.gov/gtex/GTEX2/),

which currently displays only cis effects from a large scale expression study[16]; and 3) seeQTL (http://www.bios.unc.edu/research/genomic_software/seeQTL), which includes data from 14 human studies, applies additional quality control filters and derives a consensus score by study-specific weighted meta-analyses for each SNP with adjustments for gender and ancestry principle components.[17]

There were significant associations with *TRIM47* expression (mRNA NM_033452.2) within HapMap lymphoblastoid cell lineage (seeQTL FDR adjusted $p \leq 0.01$, GTEx $p < 1x10^{-7}$, and/or mRNA by SNP Browser $p < 1x10^{-7}$). Minor alleles of rs936393, rs9894383, and rs3744017 are associated with increased *TRIM47* expression. The other three SNPs did not have significant associations but are all highly correlated with the exception of rs1055129 which is in moderate LD with the others ($r^2 \approx 0.5$). A proxy of rs1055129, rs2290771 ($r^2=0.8$), was also associated with changes in *TRIM47* expression. rs11869977 was significantly associated with a transcript of *ATP6VOA2* on chromosome 12 (transcript 235255_at) (Beta=0.32, SE=0.09, LOD=2.37, p=0.00094). SNPs within *ATP6VOA2* were not associated with white matter hyperintensity volume and there were no other putative trans-eQTLs.

5. FUNDING

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). The Stroke Association provided additional support for collection of some of the St George's, London cases. The Oxford cases were collected as part of the Oxford Vascular Study which is funded by the Medical Research Council, Stroke Association, Dunhill Medical Trust, National Institute of Health Research (NIHR) and the NIHR Biomedical Research Centre, Oxford. The Edinburgh Stroke Study was supported by the Wellcome Trust (clinician scientist award to Dr. Sudlow), and the Binks Trust. Sample processing occurred in the Genetics Core Laboratory of the Wellcome Trust Clinical Research Facility, Western General Hospital, Edinburgh. Much of the neuroimaging occurred in the Scottish Funding Council Brain Imaging Research Centre (www.sbirc.ed.ac.uk), Division of Clinical Neurosciences, University of Edinburgh, a core area of the Wellcome Trust Clinical Research Facility and part of the SINAPSE (Scottish Imaging Network – A Platform for Scientific Excellence) collaboration (www.sinapse.ac.uk), funded by the Scottish Funding Council and the Chief Scientist Office. Collection of the Munich cases and data analysis was supported by the Vascular Dementia Research Foundation. The Besta Cerebrovascular Diseases Registry (CEDIR) was supported by the Italian Ministry of Health, years 2007–2010 (Annual Research Funding; Grant Numbers: RC 2007/LR6, RC 2008/LR6; RC 2009/LR8; RC 2010/LR8).

The Cardiovascular Health Study was supported by NHLBI contracts N01-HC-85239, N01-HC-85079 through N01-HC-85086; N01-HC-35129, N01 HC-15103, N01 HC-55222, N01-HC-

75150, N01-HC-45133, HHSN268201200036C and NHLBI grants HL080295, HL087652, HL105756 with additional contribution from NINDS. Additional support was provided through AG-023629, AG-15928, AG-20098, and AG-027058 from the NIA. See also http://www.chs-nhlbi.org/pi.htm. DNA handling and genotyping was supported in part by National Center of Advancing Translational Technologies CTSI grant UL1TR000124 and National Institute of Diabetes and Digestive and Kidney Diseases grant DK063491 to the Southern California Diabetes Endocrinology Research Center.

6. REFERENCES

[1] Yatsuya H, Folsom AR, Alonso A, Gottesman RF, Rose KM; ARIC Study Investigators. Postural changes in blood pressure and incidence of ischemic stroke subtypes: the ARIC study. *Hypertension*. 2011;57:167-173.

[2] Rosamond WD, Folsom AR, Chambless LE, Wang CH, McGovern PG et al. Stroke incidence and survival among middle-aged adults: 9-year follow-up of the Atherosclerosis Risk in Communities (ARIC) cohort. *Stroke.* 1999;30:736–743

[3] Ay H, Furie KL, Singhal A, Smith WS, Sorensen AG, Koroshetz WJ. An evidence-based causative classification system for acute ischemic stroke. *Ann Neurol.* 2005;58:688–697

[4] Adams HP, Jr, Bendixen BH, Kappelle LJ, Biller J, Love BB, Gordon DL, 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.

[5] McEvoy M, Smith W, D'Este C, Duke J, Peel R, Schofield P, et al. Cohort profile: The Hunter Community Study. *Int J Epidemiol*. 2010;39:1452-1463

[6] Yadav S, Schanz R, Maheshwari A, Khan MS, Slark J, de Silva R, et al. Bio-Repository of DNA in stroke (BRAINS): a study protocol. BMC Med Genet. 2011;12:34.

[7] Fried LP, Borhani NO, Enright P, Furberg CD, Gardin JM, Kronmal RA, et al. The Cardiovascular Health Study: design and rationale. *Ann Epidemiol.* 1991;1:263-276.

[8] Gretarsdottir S, Thorleifsson G, Manolescu A, Styrkarsdottir U, Helgadottir A, Gschwendtner A, et al. Risk variants for atrial fibrillation on chromosome 4q25 associate with ischemic stroke. *Ann Neurol.* 2008;64:402-409

[9] Foulkes MA, Wolf PA, Price TR, Mohr JP, Hier DB. The Stroke Data Bank: design, methods, and baseline characteristics. *Stroke.* 1988;19:547-554.

.

[10] Heckbert SR, Wiggins KL, Glazer NL, Dublin S, Psaty BM, Smith NL, et al. Antihypertensive treatment with ACE inhibitors or beta-blockers and risk of incident atrial fibrillation in a general hypertensive population. *Am J Hypertens.* 2009;22:538–544.

[11] Smith NL, Rice KM, Bovill EG, Cushman M, Bis JC, McKnight B, et al. Genetic variation associated with plasma von Willebrand factor levels and the risk of incident venous thrombosis. Blood. 2011; 117: 6007–6011.

[12] Meschia JF, Nalls M, Matarin M, Brott TG, Brown RD Jr, Hardy J, et al. Siblings with ischemic stroke study: results of a genome-wide scan for stroke loci. *Stroke.* 2011;42:2726-2732.

[13] Meschia JF, Brott TG, Brown RD Jr, Crook RJ, Frankel M, Hardy J, et al. The Ischemic Stroke Genetics Study (ISGS) Protocol. *BMC Neurol*. 2003;3:4.

[14] International Stroke Genetics Consortium, Wellcome Trust Case-Control Consortium 2. Failure to validate association between 12p13 variants and ischemic stroke. *N.Engl.J Med*. 2010;362:1547-1550

[15] International Stroke Genetics Consortium (ISGC); Wellcome Trust Case Control Consortium 2 (WTCCC2), Bellenguez C, Bevan S, Gschwendtner A, Spencer CC, et al. Genome-wide association study identifies a variant in HDAC9 associated with large vessel ischemic stroke. *Nat Genet.* 2012;44:328-333.

[16] Stranger BE, Nica AC, Forrest MS, Dimas A, Bird CP, Beazley C, et al. Population genomics of human gene expression. Nat Genet. 2007;39:1217-1224.

[17] Xia K, Shabalin AA, Huang S, Madar V, Zhou YH, Wang W, et al. seeQTL: a searchable database for human eQTLs. *Bioinformatics.* 2012;28:451-452.

7. ACKNOWLEDGEMENTS

Wellcome Trust Case Control Consortium 2

We thank S. Bertrand, J. Bryant, S.L. Clark, J.S. Conquer, T. Dibling, J.C. Eldred, S. Gamble, C. Hind, M.L. Perez, C.R. Stribling, S. Taylor and A. Wilk of the Wellcome Trust Sanger Institute's Sample and Genotyping Facilities for technical assistance.

We acknowledge use of the British 1958 Birth Cohort DNA collection, funded by the Medical Research Council grant G0000934 and the Wellcome Trust grant 068545/Z/02, and of the UK National Blood Service controls funded by the Wellcome Trust.

Membership of Wellcome Trust Case Control Consortium 2

Management Committee

Peter Donnelly (Chair)[1,2], Ines Barroso (Deputy Chair)[3], Jennifer M Blackwell[4, 5], Elvira Bramon[6] , Matthew A Brown[7] , Juan P Casas[8] , Aiden Corvin[9], Panos Deloukas[3], Audrey

Duncanson[10], Janusz Jankowski[11], Hugh S Markus[12], Christopher G Mathew[13], Colin NA Palmer[14], Robert Plomin[15], Anna Rautanen[1], Stephen J Sawcer[16], Richard C Trembath[13], Ananth C Viswanathan[17], Nicholas W Wood[18]

Data and Analysis Group

Chris C A Spencer[1], Gavin Band[1], Céline Bellenguez[1], Colin Freeman[1], Garrett Hellenthal[1], Eleni Giannoulatou[1], Matti Pirinen[1], Richard Pearson[1], Amy Strange[1], Zhan Su[1], Damjan Vukcevic[1], Peter Donnelly[1,2]

DNA, Genotyping, Data QC and Informatics Group

Cordelia Langford[3], Sarah E Hunt[3], Sarah Edkins[3], Rhian Gwilliam[3], Hannah Blackburn[3], Suzannah J Bumpstead[3], Serge Dronov[3], Matthew Gillman[3], Emma Gray[3], Naomi Hammond[3], Alagurevathi Jayakumar[3], Owen T McCann[3], Jennifer Liddle[3], Simon C Potter[3], Radhi Ravindrarajah[3], Michelle Ricketts[3], Matthew Waller[3], Paul Weston[3], Sara Widaa[3], Pamela Whittaker[3], Ines Barroso[3], Panos Deloukas[3].

Publications Committee

Christopher G Mathew (Chair)[13], Jenefer M Blackwell[4,5], Matthew A Brown[7], Aiden Corvin[9], Chris C A Spencer[1]

[1]Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, UK; [2]Dept Statistics, University of Oxford, Oxford OX1 3TG, UK; [3]Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK; [4]Telethon Institute for Child Health Research, Centre for Child Health Research, University of Western Australia, 100 Roberts Road, Subiaco, Western Australia 6008; [5]Cambridge Institute for Medical Research, University of Cambridge School of Clinical Medicine, Cambridge CB2 0XY, UK; [6]Department of Psychosis Studies, NIHR Biomedical Research Centre for Mental Health at the Institute of Psychiatry, King's College London and The South London and Maudsley NHS Foundation Trust, Denmark Hill, London SE5 8AF, UK; [7]University of Queensland Diamantina Institute, Brisbane, Queensland, Australia; [8]Dept Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London WC1E 7HT and Dept Epidemiology and Public Health, University College London WC1E 6BT, UK; [9]Neuropsychiatric Genetics Research Group, Institute of Molecular Medicine, Trinity College Dublin, Dublin 2, Eire; [10]Molecular and Physiological Sciences, The Wellcome Trust, London NW1 2BE; [11]Department of Oncology, Old Road Campus, University of Oxford, Oxford OX3 7DQ, UK , Digestive Diseases Centre, Leicester Royal Infirmary, Leicester LE7 7HH, UK and Centre for Digestive Diseases, Queen Mary University of London, London E1 2AD, UK; [12]Clinical Neurosciences, St George's University of London, London SW17 0RE; [13]King's College London Dept Medical and Molecular Genetics, King's Health Partners, Guy's Hospital, London SE1 9RT, UK; [14]Biomedical Research Centre, Ninewells Hospital and Medical School, Dundee DD1 9SY, UK; [15]King's College London Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Denmark Hill, London SE5 8AF, UK; [16]University of Cambridge Dept Clinical Neurosciences, Addenbrooke's Hospital, Cambridge CB2 0QQ, UK; [17]NIHR Biomedical Research Centre for Ophthalmology, Moorfields Eye Hospital NHS Foundation Trust and UCL

Institute of Ophthalmology, London EC1V 2PD, UK; [18]Dept Molecular Neuroscience, Institute of Neurology, Queen Square, London WC1N 3BG, UK.

Membership of The Australian Stroke Genetics Collaborative

Christopher R. Levi, M.D

Centre for Brain and Mental Health Research, University of Newcastle and Hunter Medical Research Institute, NSW, Australia

Jane M. Maguire, Ph.D.

School of Nursing and Midwifery, University of Newcastle, New South Wales, Australia

John Attia, M.D.

Centre for Clinical Epidemiology and Biostatistics, School of Medicine and Public Health, University of Newcastle, New South Wales, Australia

Elizabeth G. Holliday, PhD

Centre for Clinical Epidemiology and Biostatistics, Hunter Medical Research Institute and School of Medicine and Public Health, University of Newcastle, New South Wales, Australia

Rodney J. Scott, Ph.D.

Centre for Bioinformatics, Biomarker Discovery and Information-Based Medicine, Hunter Medical Research Institute, Newcastle, New South Wales, Australia

Lisa F. Lincz, Ph.D.

Hunter Haematology Research Group, Calvary Mater Newcastle Hospital, Newcastle, Australia

Pablo Moscato, Ph.D.

Centre for Bioinformatics, Biomarker Discovery and Information-Based Medicine, Hunter Medical Research Institute, Newcastle, New South Wales, Australia

Simon A. Koblar, M.D.

Stroke Research Program, School of Medicine, University of Adelaide, South Australia, Australia

Jim Jannes, M.D.

Stroke Research Program, School of Medicine, University of Adelaide, Australia

Jonathan W. Sturm, M.D.

Department of Neurosciences, Gosford Hospital, Central Coast Area Health, Australia

Graeme J. Hankey, M.D.

Royal Perth Hospital, Perth, Western Australia, Australia

Ross Baker, M.D.

Royal Perth Hospital, Perth, Western Australia, Australia

Mark W. Parsons, M.D.

Centre for Brain and Mental Health Research, University of Newcastle and Hunter Medical Research Institute, NSW, Australia

Mark McEvoy, Ph.D.

Centre for Clinical Epidemiology and Biostatistics, School of Medicine and Public Health, University of Newcastle, NSW, Australia

Roseanne Peel, M.Sc.

Centre for Clinical Epidemiology and Biostatistics, School of Medicine and Public Health, University of Newcastle, NSW, Australia

Wayne Smith, Ph.D.

Hunter Medical Research Institute and University of Newcastle, NSW, Australia

Martin D. Lewis, Ph.D.

Stroke Research Program, School of Medicine, University of Adelaide, Australia

Tiffany-Jane Evans, B.Sc (Hons)

Centre for Bioinformatics, Biomarker Discovery and Information-Based Medicine, Hunter Medical Research Institute, Newcastle, NSW, Australia

Jonathan Golledge, M.D.

Vascular Biology Unit, School of Medicine and Dentistry, James Cook University, Townsville, Queensland, Australia

Erik Biros, Ph.D.

Vascular Biology Unit, School of Medicine and Dentistry, James Cook University, Townsville, Queensland, Australia

The SWISS study acknowledges the following medicals students for their contributions.

Aparna R Baheti, B.S.

School of Medicine University of Virginia Health System Charlottesville, Virginia
Rodney S Smith, B.S.
School of Medicine University of Virginia Health System Charlottesville, Virginia
Kathryn M Powell, B.S.
School of Medicine University of South Florida, Tampa, Florida