# Genetic Variation at 16q24.2 is associated with small vessel stroke.

Running head: 16q24.2 and small vessel stroke

Matthew Traylor1, Rainer Malik2, Mike A. Nalls3, Ioana Cotlarciuc4, Farid Radmanesh5,6,7, Gudmar Thorleifsson8, Ken B. Hanscombe1, Carl Langefeld9, Danish Saleheen10, Natalia S. Rost6, Idil Yet11, Tim D. Spector11, Jordana T. Bell11, Eilis Hannon12, Jonathan Mill12,13, Ganesh Chauhan14,15, Stephanie Debette14,15, Joshua C. Bis16, WT Longstreth Jr.17, M. Arfan Ikram18, Lenore J. Launer19, Sudha Seshadri20.21, METASTROKE, UK Young Lacunar DNA Study, NINDS Stroke Genetics Network, Neurology Working Group of the CHARGE Consortium, Monica Anne Hamilton-Bruce22, Jordi Jimenez-Conde23, John W. Cole24, Reinhold Schmidt25, Agnieszka Słowik26, Robin Lemmens27,28,29, Arne Lindgren30,31, Olle Melander32, Raji P. Grewal33, Ralph L. Sacco34, Tatjana Rundek34, Kathryn Rexrode35, Donna K. Arnett36, Julie A. Johnson37,38, Oscar R. Benavente39, Sylvia Wasssertheil-Smoller40, Jin-Moo Lee41, Sara L. Pulit42, Quenna Wong43, Stephen S. Rich44, Paul I.W. de Bakker45,46, Patrick F. McArdle47, Daniel Woo48, Christopher D. Anderson5,6,7,49, Huichun Xu50, Laura Heitsch51, Myriam Fornage52, Christina Jern53, Kari Stefansson8,54, Unnur Thorsteinsdottir8,54, Solveig Gretarsdottir9, Cathryn M. Lewis1,13, Pankaj Sharma4, Cathie L.M. Sudlow55, Peter M. Rothwell56, Giorgio B. Boncoraglio57, Vincent Thijs21,22,58, Chris Levi59, James F. Meschia60, Jonathan Rosand5,6,7,48, Steven J. Kittner24, Braxton D. Mitchell61,62, Martin Dichgans2,63, Bradford B. Worrall64, Hugh S. Markus65; on behalf of the International Stroke Genetics Consortium

1 Department of Medical and Molecular Genetics, King’s College London, London, UK

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

3 Laboratory of Neurogenetics, National Institute on Aging, Bethesda, MD, USA

4 Institute of Cardiovascular Research Royal Holloway University of London (ICR2UL), UK

5 Division of Neurocritical Care and Emergency Neurology, Massachusetts General Hospital, Boston, MA, USA

6 J. Philip Kistler Stroke Research Center, Department of Neurology, Massachusetts General Hospital, Boston, MA, USA

7 Center for Human Genetic Research, Massachusetts General Hospital, Boston. MA, USA

8 deCODE genetics/AMGEN, Reykjavik, Iceland

9 Center for Public Health Genomics and Department of Biostatistical Sciences, Wake Forest School of Medicine, Winston-Salem, NC, USA

10 Department of Genetics, Perelman School of Medicine, University of Pennsylvania, PA, USA

11 Department of Twin Research & Genetic Epidemiology, King's College London, London, UK

12 University of Exeter Medical School, University of Exeter, Exeter, UK

13 Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King’s College London, London, UK

14 Inserm Research Center for Epidemiology and Biostatistics (U897) – Team Neuroepidemiology, Bordeaux, France

15 University of Bordeaux, Bordeaux, France

16 Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA, USA

17 Departments of Neurology and Epidemiology, University of Washington, Seattle, WA, USA

18 Department of Neurology, Epidemiology and Radiology, Erasmus MC University Medical Center, Rotterdam, The Netherlands

19 Laboratory of Epidemiology and Population Sciences, National Institute on Aging, Bethesda, MD, USA

20 Boston University School of Medicine, Boston, MA, USA

21 Framingham Heart Study, Framingham, MA, USA

22 Department of Neurology, Royal Adelaide Hospital, Adelaide, South Australia, Australia

23 Neurovascular Research Group (NEUVAS). Neurology Department. Institut Hospital del Mar d’Investigació Mèdica. Barcelona. Spain

24 Department of Neurology, University of Maryland School of Medicine and Baltimore VAMC, MD, USA

25 Department of Neurology, Clinical Division of Neurogeriatrics, Medical University Graz, Austria

26 Department of Neurology, Jagiellonian University, Krakow, Poland

27KU Leuven - University of Leuven, Department of Neurosciences, Experimental Neurology and Leuven Research Institute for Neuroscience and Disease (LIND), Leuven, Belgium

28 VIB, Vesalius Research Center, Laboratory of Neurobiology, Department of Neurology, Leuven, Belgium

29 University Hospitals Leuven, Department of Neurology, Leuven, Belgium

30 Department of Clinical Sciences Lund, Neurology, Lund University, Lund, Sweden

31 Department of Neurology and Rehabilitation Medicine, Skåne University Hospital, Lund, Sweden

32 Department of Clinical Sciences, Lund University, Malmö, Sweden

33 Neuroscience Institute, Saint Francis Medical Center, School of Health and Medical Sciences, Seton Hall University, South Orange, New Jersey, USA

34 Department of Neurology, Miller School of Medicine, University of Miami, USA

35 Harvard Medical School, Boston, USA Center for Faculty Development and Diversity, Brigham and Women's Hospital, Boston, MA, USA

36 Department of Epidemiology, School of Public Health, University of Alabama at Birmingham, AL, USA

37 Department of Pharmacotherapy and Translational Research and Center for Pharmacogenomics, University of Florida, College of Pharmacy, Gainesville, FL, USA  
38 Division of Cardiovascular Medicine, College of Medicine, University of Florida, Gainesville, FL, USA

39 Department of Neurology, University of British Columbia, Vancouver, British Columbia, Canada

40 Department of Epidemiology and Population Health, Albert Einstein College of Medicine, NY, USA

41 Stroke Center, Department of Neurology, Washington University School of Medicine, WA, USA

42 Department of Medical Genetics, University Medical Center Utrecht, Utrecht, Netherlands

43 Department of Biostatistics, University of Washington, WA, USA

44 Center for Public Health Genomics, University of Virginia School of Medicine, VA, USA

45 Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, the Netherlands;

46 Department of Medical Genetics, Center for Molecular Medicine, University Medical Center Utrecht, Utrecht, the Netherlands.

47 Department of Medicine, University of Maryland School of Medicine, MD, USA

48 University of Cincinnati College of Medicine, Cincinnati, OH, USA

49 Program in Medical and Population Genetics, Broad Institute, Boston, MA, USA

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

51Division of Emergency Medicine, Washington University School of Medicine, St Louis, MO, USA

52 The University of Texas Health Science Center at Houston, Houston, TX, USA

53 Institute of Biomedicine, the Sahlgrenska Academy at University of Gothenburg, Sweden

54 Faculty of Medicine, University of Iceland, Reykjavik, Iceland

55 Center for Clinical Brain Sciences & Institute of Genetics and Molecular Medicine, University of Edinburgh, UK

56 Nuffield Department of Clinical Neurosciences, University of Oxford, UK

57 Department of Cerebrovascular Diseases, Fondazione IRCCS Istituto Neurologico "Carlo Besta", Milano, Italy

58 Department of Neurology, Austin Health and Florey Institute of Neuroscience and Mental Health, Heidelberg, Australia

59 John Hunter Hospital, Hunter Medical Research Institute and University of Newcastle, Newcastle, NSW, Australia

60 Department of Neurology, Mayo Clinic Jacksonville, Jacksonville, FL, USA

61 Department of Medicine, University of Maryland School of Medicine, MD, USA

62 Geriatrics Research and Education Clinical Center, Baltimore Veterans Administration Medical Center, Baltimore, MD, USA

63 Munich Cluster of Systems Neurology, SyNergy, Munich, Germany

64 Departments of Neurology and Public Health Sciences, University of Virginia School of Medicine, Charlottesville, VA, USA

65 Stroke Research Group, Division of Clinical Neurosciences, University of Cambridge, UK

Corresponding Author:

Matthew Traylor ([matthew.traylor@kcl.ac.uk](mailto:matthew.traylor@kcl.ac.uk))

Department of Medical and Molecular Genetics,

King’s College London

8th Floor, Tower Wing,

Guys Hospital, Great Maze Pond

London SE1 9RT

Title Characters: 76; Running Head Characters: 31

Word Count: 4,090; Abstract: 247; Tables: 2; Figures: 5; References: 50

Running head: 16q24.2 and small vessel stroke

# Abstract

Objective: Genome-wide association studies (GWAS) have been successful at identifying associations with stroke and stroke subtypes, but have not yet identified any associations solely with small vessel stroke (SVS). SVS comprises a quarter of all ischaemic stroke and is a major manifestation of cerebral small vessel disease, the primary cause of vascular cognitive impairment. Studies across neurological traits have shown younger onset cases have an increased genetic burden. We leveraged this increased genetic burden by performing an age-at-onset informed GWAS meta-analysis, including a large younger onset SVS population, to identify novel associations with stroke.

Methods: We used a three-stage age-at-onset informed GWAS to identify novel genetic variants associated with stroke. On identifying a novel locus associated with SVS, we assessed its influence on other small vessel disease phenotypes, as well as on mRNA expression of nearby genes, and on DNA methylation of nearby CpG sites in whole blood and in the fetal brain.

Results: We identified an association with SVS in 4,203 cases and 50,728 controls on chromosome 16q24.2 (OR(95% CI)=1.16(1.10-1.22); p=3.2x10-9). The lead SNP (rs12445022) was also associated with cerebral white matter hyperintensities (OR(95% CI)=1.10(1.05-1.16); p=5.3x10-5; N=3,670), but not intracerebral haemorrhage (OR(95% CI)=0.97(0.84-1.12); p=0.71; 1,545 cases, 1,481 controls). rs12445022 is associated with mRNA expression of *ZCCHC14* in arterial tissues (p=9.4x10-7), and DNA methylation at probe cg16596957 in whole blood (p=5.3x10-6)

Interpretation: 16q24.2 is associated with SVS. Associations of the locus with expression of *ZCCHC14* and DNA methylation suggest the locus acts through changes to regulatory elements.

# Introduction

Genome-wide association studies (GWAS) enable identification of common genetic variants that influence disease risk and have proved successful in elucidating pathophysiological mechanisms underlying diseases with a genetic influence.1 A number of GWAS associations have recently been identified with ischaemic stroke, almost all of which have been associated with specific stroke subtypes.2-4 A number of genetic associations have been reported with cardioembolic (CE) and large artery stroke (LAS), but in contrast there have been no robust associations solely with small vessel stroke (SVS). This is despite epidemiological data that suggest genetic factors are particularity important for SVS. For example, there are a number of monogenic stroke disorders associated with SVS,5 and family history studies have shown a strong association between SVS and a family history of stroke.6 Similarly, related traits, including white matter hyperintensities, have been shown to have high heritability.7

SVS itself comprises a quarter of all ischaemic stroke and is one of the clinically overt manifestations of cerebral small vessel disease (SVD), the major cause of vascular cognitive impairment. Other radiological features of SVD include white matter hyperintensities (WMH) best seen on T2-weighted MRI, cerebral microbleeds - seen on gradient echo MRI, and intracerebral hemorrhages (ICH).8 Despite its importance, the pathogenesis of SVD remains poorly understood and this limits the development of proven treatments for established disease.

One consistent finding across adult-onset neurological complex diseases including Parkinson’s disease,9 Alzheimer’s disease,10 and stroke,11 is that younger onset cases have a stronger genetic burden from common disease-associated SNPs. Leveraging this increased burden, by focussing on younger onset cases in analysis of genetic data, can lead to detection of novel trait-associated variants.11 This may be particularly relevant for SVS, as epidemiological studies have shown stronger associations with SVS and a family history of stroke in younger stroke cases.6

Here, we perform an age-at-onset informed GWAS meta-analysis in stroke, including a large population of younger onset (age<70) small vessel stroke (SVS) cases. We perform analysis for all ischaemic stroke (IS) and its three subtypes: cardioembolic (CE), large artery stroke (LAS) and SVS. Using this approach we identify a novel association with SVS, seek further validation of the locus in other SVD phenotypes, and assess the influence of SNPs at the locus on mRNA expression of nearby genes and DNA methylation at nearby CpG sites.

# Methods

*Study design*

We employed a three-stage design for the association analysis (Figure 1). In brief, in stage I we performed association analysis of stroke phenotypes in 10,210 cases and 12,285 controls of European ancestry from Europe, United States, and Australia; most of which contributed to the METASTROKE ischaemic stoke GWAS meta-analysis – and all of which have been described previously (Table 1). 2, 12, 13 In all cases, diagnosis of stroke was based on clinical evaluation with radiological confirmation. Subtyping of stroke cases was based on the TOAST criteria; in this analysis we considered the CE, LAS and SVS subtypes.14 Of note, our SVS analysis included a large sample (1,012 cases, 970 controls) of younger onset (age<70) MRI-confirmed lacunar strokes, meaning that although we investigated all subtypes, we had most power to identify associations with SVS.

In stage II, we took 3 SNPs from the top 25 loci from each phenotype forward for a first in silico replication in the NINDS Stroke Genetics Network (SiGN), 2 which consisted of 7,743 cases and 17,790 controls. We meta-analysed stages I and II together and identified 3 loci with p<5x10-7. Finally in stage III, we determine whether these 3 SNPs were associated with the phenotype in which they were identified (CE or SVS) by *in silico* replication in a large Icelandic population (deCODE; 520 SVS cases, 1,100 CE cases, 50,728 controls; stage III).

*Genotyping and Imputation*

Genotyping, quality control and imputation of all studies has been described previously.2, 3, 13 All studies were genotyped on commercially available arrays from Illumina or Affymetrix and imputed to 1000 Genomes phase 1 reference panels using IMPUTE or MACH.15 Imputation quality score was assessed by calculating the ratio of the observed to the expected binomial variance of the allele dosage.

*Association analysis*

Association analysis was performed using a covariate-informed approach,11, 16 which we,and others, have implemented previously.11, 17 Briefly, the approach uses case/control status and a covariable – in this case age-at-onset – to estimate each individual’s stroke liability, which can be interpreted as their underlying propensity to stroke, on a normally-distributed scale. In this analysis cases with an earlier age-at-onset take more extreme positive values than late onset cases as, due to the lower prevalence of stroke at younger ages, they are assumed to have higher stroke liability. Conversely, controls who are older and stroke-free at age-at-observation take more extreme negative value than younger controls as they have been stroke-free for a longer time and are therefore assumed to have a lower stroke liability.

In this analysis, the approach was implemented in our software, CIAO (provided at https://sites.google.com/site/mtraylor263/software/covariate-informed-gwas-analysis). Specifically, the approach taken is to model phenotype data using a continuous unobserved normally distributed quantitative trait, called the disease liability (), where and denotes the genetic effects. Then an individual is a case (*z=1*) if and only if and is a control (*z=0*) otherwise. is a parameter estimating the effect of a given covariate *j* on the liability scale. denotes the disease prevalence *p* at the covariate mean under a normal cumulative distribution function . This model is used to approximate the effect of a disease covariate – in this case age-at-onset - on the liability scale, based on estimates of risk of ischaemic stroke by age from epidemiological data, thereby estimating . For this analysis, the gender-specific risk of ischaemic stroke by age from an index age of 55 was obtained from population-based estimates (1.8%, 5.4%, and 12.1% before 65, 75, and 85, respectively in women; 2.4%, 7.3%, and 12.6% before 65, 75, and 85, respectively in men).18 We assumed that 20% of ischaemic stroke cases had each of the cardioembolic, small vessel or large vessel stroke subtypes, approximating proportions observed in population-based studies..19 We developed two models for our analysis; one based on the risk rates for all ischaemic stroke, and secondly for the three stroke subtypes. We used these models to calculate posterior mean liabilities after conditioning on age-at-onset for the four stroke phenotypes separately (). Controls were modelled in the same way, but were assumed to take the posterior mean from the lower (unaffected) portion of the distribution in the liability threshold model (. Where age data was missing, individuals were assigned the median age value (<1% of cases). Regression was then performed on posterior liabilities (by multiplying the number of samples by the squared correlation between the expected genotype dosage and posterior mean liabilities for each of the discovery cohorts in the four ischaemic stroke phenotypes (all IS, CE, LAS, SVS). Ancestry-informative principal components were included where appropriate, using the EIGENSTRAT procedure.20 Any residual inflation was accounted for by adjusting results by the genomic inflation factor, λ.21 In all analyses, SNPs with imputation quality score<0.7 or minor allele frequency<0.01 were excluded and meta-analysis was performed using Stouffer’s method in METAL.22

*Further analysis of a novel locus associated with small vessel stroke*

For a novel variant associated with SVS, we performed further analysis to elucidate the association for different groups based on age-at-onset. Firstly, for datasets in stage I and II, we divided the cases into quartiles based on age-at-onset and estimated the association of the SNP with each quartile using logistic regression with all controls, meta-analysing using a fixed-effects inverse variance weighted approach (data not available in BRAINS, MGH-GASROS, ISGS/SWISS). Secondly, we interrogated associations at the locus in non-European ancestry populations, comprising 657 small vessel African-American stroke cases and 3,251 matched controls from the NINDS Stroke Genetics Network and African or African-Caribbean ancestry individuals from the South London Ethnicity and Stroke Study (SLESS), 2, 23 and 314 SVS cases and 5,193 controls of Pakistani ancestry from the RACE study.3 We used logistic regression to evaluate the association within each group, and evaluated the overall transethnic association by meta-analysing using Stouffer’s method.

In addition, we explored association of the SNP with other SVD phenotypes. We evaluated association of the SNP with 1) white matter hyperintensity volumes (WMHV) measured on T2-weighted MRI in 3,670 ischaemic stroke patients of European ancestry,24 2) in MRI-defined small subcortical brain infarcts (SSBI) brain infarcts in 17,197 trans-ethnic individuals (85.7% European; 8.8% African-American; 3.5% Hispanic; 1.0% Chinese; 1.0% Malay) from community studies recruited within the neuro-CHARGE consortium (mean age 68.90 ± 10.31; 1,986 with infarcts). SSBI were defined as MRI-defined brain infarcts of 3-15 mm or 3-20 mm in size, located in the basal ganglia, the white matter, or the brainstem. Association analysis was performed overall, and for the subset of cases with extensive WMH burden – defined as the top age-specific quartile of WMH volume on a quantitative scale or above the age-specific median by 5-year age-categories for studies using semi-quantitative measurements of WMH burden; N=549; 3) ICH in 1,545 European ancestry cases and 1,481 controls, described previously,25 and stratified according to lobar or nonlobar location.

*Evaluation of regulatory chromatin states, mRNA expression and DNA methylation*

To investigate a novel locus, we used existing resources and performed some further analyses to characterize its regulatory potential. We interrogated chromatin states and regulatory motifs from ENCODE and Epigenomics Roadmap using Haploreg v4.1.26 We also evaluated whether the associated SNP influences gene expression using GTEx portal.27 Upon identifying an association between the SNP and expression of a nearby gene, we evaluated the evidence that the association signal for SVS and gene expression derives from the same causal variant using a Bayesian colocalisation test.28 Using the R coloc package (<http://cran.r-project.org/web/packages/coloc>), we compared 5 models for SNPs with 50Kb of our lead SNP using the approach (H0: No association with either trait; H1: Association with SVS, not with expression; H2: Association with expression, not with SVS; H3: Association with SVS and expression, two independent SNPs; H4: Association with SVS and expression, one shared SNP).

Next, we assessed whether the lead SNP (rs12445022), or 3 SNPs in linkage disequilibrium (LD) (rs4843625, rs12920915, rs12444224), influence DNA methylation levels in whole blood. We evaluated genetic associations of whole blood DNA methylation levels at selected CpG-sites profiled on the Illumina Infinium HumanMethylation450 BeadChip array in a group of 660 monozygotic (MZ) female twins (mean age 59, age range 18 to 79). These individuals were research volunteers from the TwinsUK cohort in the United Kingdom.29 All were of European ancestry. For each CpG-site of interest we calculated the normalised methylation means for the 330 MZ twin pairs as a phenotype in the genetic analysis, and took into account covariates including smoking, BMI, age, methylation plate, and blood cell count estimates. TwinsUK imputed genotypes were obtained for the 1000 genomes reference set,30 where we excluded SNPs with Hardy–Weinberg p<1x10-4, Minor allele frequency (MAF) <5% and those with IMPUTE info value <0.8. We tested for association with our SNP, or SNPs in close LD (r2>0.6) with DNA methylation at CpG-sites. We used p<4x10-5, equivalent to a false discovery rate (FDR) <5%,31 to identify significant cis-mQTL associations.

Finally, we explored genetic associations at 16q24.2 (defined as within 50Kb of rs12445022) with DNA methylation profiles in 166 human fetal brain samples (92 male, 74 female) ranging from 56–166 days post-conception initially using publically available data – which holds results for mQTL associations reaching the study-wide significance threshold (http://epigenetics.essex.ac.uk/mQTL/). Methods for this study have been published in detail elsewhere.32 Briefly, DNA methylation levels were profiled on the Illumina Infinium HumanMethylation450 BeadChip array and SNP genotypes were obtained from the Illumina HumanOmniExpress BeadChip and imputed to 1000 Genomes phase 3 using SHAPEIT and Minimac3 via the Michigan Imputation Server.15, 33 SNP-methylation probe pairs were tested using the R package MatrixEQTL,34 including covariates to control for age, sex and ancestry-informative principal components. Upon identifying a significant association at 16q24.2, we performed additional analyses (not publicly available: we gained access to the data) to test whether any of our 4 SNPs (rs12445022, rs4843625, rs12920915, rs12444224) were associated with methylation at the identified probe. We again used p<4x10-5, equivalent to a false discovery rate (FDR) <5%,31 to identify significant cis-mQTL associations.

# Results

*Association analysis*

In phase I association analysis we confirmed previous associations between *HDAC9* and LAS (rs2107595, p=3.0x10-8) and between *PITX2* and CE (rs192172299, p=2.0x10-9).3, 4 Previous associations between *ZFHX3* and CE and between *MMP12* and LAS did not reach genome-wide significance in this analysis (rs879324, p=5.0x10-7and rs586701, p=0.0014; respectively).11 A SNP in a region close to *HABP2* previously associated with young onset ischaemic stroke was also significant, albeit not genome-wide, in this analysis (rs11196288; p=2.4x10-4).35 Genomic inflation λ and the equivalent values scaled to 1000 cases and 1000 controls (λ­1000),36 were well controlled across all analyses (IS, λ (λ­1000) =1.05 (1.00);CE, λ (λ­1000) =1.02 (1.00); LAS, λ (λ­1000) =1.02 (1.00); SVS. λ (λ­1000) =1.01 (1.00)).

We took 25 independent loci forward (3 SNPs in LD from each locus selected on p-value) from each analysis (IS, CE, LAS, SVS) for *in silico* replication in the NINDS Stroke Genetics Network study (stage II). Information on these SNPs is provided in Supplementary Tables 1-4. Following this analysis , excluding previously reported associations, three loci showed significance at p<5x10-7 (two with SVS, one with CE) and one was genome-wide significant (rs12445022, p=4.4x10-8, associated with SVS). We followed up all three loci in a second *in silico* replication (stage III) in a large Icelandic population (deCODE). A single SNP, rs12445022, showed evidence of replication (p=0.011). When performing a meta-analysis across all populations, rs12445022 was associated with SVS at genome-wide significance (p=3.2x10-9; Figure 2). The SNP was either genotyped or well imputed (info>0.9) in all cohorts and lies in an intergenic region between Junctophilin 3 (*JPH3*) and Zinc Finger, CCHC Domain Containing 14 (*ZCCHC14*). To confirm the association with rs12445022, we repeated the analysis using logistic regression; the approach taken in a conventional GWAS. The association was validated using this method, and associations were consistent across populations (OR(95% CI)=1.16(1.10-1.22); p=1.3x10-8; heterogeneity p=0.56; Figure 3).

*Further analysis of a 16q24.2 novel locus associated with small vessel stroke*

We evaluated association of the lead SNP in different quantiles of age at stroke onset, using all controls in each analysis. The strongest associations were observed in younger onset cases, suggesting the influence of the SNP might be greatest in these individuals (Figure 4).However, this was not demonstrated statistically (p>0.05).

We performed further analysis to assess whether the SVS-associated SNP influenced other manifestations of cerebral SVD. The SNP (rs12445022) was also associated with increased T2-WMHV (OR(95% CI)=1.10(1.05-1.16); p=5.3x10-5; Figures 2,5), and showed little heterogeneity across study groups (heterogeneity p=0.58). Conversely, the SNP was not associated with ICH – neither overall, nor in subgroups divided by lobar/non-lobar location. For SSBI, the direction of effect was the same as for SVS, but the effect was weaker and non-significant (OR(95% CI)=1.05(0.97-1.14); p=0.28). For the subgroup with WMH, the effect was stronger – and similar to that observed for SVS, but was again non-significant (OR(95% CI)=1.15(0.99-1.33); p=0.076).

We next evaluated the identified locus in non-European ancestry populations. The SNP had a similar frequency to Europeans in South Asians from RACE (MAF=37%), but was rarer in African ancestry populations, consisting of African-Americans from NINDS Stroke Genetics Network and United Kingdom individuals of African or African-Caribbean ethnicity from SLESS (MAF=14%). Associations with the SNP were in the same direction as in European ancestry populations (Figure 4), but did not reach statistical significance in either ancestry, reflecting the much smaller sample sizes. However, when combining data from all populations, evidence for association at the SNP (p=1.4x10-9) was stronger than in European ancestry populations alone, which might suggest a common association across populations. Indeed, there was not evidence of a significant difference in the strength of association between the European and non-European ancestry individuals (p=0.64).

*Regulatory chromatin states, mRNA expression and DNA methylation related to 16q24.2*

We used existing databases to assess the functional consequences of SNPs in the 16q24.2 region. Firstly, we used the Haploreg v4.1 database to interrogate chromatin states and regulatory motifs from ENCODE and NIH Roadmap Epigenomics Mapping Consortium. 26, 37, 38 The database showed that our lead SNP influences chromatin states in multiple tissues. The SNP is classified as a genic promoter in 9 tissues, an enhancer in 13 tissues and overlaps DNAse1 hypersensitivity sites in 21 tissues. Output from Haploreg is presented as Supplementary Data.

Secondly, we used publicly available databases to evaluate the evidence that the lead SNP influences expression of nearby genes using the GTEx portal.27 The implicated A allele of our lead SNP (rs12445022) was associated with decreased expression of *ZCCHC14* in tibial arterial tissue (p=9.4x10-7; Figure 2). We used a Bayesian colocalisation technique to assess whether the same variant drives the both the SVS association signal and mRNA expression of *ZCCHC14*.28 There was overwhelming evidence in support of H4 (Posterior probability = 99.7%), strongly indicating that a single variant - most likely to be rs12445022 - influences both SVS and expression of *ZCCHC14*.

Finally, we performed analyses to assess whether the lead SNP, or the 3 SNPs in LD, influence DNA methylation at CpG probes in whole blood. We found evidence that the lead SNP, and 3 SNPs in close LD (r2>0.6), influence DNA methylation at 4 nearby CpG sites (cg16596957, cg10312981, cg03020503, cg00555085; all p<4.0x10-5, Table 2). The implicated A allele of rs12445022 was associated with decreased methylation at the cg16596957 probe (beta(SE)=-0.38(0.082); p=5.3x10-6). The same 16q24.2 region by CpG probe (cg16596957) association was also recently reported in another study in whole blood.40 In addition, we looked for an association between SNPs at the 16q24.2 locus and DNA methylation levels in fetal brains, initially using publicly available data (http://epigenetics.essex.ac.uk/mQTL/). There was a strong association with SNPs in distant LD with our lead SNP (rs8047314 ~ cg08031982; p=7.1x10-14; r2=0.16 with rs12445022). We then performed additional analyses (not publicly available: we gained access to the data) to test if our lead SNP, or the SNPs in close LD, were associated with methylation at cg08031982. We could identify no associations that reached our significance threshold (p<4.0x10-5). However, there was a near-significant association of rs12920915 and rs4843625 with methylation at the cg08031982 probe (both p=7.8x10-5). Our lead SNP, rs12445022 was not associated (p=9.9x10-4).

# Discussion

Genome-wide association studies in SVS have largely been disappointing. Some studies have suggested that an association with all IS at the highly pleiotropic 12q24.12 is driven by an association with SVS, 2 but no genome-wide significant associations specifically with SVS have yet been identified. Using an age-of–onset informed analysis approach we identified a novel locus at 16q24.2 associated with SVS. The SNP was also associated using a standard logistic regression approach, but was less significant – a difference of almost an order of magnitude (p=3.2x10-9 compared to p=1.3x10-8).In addition, the association was stronger with younger onset SVS, suggesting a greater influence in these individuals. We tested whether the 16q24.2 association extends to other cerebral SVD related phenotypes. We showed that the same locus also influences WMH, and may have a similar effect on MRI-defined subcortical brain infarcts from prospective studies, although the association did not reach significance in our analysis. However, the locus does not appear to influence risk of ICH. A SNP in the same 16q24.2 region (rs4081947), in partial LD with our SNP (r2=0.28), was also recently reported to be associated with migraine in a large GWAS meta-analysis. 41 These data provide strong supportive evidence that this 16q24.2 locus harbours variants that influence diseases of the cerebral vasculature.

Identifying the mechanisms by which GWAS associations influence disease risk presents additional challenges. In this case, the underlying mechanism and the specific genes implicated remains uncertain. Interrogation of mRNA expression data points to the lead SNP influencing expression of the nearest gene, *ZCCHC14*. This gene is ubiquitously expressed, but is highly expressed in arterial tissues and in the brain. However, its function is not well characterized. Zinc fingers of the CCHC-type contain an 18 digit residue found in the nucleocapsid of retroviruses, and therefore may be important in viral response. Other plausible candidate genes reside nearby. The locus lies around 1Mb away from genes encoding forkhead box proteins including *FOXC2, FOXL1,* and *FOXF1.* These proteins, particularly the closely related *FOXC1 –* a paralogue of *FOXC2*, have been implicated in Mendelian forms of SVS.42 We found no evidence linking our SNP to expression of these genes. However, the function of these proteins changes dramatically between early development and in adult tissues,43 which might explain the absence of an association. This, coupled with the fact that *FOXF2* variants have also recently been implicated in ischaemic stroke,44 make forkhead box proteins exciting targets for follow-up experiments.

Assessing DNA methylation, the process by which methyl groups are added to DNA thereby modifying its function, offers another potential method for mechanistic insight. This epigenetic process influences gene expression and regulation in humans, and may be particularly relevant for diseases such as stroke where gene-environment interactions are likely play an important role in pathogenesis.45 Substantial inter-individual variation exists with respect to age and tissue type.46 However, an important emerging mechanism influencing methylation is local sequence content.47 Notably, recent studies have shown that GWAS findings from stroke-relevant traits such as blood pressure are likely to act by influencing DNA methylation.48 This may be particularly relevant for SVS, in which environmental and other vascular risk factors such as hypertension are important and have been shown to interact with disease risk.49 We evaluated whether our associated SNP (rs12445022), or SNPs in close LD, influence methylation of nearby CpG sites. We found evidence from whole blood that the same genetic variation influences DNA methylation. SNPs in distant LD also influenced DNA methylation at a different probe (cg08031982) in the fetal brain. Further evidence comes from published studies in lung, breast, and kidney tissues,31 as well as in utero,50 all of which have shown that the genetic variation at the same 16q24.2 region influences methylation at the cg08031982 probe. Interestingly, the CpG sites influenced by the locus appear to differ by tissue, with different probes affected in whole blood compared to fetal brain. This might imply tissue-specific functional consequences of the locus and therefore highlights the importance of performing follow-up experiments in appropriate tissues. Based on the evidence presented here, we can only speculate on how genetic variation at the locus leads to increased risk of SVS. One hypothesis is that expression of *ZCCHC14*, or other proteins, is mediated through altered methylation of the probes identified. This might occur, in part, in response to environmental stimuli. Evaluating these hypotheses in a relevant tissue type will be an important future analysis to identify the causal mechanisms leading to SVS.

This study has limitations. Our results suggested that the association may be present in other ethnicities, but we had an insufficient number of cases to establish common risk conclusively. Follow-up studies are therefore required in other ethnic groups. In addition, downstream functional experiments will be required to determine the consequences of the identified association. The mRNA expression and methylation analyses presented herein were constrained by available tissue types. Validation of the findings in more disease relevant tissue types such as cerebral small vessels therefore represent important follow-up analyses, although obtaining such tissue in a state to allow mRNA studies is very challenging. We performed mRNA expression and methylation analyses using either the lead SNP (rs12445022) or 3 LD SNPs. The results should be interpreted with the limitation that we cannot be certain that any of these SNPs is the causal variant. Radiological confirmation of SVS in this study was performed using either CT or MRI. Evidence shows that MRI is considerably more reliable at identifying SVS. Replication of the association in an MRI-confirmed population may therefore provide a more accurate estimate of the effect of the locus on SVS risk. Similarly, interrogation of causative classification system (CCS) definitions of SVS may provide further insights. 51 Another method of interrogating the combined influence of age and genotype is by testing for an interaction. In this analysis, we were unable to do this as age was not available in some sets of controls (e.g WTCCC2).

In this large genome-wide meta-analysis using an age-at-onset informed approach, we have identified the first genome-wide significant locus that is associated solely with SVS. Our findings, which point to subtle changes in gene expression and DNA methylation influencing disease risk, show that strategies that account for different liability across disease related covariates such as age can identify novel associations with disease.

# Acknowledgements

Matthew Traylor is funded by the NIHR Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust and King's College London. Hugh Markus is supported by an NIHR Senior Investigator award and his work is supported by NIHR Comprehensive Biomedical Research Unit funding awarded to Cambridge University Hospitals Trust. Cathryn Lewis receives salary support from the National Institute for Health Research (NIHR) Mental Health Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King’s College London. 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). Robin Lemmens is a senior clinical investigator of FWO Flanders. The TwinsUK study was funded in part by the European Research Council (ERC 250157), and from the TwinsUK resource, which receives support from the Wellcome Trust and the National Institute for Health Research (NIHR)-funded BioResource, Clinical Research Facility and Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust in partnership with King's College London. SNP Genotyping was performed by The Wellcome Trust Sanger Institute and National Eye Institute via NIH/CIDR. The SiGN study was funded by a cooperative agreement grant from the US National Institute of Neurological Disorders and Stroke, National Institutes of Health (U01 NS069208)

**Author Contributions**

M.T, R.M, C.M.L, B.B.W, H.S.M conceived and designed the study. M.T, B.B.W, H.S.M drafted the manuscript. M.T, K.B.H drew the figures. M.T, R.M, M.A.N, I.C, F.R, P.S, D.S, M.A.H-B, C.L.M.S, P.M.R, G.B, V.T, R.L, C.L, J.F.M, J.R, M.D, B.B.W, H.S.M contributed acquisition and analysis of METASTROKE datasets. H.X, L.H, M.F, C.J, J.F.M, B.D.M, S.J.K, M.D, B.B.W J.J-C, J.W.C, R.S, A.S, R.L, A.L, O.M, R.P.G, R.L.S, T.R, K.R, D.K.A, J.A.J, O.R.B, S.W-S, J-M.L, M.T, S.S.R, P.D.B, S.L.P, Q.W, P.F.M, D.W, C.D.A, J.R contributed acquisition and analysis of the NINDS-SIGN data. J.R, D.W, C.L, C.D.A contributed acquisition and analysis of ICH data. G.C, S.D, L.J.L, S.S, J.C.B, W.T.L Jr, contributed acquisition and analysis of neuro-CHARGE data. I.Y, T.D.S, J.T.B, E.H., J.M contributed acquisition and analysis of DNA methylation data. M.T, H.S.M, N.S.R contributed acquisition and analysis of WMH data. G.T, K.S, U.T, S.G contributed acquisition and analysis of deCODE data.

**Conflicts of Interest**

Authors whose affiliations are listed as deCODE/Amgen are employees of deCODE/ Amgen. The remaining authors declare no conflicts of interest.

# Table 1 – Ischaemic Stroke Study participants

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Population | IS | CE | LAS | SVS | Controls | % cases with MRI | Age of cases (mean(s.d)) |
| Stage I Populations | | | | | | | |
| ASGC | 1,162 | 240 | 421 | 310 | 1,244 | 43.0% | 72.9 (13.2) |
| WTCCC2-Germany | 1,174 | 330 | 346 | 106 | 797 | 83.0% | 66.7 (12.9) |
| WTCCC2-UK | 2,374 | 474 | 498 | 460 | 5,175 | 37.2% | 72.2 (12.5) |
| Milano | 366 | 64 | 73 | 25 | 407 | 86.7% | 57.4 (15.6) |
| DNA-lacunar / GENESIS | 1,287 | 80 | 64 | 1,012 | 970 | 100.0% | 59.6 (12.0) |
| LSS | 455 | 157 | 70 | 55 | 455 | 89.0% | 67.7 (14.5) |
| ISGS / SWISS | 1,014 | 235 | 217 | 187 | 1,370 | 83.0% | 66.5 (13.6) |
| BRAINS | 361 | 29 | 120 | 97 | 444 | 30.8% | 74.4 (14.2) |
| MGH-GASROS | 294 | 106 | 68 | 23 | 376 | 60.0% | 66.7 (14.5) |
| VISP | 1,723 | - | - | - | 1,047 | 47.0% | 68.0 (10.7) |
| Total (discovery) | 10,210 | 1,715 | 1,877 | 2,275 | 12,285 |  |  |
| Stage II populations | | | | | | | |
| NINDS Stroke Genetics Network | 7,743 | 2,001 | 1,130 | 1,408 | 17,970 | 62.0% | 66.3 (14.8) |
| Stage III populations | | | | | | | |
| deCODE | - | 1,100 | - | 520 | 20,473 | NA | 72.7 (11.6) |
| TOTAL | 17,953 | 4,816 | 3,007 | 4,203 | 50,728 |  |  |

IS, all ischaemic stroke; CE, cardioembolic stroke; LAS, large artery stroke; SVS, small vessel stroke; ASGC, Australian Stroke Genetics Collaborative; WTCCC2, Wellcome Trust Case Control Consortium 2; LSS, Leuven Stroke Study; BRAINS, Bio-repository of DNA in stroke; MGH-GASROS, The MGH Genes Affecting Stroke Risk and Outcome Study; VISP, The Vitamin Intervention for Stroke Prevention Trial; NA, information not available.

# Table 2 – Significant associations between rs12445022 and LD SNPs (r2>0.6) with cis-methylation probes in whole blood

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| SNP Variant | SNP BP | CpG Probe | Probe BP | RA | beta (SE) | P-value |
| rs12445022 | 87,575,332 | cg16596957 | 87,575,151 | A | -0.38 (0.082) | 5.3x10-6 |
| rs4843625 | 87,576,996 | cg16596957 | 87,575,151 | C | -0.33 (0.075) | 1.3x10-5 |
| rs4843625 | 87,576,996 | cg10312981 | 87,577,304 | C | 0.39 (0.074) | 1.9x10-7 |
| rs4843625 | 87,576,996 | cg03020503 | 87,577,656 | C | 0.35 (0.075) | 5.0x10-6 |
| rs4843625 | 87,576,996 | cg00555085 | 87,616,248 | C | 0.34 (0.075) | 6.6x10-6 |
| rs12920915 | 87,577,521 | cg16596957 | 87,575,151 | T | -0.38 (0.075) | 7.3x10-7 |
| rs12920915 | 87,577,521 | cg10312981 | 87,577,304 | T | 0.38 (0.075) | 1.0x10-6 |
| rs12920915 | 87,577,521 | cg03020503 | 87,577,656 | T | 0.34 (0.076) | 1.1x10-5 |
| rs12920915 | 87,577,521 | cg00555085 | 87,616,248 | T | 0.33 (0.076) | 2.2x10-5 |
| rs12444224 | 87,580,855 | cg16596957 | 87,575,151 | T | -0.38 (0.075) | 8.0x10-7 |
| rs12444224 | 87,580,855 | cg10312981 | 87,577,304 | T | 0.38 (0.075) | 8.7x10-7 |
| rs12444224 | 87,580,855 | cg03020503 | 87,577,656 | T | 0.35 (0.076) | 1.1x10-5 |
| rs12444224 | 87,580,855 | cg00555085 | 87,616,248 | T | 0.32 (0.076) | 2.6x10-5 |

# BP, base position; RA, reference allele

# Figure Legends

# Figure 1 – Flow chart of analyses performed

# Figure 2 – Associations at 16q24.2 with A) small vessel stroke, B) cerebral white matter hyperintensities, C) mRNA expression of ZCCHC14; and D) Gene Locations and associations of the locus with DNA methylation

SVS, small vessel stroke; WMH, white matter hyperintensities; ZCCHC14, Zinc Finger, CCHC Domain Containing 14; JPH3, Junctophilin 3; meQTL, methylation quantitative trait locus.

# Figure 3 – Forest Plot of Associations with rs12445022 under a logistic regression model

# Figure 4 – Association of rs12445022 with small vessel stroke by quartiles of age-at-stroke onset in Europeans

# Figure 5 – Associations with rs12445022 for stroke and cerebral small vessel disease phenotypes

# References

1. Visscher PM, Brown MA, McCarthy MI, Yang J. Five years of GWAS discovery. Am J Hum Genet. 2012 Jan 13;90(1):7-24.

2. Loci associated with ischaemic stroke and its subtypes (SiGN): a genome-wide association study. Lancet Neurol. 2015 Dec 18.

3. Traylor M, Farrall M, Holliday EG, 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 Nov;11(11):951-62.

4. Bellenguez C, Bevan S, Gschwendtner A, et al. Genome-wide association study identifies a variant in HDAC9 associated with large vessel ischemic stroke. Nat Genet. 2012 Mar;44(3):328-33.

5. Tan RY, Markus HS. Monogenic causes of stroke: now and the future. J Neurol. 2015 Dec;262(12):2601-16.

6. Jerrard-Dunne P, Cloud G, Hassan A, Markus HS. Evaluating the genetic component of ischemic stroke subtypes: a family history study. Stroke. 2003 Jun;34(6):1364-9.

7. Turner ST, Jack CR, Fornage M, Mosley TH, Boerwinkle E, de Andrade M. Heritability of leukoaraiosis in hypertensive sibships. Hypertension. 2004 Feb;43(2):483-7.

8. Wardlaw JM, Smith EE, Biessels GJ, et al. Neuroimaging standards for research into small vessel disease and its contribution to ageing and neurodegeneration. Lancet Neurol. 2013 Aug;12(8):822-38.

9. Escott-Price V, Nalls MA, Morris HR, et al. Polygenic risk of Parkinson disease is correlated with disease age at onset. Ann Neurol. 2015 Apr;77(4):582-91.

10. Sleegers K, Bettens K, De Roeck A, et al. A 22-single nucleotide polymorphism Alzheimer risk score correlates with family history, onset age, and cerebrospinal fluid Abeta. Alzheimer's & dementia : the journal of the Alzheimer's Association. 2015 Jun 15.

11. Traylor M, Makela KM, Kilarski LL, et al. A novel MMP12 locus is associated with large artery atherosclerotic stroke using a genome-wide age-at-onset informed approach. PLoS Genet. 2014 Jul;10(7):e1004469.

12. Malik R, Traylor M, Pulit SL, et al. Low-frequency and common genetic variation in ischemic stroke: The METASTROKE collaboration. Neurology. 2016 Mar 2.

13. Traylor M, Bevan S, Baron JC, Hassan A, Lewis CM, Markus HS. Genetic Architecture of Lacunar Stroke. Stroke. 2015 Sep;46(9):2407-12.

14. Adams HP, Jr., Bendixen BH, Kappelle LJ, et al. Classification of subtype of acute ischemic stroke. Definitions for use in a multicenter clinical trial. TOAST. Trial of Org 10172 in Acute Stroke Treatment. Stroke. 1993 Jan;24(1):35-41.

15. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nat Genet. 2012 Aug;44(8):955-9.

16. Zaitlen N, Lindstrom S, Pasaniuc B, et al. Informed conditioning on clinical covariates increases power in case-control association studies. PLoS Genet. 2012;8(11):e1003032.

17. Williams AL, Jacobs SB, Moreno-Macias H, et al. Sequence variants in SLC16A11 are a common risk factor for type 2 diabetes in Mexico. Nature. 2014 Feb 6;506(7486):97-101.

18. Seshadri S, Wolf PA. Lifetime risk of stroke and dementia: current concepts, and estimates from the Framingham Study. Lancet Neurol. 2007 Dec;6(12):1106-14.

19. Lovett JK, Coull AJ, Rothwell PM. Early risk of recurrence by subtype of ischemic stroke in population-based incidence studies. Neurology. 2004 Feb 24;62(4):569-73.

20. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006 Aug;38(8):904-9.

21. Devlin B, Roeder K. Genomic control for association studies. Biometrics. 1999 Dec;55(4):997-1004.

22. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010 Sep 1;26(17):2190-1.

23. Markus HS, Khan U, Birns J, et al. Differences in stroke subtypes between black and white patients with stroke: the South London Ethnicity and Stroke Study. Circulation. 2007 Nov 6;116(19):2157-64.

24. Traylor M, Zhang C, Adib-Samii P, et al. Genome-wide Meta-analysis of Cerebral White Matter Hyperintensities in Patients with Stroke Neurology (in press).

25. Woo D, Falcone GJ, Devan WJ, et al. Meta-Analysis of Genome-Wide Association Studies Identifies 1q22 as a Susceptibility Locus for Intracerebral Hemorrhage. Am J Hum Genet. 2014 Mar 18;94(4):511-21.

26. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res. 2012 Jan;40(Database issue):D930-4.

27. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science. 2015 May 8;348(6235):648-60.

28. Giambartolomei C, Vukcevic D, Schadt EE, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet. 2014 May;10(5):e1004383.

29. Moayyeri A, Hammond CJ, Valdes AM, Spector TD. Cohort Profile: TwinsUK and healthy ageing twin study. Int J Epidemiol. 2013 Feb;42(1):76-85.

30. Abecasis GR, Auton A, Brooks LD, et al. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012 Nov 1;491(7422):56-65.

31. Shi J, Marconett CN, Duan J, et al. Characterizing the genetic basis of methylome diversity in histologically normal human lung tissue. Nature communications. 2014;5:3365.

32. Hannon E, Spiers H, Viana J, et al. Methylation QTLs in the developing brain and their enrichment in schizophrenia risk loci. Nat Neurosci. 2016 Jan;19(1):48-54.

33. Delaneau O, Zagury JF, Marchini J. Improved whole-chromosome phasing for disease and population genetic studies. Nature methods. 2013 Jan;10(1):5-6.

34. Shabalin AA. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. Bioinformatics. 2012 May 15;28(10):1353-8.

35. Cheng YC, Stanne TM, Giese AK, et al. Genome-Wide Association Analysis of Young-Onset Stroke Identifies a Locus on Chromosome 10q25 Near HABP2. Stroke. 2016 Feb;47(2):307-16.

36. de Bakker PI, Ferreira MA, Jia X, Neale BM, Raychaudhuri S, Voight BF. Practical aspects of imputation-driven meta-analysis of genome-wide association studies. Hum Mol Genet. 2008 Oct 15;17(R2):R122-8.

37. Dunham I, Kundaje A, Aldred SF, et al. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012 Sep 6;489(7414):57-74.

38. Kundaje A, Meuleman W, Ernst J, et al. Integrative analysis of 111 reference human epigenomes. Nature. 2015 Feb 19;518(7539):317-30.

39. Westra HJ, Peters MJ, Esko T, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. Nat Genet. 2013 Oct;45(10):1238-43.

40. Liu Y, Li X, Aryee MJ, et al. GeMes, clusters of DNA methylation under genetic control, can inform genetic and epigenetic analysis of disease. Am J Hum Genet. 2014 Apr 3;94(4):485-95.

41. Gormley P, Anttila V, Winsvold BS, et al. Meta-analysis of 375,000 individuals identifies 38 susceptibility loci for migraine. bioRxiv.doi: <http://dx.doi.org/10.1101/030288>.

42. French CR, Seshadri S, Destefano AL, et al. Mutation of FOXC1 and PITX2 induces cerebral small-vessel disease. J Clin Invest. 2014 Nov;124(11):4877-81.

43. Lehmann OJ, Sowden JC, Carlsson P, Jordan T, Bhattacharya SS. Fox's in development and disease. Trends Genet. 2003 Jun;19(6):339-44.

44. Identification of additional risk loci for stroke and small vessel disease: a meta-analysis of genome-wide association studies. Lancet Neurol. 2016 Apr 7.

45. Bell JT, Pai AA, Pickrell JK, et al. DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. Genome Biol. 2011;12(1):R10.

46. Horvath S. DNA methylation age of human tissues and cell types. Genome Biol. 2013;14(10):R115.

47. Banovich NE, Lan X, McVicker G, et al. Methylation QTLs are associated with coordinated changes in transcription factor binding, histone modifications, and gene expression levels. PLoS Genet. 2014 Sep;10(9):e1004663.

48. Kato N, Loh M, Takeuchi F, et al. Trans-ancestry genome-wide association study identifies 12 genetic loci influencing blood pressure and implicates a role for DNA methylation. Nat Genet. 2015 Nov;47(11):1282-93.

49. Adib-Samii P, Devan W, Traylor M, et al. Genetic architecture of white matter hyperintensities differs in hypertensive and nonhypertensive ischemic stroke. Stroke. 2015 Feb;46(2):348-53.

50. Teh AL, Pan H, Chen L, et al. The effect of genotype and in utero environment on interindividual variation in neonate DNA methylomes. Genome Res. 2014 Jul;24(7):1064-74.

51. Ay H, Benner T, Arsava EM, et al. A computerized algorithm for etiologic classification of ischemic stroke: the Causative Classification of Stroke System. Stroke. 2007 Nov;38(11):2979-84.