The Lancet Neurology

Twelve loci provide insights into the genetic basis of lacunar stroke and small vessel disease: a meta-analysis of genome-wide association studies

--Manuscript Draft--

Manuscript Number:	THELANCETNEUROLOGY-D-20-00873R2
Article Type:	Article (Original Research)
Keywords:	STROKE; Lacunar Stroke; Small vessel disease; genome wide association study
Corresponding Author:	Markus Hugh Clinical Neurosciences Cambridge, UNITED KINGDOM
First Author:	Matthew Traylor, PhD
Order of Authors:	Matthew Traylor, PhD
	Elodie Persyn
	Liisa Tomppo
	Sofia Klasson
	Vida Abedi
	Mark Bakker
	Nuria Torres
	Linxin Li
	Steven Bell
	Loes Rutten-Jacobs
	Daniel Tozer
	Christoph Griessenauer
	Yanfei Zhang
	Annie Pedersen
	Pankaj Sharma
	Jordi Jimenez-Conde
	Tatjana Rundek
	Raji Grewal
	Arne Lindgren
	James Meschia
	Veikko Salomaa
	Aki Havulinna
	Christina Kourkoulis
	Katherine Crawford
	Sandro Marini
	Braxton Mitchell
	Steven Kittner
	Jonathan Rosand
	Martin Dichgans

	Christina Jern
	Daniel Strbian
	Israel Fernandez-Cadenas
	Ramin Zand
	Ynte Ruigrok
	Natalia Rost
	Robin Lemmens
	Peter Rothwell
	Christopher Anderson
	Joanna Wardlaw
	Cathryn Lewis
	Markus Hugh
Manuscript Region of Origin:	UNITED KINGDOM
Abstract:	Background Despite causing up to a quarter of all strokes, the genetic basis of lacunar stroke remains poorly understood, with a single locus on 16q24 identified to date. We performed a genome-wide association study (GWAS) of lacunar stroke, expanding the sample size by recruiting robustly phenotyped cases using MRI in diagnosis, to identify novel associations and provide mechanistic insights into the disease. Methods We performed a GWAS of 7,338 cases and 225,258 controls, including newly recruited patients and reanalysis of existing studies, of which 2,987 cases (matched with 29,540 controls) were confirmed using MRI. We used multi-trait analysis of GWAS (MTAG), performing a joint analysis with a study of cerebral white matter hyperintensities (N=42,310) - an etiologically related radiological trait, to uncover additional genetic associations. We performed a transcriptome-wide association study (TWAS), to determine genes for which expression is associated with lacunar stroke, identified significantly enriched pathways using MAGMA and determined cardiovascular risk factors causally associated with the disease using Mendelian Randomization. Findings 5 loci were associated with lacunar stroke in European or Transethnic meta-analysis. A further 7 loci were associated in multi-trait analysis. Two loci contain genes (COL4A2 and HTRA1) involved in monogenic lacunar stroke. Pathway analyses implicated disruption of the extracellular matrix, phosphatidylinositol 5 phosphate binding, and roundabout (ROBO) binding at FDR<0.05, while Mendelian randomization linked elevated blood pressure, history of smoking, and type 2 diabetes in the etiology of lacunar stroke. Interpretation Lacunar stroke has a substantial heritable component, with 12 loci now identified that may represent future treatment targets. These loci provide insights into lacunar stroke pathogenesis, highlighting disruption of the vascular ECM (COL4A2, LOX, SH3PXD2A, GPR126, HTRA1), pericyte differentiation (FOXF2, GPR126), TGF-beta signaling (HTRA1), and myelination (ULK4, GPR126) in disease risk.

Twelve loci provide insights into the genetic basis of lacunar stroke and small vessel disease: a meta-analysis of genome-wide association studies

Matthew Traylor PhD 1-2, Elodie Persyn PhD 3, Liisa Tomppo MD 4, Sofia Klasson PhD 5, Vida Abedi PhD 6, Mark K Bakker MSc 7, Nuria Torres PhD 8, Linxin Li DPhil 9, Steven Bell PhD 10, Loes Rutten-Jacobs PhD 11, Daniel J Tozer PhD 10, Christoph J Griessenauer MD 12-13, Yanfei Zhang PhD 14, Annie Pedersen MD 5, Prof Pankaj Sharma MD 15, Jordi Jimenez-Conde MD 16, Prof Tatjana Rundek MD 17, Raji P. Grewal MD 18, Prof Arne Lindgren MD 19-20, Prof James F Meschia MD 21, Prof Veikko Salomaa MD 22, Aki Havulinna DSc 22-23, Christina Kourkoulis BS 24-26, Katherine Crawford BS 24-25, Sandro Marini MD 24-25, Prof Braxton D. Mitchell PhD 27-28, Prof Steven J. Kittner MD 28-29, Prof Jonathan Rosand MD 24-26,30, Prof Martin Dichgans MD 31-32, Prof Christina Jern MD 5, Daniel Strbian MD 4-33, Israel Fernandez-Cadenas PhD 8,34, Ramin Zand MD 35, Ynte Ruigrok MD 7, Prof Natalia Rost MD 36, Robin Lemmens MD 37, Prof Peter M Rothwell FMed Sci 9, Christopher D Anderson MD 24-26,30, Prof Joanna Wardlaw MD 38, Prof Cathryn M Lewis PhD 39,3, Prof Hugh S Markus FMed Sci 10, Helsinki Stroke Study, The Dutch Parelsnoer Institute-Cerebrovascular accident (CVA) Study Group, NINDS Stroke Genetics Network (SiGN), UK DNA Lacunar Stroke Study Investigators, International Stroke Genetics Consortium (ISGC).

1, Clinical Pharmacology, William Harvey Research Institute, Queen Mary University of London, London, UK

2, The Barts Heart Centre and NIHR Barts Biomedical Research Centre - Barts Health NHS Trust, The William Harvey Research Institute, Queen Mary University London, London, UK

3, Department of Medical and Molecular Genetics, King's College London, London, UK

4, Department of Neurology, Helsinki University Hospital, Helsinki, Finland

5, Department of Laboratory Medicine, Institute of Biomedicine, Sahlgrenska Academy at University of Gothenburg, Gothenburg, Sweden

6, Department of Molecular and Functional Genomics, Weis Center for Research, Geisinger Health System, Danville, PA, USA

7, Department of Neurology and Neurosurgery, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, Netherlands

8, Stroke Pharmacogenomics and Genetics, Sant Pau Institute of Research, Hospital de la Santa Creu I Sant Pau, Barcelona, Spain

9, Centre for the Prevention of Stroke and Dementia, Nuffield Department of Clinical Neuroscience, University of Oxford, Oxford, UK

10, Clinical Neurosciences, University of Cambridge, Cambridge, UK

11, Product Development Personalized Health Care, F. Hoffmann-La Roche Ltd., Basel, Switzerland

12, Neuroscience Institute, Geisinger Health System, Danville, PA, USA

13, Institute of Neurointervention, Paracelsus Medical University, Salzburg, Austria

14, Genomic Medicine Institute, Geisinger Health System, Danville, PA, USA

15, Institute of Cardiovascular Research, Royal Holloway University of London, London, UK

16, Neurovascular Research Group, Department of Neurology of Hospital del Mar-IMIM (Institut Hospital del Mar d'Investigacions Mediques), Universitat Autonoma de Barcelona/DCEXS-Universitat Pompeu Fabra, Barcelona, Spain

17, Evelyn F. McKnight Brain Institute, Department of Neurology, University of Miami Miller School of Medicine, Miami, USA

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

19, Department of Neurology, Skane University Hospital, Lund, Sweden

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

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

22, Department of Public Health Solutions, Finnish Institute for Health and Welfare, Helsinki, Finland

23, , Institute for Molecular Medicine Finland - FIMM HiLIFE, Helsinki, Finland

24, Center for Genomic Medicine, Massachusetts General Hospital, Boston, USA

25, Program in Medical & Population Genetics, Broad Institute of Harvard and MIT, Cambridge, USA

26, Henry and Allison McCance Center for Brain Health, Massachusetts General Hospital, Boston, USA

27, Division of Endocrinology, Diabetes and Nutrition, Department of Medicine, University of Maryland School of Medicine, Baltimore, USA

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

29, Department of Neurology, University of Maryland School of Medicine, Baltimore, USA

30, Department of Neurology, Massachusetts General Hospital, Boston, USA

31, Institute for Stroke and Dementia Research (ISD), LMU Munich, Munich, Germany

32, Munich Cluster for Systems Neurology (SyNergy), Munich, Germany

33, Clinical Neurosciences, University of Helsinki, Helsinki, Finland

34, Neurovascular Research Laboratory and Neurovascular Unit, Institut de Recerca, Hospital Vall d'Hebron, Universitat Autonoma de Barcelona, Barcelona, Spain

35, Geisinger Neuroscience Institute, Geisinger Health System, Danville, PA, USA

36, J. Philip Kistler Stroke Research Center, Massachusetts General Hospital, Boston, USA

37, KU Leuven - University of Leuven, Department of Neurosciences, Experimental Neurology; VIB Center for Brain & Disease Research, University Hospitals Leuven, Department of Neurology, Leuven, Belgium

38, Centre for Clinical Brain Sciences, UK Dementia Research Institute and Row Fogo Centre for Research into the Ageing Brain, University of Edinburgh, Edinburgh, UK

39, Social, Genetic, and Developmental Psychiatry Centre, King's College London, London, UK

75 Corresponding Author:

76 Professor Hugh Markus. Stroke Research Group, Clinical Neurosciences, University of Cambridge,

77 CB2 0QQ. hsm32@medschl.cam.ac.uk

Abstract

Background. Despite causing up to a quarter of all strokes, the genetic basis of lacunar stroke remains poorly understood, with a single locus on 16q24 identified to date. We performed a genome-wide association study (GWAS) of lacunar stroke, expanding the sample size by recruiting robustly phenotyped cases using MRI in diagnosis, to identify novel associations and provide mechanistic insights into the disease.

Methods. We performed a GWAS of 7,338 cases and 225,258 controls, including newly recruited patients and reanalysis of existing studies, of which 2,987 cases (matched with 29,540 controls) were confirmed using MRI. We used multi-trait analysis of GWAS (MTAG), performing a joint analysis with a study of cerebral white matter hyperintensities (N=42,310) - an etiologically related radiological trait, to uncover additional genetic associations. We performed a transcriptome-wide association study (TWAS), to determine genes for which expression is associated with lacunar stroke, identified significantly enriched pathways using MAGMA and determined cardiovascular risk factors causally associated with the disease using Mendelian Randomization.

Findings. 5 loci were associated with lacunar stroke in European or Transethnic meta-analysis. A further 7 loci were associated in multi-trait analysis. Two loci contain genes (*COL4A2* and *HTRA1*) involved in monogenic lacunar stroke. Pathway analyses implicated disruption of the extracellular matrix, phosphatidylinositol 5 phosphate binding, and roundabout (ROBO) binding at FDR<0.05, while Mendelian randomization linked elevated blood pressure, history of smoking, and type 2 diabetes in the etiology of lacunar stroke.

Interpretation. Lacunar stroke has a substantial heritable component, with 12 loci now identified that may represent future treatment targets. These loci provide insights into lacunar stroke pathogenesis, highlighting disruption of the vascular ECM *(COL4A2, LOX, SH3PXD2A, GPR126, HTRA1*), pericyte differentiation (*FOXF2, GPR126*), TGF-beta signaling *(HTRA1),* and myelination (*ULK4, GPR126*) in disease risk.

Funding. This work was supported by a British Heart Foundation Programme Grant.

Introduction

Lacunar strokes are small subcortical infarcts that arise from ischemia in the territory of the deep perforating arteries of the brain.[1,2] They comprise up to a quarter of all ischaemic strokes and are usually due to cerebral small vessel disease (SVD), which is also the most common pathology underlying intracerebral haemorrhage and vascular cognitive impairment. Radiologically, SVD is also characterized by the presence of cerebral white matter hyperintensities (WMH), enlarged perivascular spaces, microbleeds and brain atrophy. [3] Despite its widespread importance, few therapeutic interventions have been proven to reduce SVD. One obstacle to developing new therapeutic approaches has been a lack of understanding of the underlying pathophysiology. One method which has been successfully used to discover pathophysiological processes and uncover potential treatment targets in other complex disease is the use of genetic data derived from genome wide association studies (GWAS). Recent GWAS have identified 35 loci associated with risk of ischaemic stroke and its major subtypes; [4-6] but while many loci have been identified with the other major stroke subtypes (cardioembolic and large artery stroke) only one locus has robustly associated with lacunar stroke so far. [6] This is surprising because lacunar stroke is the stroke subtype most likely to be caused by monogenic disease, [7] and sporadic lacunar stroke has been strongly associated with a family history of stroke. [8] Additionally, studies of other MRI markers of CSVD have demonstrated a substantial genetic component; a recent GWAS identified 31 loci across 3 phenotypes. [9]

We formed a collaboration to perform GWAS of lacunar stroke to identify novel associations and provide mechanistic insights into the disease. We recruited cases of lacunar stroke from hospitals across the UK as part of the UK DNA lacunar stroke studies 1 and 2, and from collaborators within the International Stroke Genetics Consortium (ISGC); and re-analysed data from previous studies. [10-13] As MRI confirmation of lacunar stroke is more reliable, [14,15] we focused on recruiting MRI confirmed cases. We first performed a GWAS of this data to identify novel genetic loci associated with lacunar stroke. Secondly, we used a multi-trait approach to detect additional genetic variation associated with lacunar stroke in a joint analysis with cerebral WMH from a large-scale GWAS.[9] Thirdly, we followed up our initial analyses using the transcriptome-wide association study (TWAS) approach to identify transcribed genes whose expression is associated with lacunar stroke, and used Mendelian randomization to assess common cardiovascular risk factors that contribute to the disease. We performed analyses

separately in MRI confirmed and standard phenotyping groups to assess whether MRI confirmation improves power to detect genetic associations.

Methods

Study Design and Phenotype Definitions

The workflow and design of the study is presented in Figure 1. Lacunar stroke cases were recruited from a combination of acute stroke admissions and outpatient services from Europe, United States, South America, and Australia. Study inclusion criteria are detailed in the appendix, and lacunar stroke classification is given below. For each contributing study, approval for inclusion in this analysis complied with local ethical standards and with local institutional review board/ethics committee oversight. All cases and controls provided informed consent for genetic studies.

For the purposes of this analysis, lacunar stroke samples were divided into two strata: the MRI confirmed group and the standard phenotyping group.

In the *MRI confirmed group,* lacunar stroke was defined as a clinical lacunar syndrome, [16] with an anatomically compatible lesion on MRI (subcortical infarct, ≤15 mm in diameter) - either as a region of high intensity on diffusion weighted imaging (DWI) for acute infarcts, or as a region of low intensity on FLAIR or T1 imaging for non-acute infarcts, [3] and absence of other non-SVD causes of stroke. MRI were centrally reviewed according to a standard proforma to confirm the diagnosis of lacunar stroke and identify any exclusion criteria. All patients underwent comprehensive stroke investigation, including brain MRI, imaging of the carotid arteries, and electrocardiogram. Echocardiography was performed where appropriate. Exclusion criteria were stenosis >50% in the extra or intracranial cerebral vessels, or previous carotid endarterectomy; cardioembolic source of stroke, defined according to the TOAST criteria as high or moderate probability;[10] cortical infarct on MRI; subcortical infarct >15 mm in diameter, as these can be caused by embolic mechanisms (striatocapsular infarcts); and any other specific cause of stroke (e.g. lupus anticoagulant, cerebral vasculitis, dissection, and monogenic cause of stroke).

In the *standard phenotyping group*, lacunar stroke was defined according to the TOAST criteria, based on a clinical lacunar syndrome, and the absence of other causes of stroke, or non-lacunar infarction on computed tomography. [10]

Genotyping and Imputation

Genotyping arrays, quality control filters, and imputation reference panels are listed in the appendix (pp 7-8). All studies inferred the genetic ancestry of samples by comparison with reference populations using principal components analysis. European samples in this study are defined as those which segregated with European ancestry reference samples. The majority of studies were imputed to the Haplotype Reference Consortium build. Where this was not possible due to logistical or ethical reasons, imputation to 1000 Genomes Phase 3 (All ancestry groups) panels was used. [13,17]

Statistical Analysis

All studies used logistic regression to assess the association of single nucleotide polymorphisms (SNP) allele dosages with lacunar stroke, including ancestry informative principal components as covariates as appropriate. All studies included cases with geographically matched controls, as confirmed by principal components analysis. Some studies had a combination of cases based on TOAST diagnosis of lacunar stroke which were re-analysed for MRI confirmation of lacunar stroke, and cases based on TOAST diagnosis of lacunar stroke only for which MRI was either not acquired or we were not able to access. In these circumstances we analysed the MRI confirmed and TOAST only (standard phenotyping) groups separately, and divided the study controls between the two groups to avoid any sample overlap. Any cases with subsequent MRI confirmation of lacunar stroke were omitted from the TOAST only group; all individuals were only analyzed once.

Meta-analysis was performed based on a fixed effects inverse-variance weighted procedure using *METAL*. [18] Meta-analysis was performed in the MRI confirmed and standard phenotyping groups separately, and in all studies combined. We used the principles in Winkler et al. to scrutinize datasets used in the meta-analysis. [19] Per study, we filtered out SNPs with imputation INFO scores < 0.7 or

minor allele frequency (MAF) < 0.01. Additionally, we removed low frequency or poorly imputed SNPs in smaller studies by removing variants with INFO x MAF x Number of cases < 2. [5] Genomic control correction based on genomic inflation lambda was used per study to adjust for any residual inflation. [20] LDSCORE intercept values were used to assess whether population structure had been sufficiently resolved at the meta-analysis level. [21] After meta-analysis we excluded SNPs not present in at least 50% of cases.

We defined significant loci as those containing SNPs reaching $p<5x10^{-8}$ and being in linkage equilibrium ($R^2>0.1$) with other lead SNPs. Where multiple loci met these criteria within 1Mb we used conditional and joint multiple SNP analysis (GCTA-cojo) to determine whether these SNPs remained genome-wide significant in a joint modelling scenario. [22] We used Nagelkerke R^2 values to calculate the proportion of variance explained by genome-wide significant SNPs using the NagelkerkeR2 function in the R fmsb library.[23] We subtracted the R^2 estimate for the model including only principal components from the model also containing the genome-wide significant SNPs to obtain an estimate of R^2. We used a genome-based restricted maximum likelihood (GREML) approach, implemented in GCTA, [24,25] and LD Score regression, [21] to estimate heritability of lacunar stroke (MRI confirmed and non confirmed).

Multi-Trait Analysis

We applied multi-trait analysis of GWAS (MTAG), [26] performing a joint analysis with a large study of cerebral WMH on MRI (N=42,310), [9] which shares a common etiology with lacunar stroke through cerebral SVD, to uncover additional genetic variation associated with lacunar stroke. We considered associations significant if they attained a p-value < $5x10^{-8}$ in MTAG analysis, had a p-value <0.05 for association with WMH and lacunar stroke in univariate analysis, and showed greater significance in MTAG analysis than in univariate analyses for WMH or lacunar stroke. To confirm our associations, we used an alternative approach, Bayesian multivariate analysis of summary statistics (BMASS). [27]

Transcriptome-Wide Association Study

We used the transcriptome-wide association study (TWAS) approach to identify genes for which genetically altered expression was associated with lacunar stroke. Analyses were performed using FUSION, [28] from gene expression models derived from the GTEx v7, [29] CommonMind Consortium (CMC), [30] and Young Finns Study (YFS) datasets. [31] The CMC gene expression tissues (labelled as CMC-brain) were collected from the dorsolateral prefrontal cortex of individuals with schizophrenia or controls (TWAS-N=452). In the YFS study (labelled as YFS-whole blood), peripheral blood gene expression has been collected for 1,650 participants (TWAS-N=1,264). Among the available GTEx tissues, we focused our TWAS analysis on the aortic artery (TWAS-N=267), coronary artery (TWAS-N=152), tibial artery (TWAS-N=388) and whole blood (TWAS-N=369), based on the assumption that these tissues would be the most relevant for lacunar stroke pathogenesis. Bonferroni correction for multiple testing was applied taking into account the total number of tested genes across the tissues. TWAS results were further investigated with colocalization analysis of expression quantitative trait loci (eQTLs) and GWAS signals with the R package COLOC, [32] to assess whether the observed eQTL and GWAS associations were consistent with a common shared association.

Bioinformatics Analyses for Novel Associations

We used Phenoscanner to query whether our genome-wide significant SNPs have been associated with DNA methylation, [33,34] metabolite or protein levels from genome-wide studies at genome-wide significance ($p<5x10^{-8}$) in other GWAS. We scanned DrugBank and DGIdb to assess the therapeutic potential of targeting associated genes. [35,36]

Pathway Analysis

To identify biological pathways significantly associated with risk of lacunar stroke, we used MAGMA. [37] We first used MAGMA to calculate significance of each gene based on association results, and then used these gene-level statistics to estimate enrichment of Gene Ontology (GO) pathways from the Molecular Signatures Database (MSigDB) using a gene-set enrichment analysis approach. [38] We

investigated only GO terms containing at least 4 and less than 200 genes and considered pathways attaining a false discovery rate (FDR) < 0.05 as being significantly associated with lacunar stroke.

Mendelian Randomization Analysis

We performed Mendelian Randomization analyses to determine whether any lipid (low density lipoprotein, high density lipoprotein, triglycerides), [39] blood pressure (systolic blood pressure, diastolic blood pressure, pulse pressure), [40] metabolic (type 2 diabetes, body mass index), [41,42] and lifestyle risk factors (ever smoking) have a causal impact on lacunar stroke based on genetics. [43,44] Instrumental variables were independent (LD $r^2<0.01$) genome-wide significant ($p<5x10^{-8}$) variants associated with each trait from previous analyses, and are listed in appendix pp 9-74. For blood pressure traits, we included SNPs associated at genome-wide significance with any of the three traits in all analyses. For body mass index, we used the set of independent SNPs provided by study authors. [41] We calculated the ratio of the SNP risk factor effect size by the corresponding effect size for lacunar stroke and aggregated effects across all risk factor-associated SNPs using an inverse-variance weighted procedure. As secondary analyses, we used median, weighted median and MR-Egger approaches to aggregate across SNPs. We used the MR-Egger intercept to assess evidence of directional pleiotropy. In all analyses we used the MendelianRandomization package in R.[45] Results are presented as odds ratios per genetically predicted increase in each risk factor (original scale).

Role of the Funding Source

The funder had no role in the design or execution of the study, interpretation of data, writing of the report, or the decision to submit the paper for publication. Study authors had full access to data included in the study and accept responsibility to submit for publication.

Results

We meta-analysed studies from Europe, United States, and Australia, giving 6,030 cases and 248,929 controls of European ancestry, and 7,338 cases and 225,258 controls in Transethnic analysis. 2,987 (40.7%) cases (matched with 29,540 controls) had confirmation by MRI. Study cohorts, including genomic inflation λ values, are described in the appendix (pp 5-6, 100-111). Following meta-analysis, in the European analysis LDSCORE intercept values were equal to 1.046 (s.e = 0.008) and the λ_{1000} value was 1.007, while in the Transethnic analysis the λ_{1000} value was 1.005, indicating no substantial inflation. SNP-heritability of MRI confirmed lacunar stroke, calculated using GREML methods in a European ancestry subset of 1,693 cases and 10,171 controls genotyped on the same array, was h^2=0.17-0.21, standard error=0.02 assuming stroke prevalence of 1-3%, and that 20% of these are lacunar strokes. Using LD Score regression estimates of SNP-heritability were lower then GREML estimates, but were higher in the MRI confirmed (h^2=0.065, s.e=0.017) than in the non MRI confirmed (h^2=0.0081, s.e=0.0025). The genetic correlation between MRI confirmed and non MRI confirmed groups using LD Score regression was significant (rg=0.61, s.e=0.21, p=0.0033).

Three loci were associated with lacunar stroke in European samples, while 3 were associated in Transethnic analysis, giving 5 loci overall (Table 1, Figure 2). Regional association plots and forest plots for these loci are provided in the appendix (pp 115-134). Four of the loci were novel, while one was the previously associated 16q24 locus. [6] One other locus (*ICA1L-WDR12-CARF-NBEAL1*) was associated in gene-based analyses in MEGASTROKE, and was associated in a recent multi-trait analysis of intracerebral haemorrhage and lacunar stroke. [5,46]

We next applied MTAG to identify additional genetic variation underlying lacunar stroke in a joint analysis with an etiologically related trait, cerebral WMH. Genetic correlation between lacunar stroke and cerebral WMH, calculated using LDSCORE regression, [21] was substantial for the MRI confirmed group (rG(SE)=0.46(0.10) p=4.6×10^{-6}) and slightly lower when including all lacunar strokes (rG(SE)=0.37(0.09) p=4.0×10^{-5}). In the joint analysis with cerebral WMH, variants in seven additional loci reached genome-wide significance for lacunar stroke overall (Table 1, Figure 2). Four of these loci (SLC25A44-PMF1-BGLAP, LOX-ZNF474-LOC100505841, SH3PXD2A, COL4A2) have previously been associated with WMH. [47,48] Regional association plots and Forest plots for the loci are provided in the appendix (pp 120-134).

None of the 12 loci reaching genomewide significance showed evidence of heterogeneity (p=0.05 to 0.98; appendix pp 123-134). In two regions (*SH3PXD2A and HTRA1-ARMS2)* multiple apparently independent (LD r^2<0.1) SNPs reached genome-wide significance. However, in a joint modelling scenario employed using GCTA-cojo, only a single SNP at each of these regions remained genome-wide significant showing that a single variant remains the most parsimonious explanation of the association at this locus. [22] We discarded two regions according to our protocol. One region on chromosome 17q25 showed an association solely with WMH, with no association with lacunar stroke (lead SNP p=0.39). A second region on chromosome 14 (*EVL-DEGS2*), was not as significant in MTAG analysis ($p=1.2x10^{-9}$) as in WMH alone ($p=1.2x10^{-12}$) so an independent contribution of lacunar stroke to the association could not be determined. Further evidence is required to determine that these regions are associated with lacunar stroke, so each was discarded from this analysis. The *ZBTB14-EPB41L3* locus that was associated with lacunar stroke was not associated with WMH (p=0.33 and effect in the opposite direction). Similarly, for the *ULK4* locus associated with lacunar stroke, the lead SNP did not reach significance for WMH (p=0.12), but was in the consistent effect direction and thus could reflect a lack of study power.

The 12 loci showed stronger effects in the MRI confirmed group compared to the standard phenotyping group (in European ancestry analysis, appendix pp 86), although not significantly so (one-tailed p-value=0.07), with a median proportional increase in odds ratio of 3.4%. The 12 loci explained 1.4% of the overall heritability, and 6.5-8.1% of the lacunar stroke heritability from GWAS arrays, as calculated in this study.

We performed a TWAS to identify genes for which expression was associated with lacunar stroke (Figure 3). Genetically elevated *SLC25A44* was associated with lacunar stroke in multi-trait analysis in arterial tissues, while genetically decreased *ULK4* was associated with lacunar stroke in arterial tissues, whole blood, and brain. At the 2q33.2 locus, genetically elevated *CARF*, *FAM117B*, *ICA1L*, and *NBEAL1* were all associated with lacunar stroke. All associations were confirmed by colocalization analysis between the gene expression and lacunar stroke associations (posterior probability (pp)>0.7). Five other associations were identified in the TWAS, but were not confirmed by colocalization analysis (pp<0.7, Figure 3).

We used Phenoscanner to interrogate whether the 12 lead SNPs were associated with DNA methylation, metabolite or protein levels from large scale studies. [49-51] Eleven of the 12 lead SNPs showed associations with DNA methylation at genome-wide significance: more than expected by chance based on randomly selected SNPs across the genome ($p<0.01$), and 10 of which were associated in multiple independent studies (appendix pp 87-97). Conversely, none of the 12 SNPs were associated with metabolite or protein levels.

Querying databases that catalogue drug-gene relationships showed that 11 of the genes listed in Table 1 are categorized as 'druggable' indicating they have potential for therapeutic development (appendix pp 98). However, no existing drugs target any of the genes identified in this study.

A pathway analysis based on the multi-trait analysis results using MAGMA revealed five significantly associated Gene Ontology gene sets: Phosphatidylinositol 5 Phosphate Binding ($p=2.2x10^{-6}$, FDR=0.020), Extracellular Matrix Structural Constituent ($p=6.2x10^{-6}$, FDR=0.027), Extracellular Matrix Constituent Conferring Elasticity ($p=8.9x10^{-6}$, FDR=0.027), Middle Ear Morphogenesis ($p=2.3x10^{-5}$, FDR=0.049), and Roundabout (ROBO) Binding ($p=2.7x10^{-5}$, FDR=0.049). No pathways were significant when based solely on lacunar stroke results. Results for all pathways with FDR<0.5 are presented in the appendix (pp 75-76).

Mendelian randomization analyses using an inverse variance weighted approach found positive associations with diastolic, systolic, and pulse pressure, type 2 diabetes, and ever smoking with lacunar stroke (Figure 4). No significant finding showed any evidence of pleiotropy, as assessed using the MR-Egger intercept. There was evidence not reaching Bonferroni corrected significance, for a protective effect of increased high-density lipoprotein on risk of lacunar stroke. There was no evidence of association with body mass index, low density lipoprotein or triglycerides. Secondary analysis for all risk factors using median, weighted median, and MR-Egger approaches are presented in the appendix (pp 135-143).

GWAS Summary statistics from the primary analyses are available at GWAS Catalog (https://www.ebi.ac.uk/gwas/summary-statistics) and on the Cerebrovascular Portal (http://www.cerebrovascularportal.org).

Discussion

Despite its public health importance as the cause of a quarter of all strokes, previous GWAS studies have only identified one genetic locus for lacunar stroke, in contrast to the 35 identified for ischaemic stroke and its major subtypes.[5] We performed a GWAS of lacunar stroke, including the largest number of cases with MRI confirmation to date, identifying 11 novel loci in addition to replicating the only previously reported locus.

The primary analysis identified four novel loci. One association on chromosome 11, encompassing *SPI1-SLC39A13-PSMC3-RAPSN* was identified in both European and Transethnic analyses. The lead SNP is a synonymous variant in *SLC39A13* (Solute Carrier Family 39 Member 13), a transmembrane protein with roles in zinc transport. Mutations in this gene cause a form of Ehlers-Danlos syndrome, a group of connective tissues disorders which influence the vasculature and can cause stroke; [52] vascular abnormalities have been reported in *SLC39A13* knockout mice.[53] We additionally identified a locus for which the lead SNP resides in an intron of *ULK4* (UNC-51 Like Kinase 4) on chromosome 3. The TWAS analysis suggests *ULK4* is the most likely implicated gene, with genetically decreased expression of *ULK4* being associated with lacunar stroke. Variants in close LD with the lead SNP have been implicated in diastolic blood pressure in large scale GWAS. [54] However, the direction of effect was opposite to that for lacunar stroke, suggesting this likely reflects pleiotropy rather than a causal pathway. Variants in close LD have also been associated with another cardiovascular disease, Acute Aortic Dissection. [55] *ULK4* belongs to the family of serine/threonine protein kinases, a group of phosphorylating kinases involved in diverse processes including cell proliferation and differentiation, apoptosis and embryonic development. Its deficiency leads to hypomyelination, [56] and it has been associated with neuropsychiatric traits. [57] Finally, we report a novel association on chromosome 18, located between *ZBTB14* (Zinc Finger and BTB Domain Containing 14), a zinc finger transcription factor, and *EPB41L3* (Erythrocyte Membrane Protein Band 4.1 Like 3), a membrane protein that inhibits cell proliferation and promotes apoptosis.

In multi-trait analysis we identified 7 further loci, all of which are reported as associated with lacunar stroke at genome-wide significance for the first time. Two have not been reported as being associated with any cerebrovascular disease previously. One lies in an intergenic region between the *VTA1*

(Vesicle Trafficking 1) and *GPR126* (G Protein-Coupled Receptor 126) genes. *GPR126* is a G-Protein Coupled Receptor which is activated by type IV collagen and has an important role in myelination. [58] *GPR126* binds laminin-211, [59] an extracellular matrix protein produced by astrocytes and present in the brain, with key roles in development and function of the blood-brain barrier, [60] in part through regulation of pericyte differentiation – a mechanism previously implicated through the *FOXF2* gene. [61,62] SVD-related endothelial dysfunction has also been shown to prevent oligodendrocyte precursor cell maturation, contributing to impaired myelination. [63] One hypothesis is that the *GPR126* variant might exacerbate this process, inhibiting repair from myelin damage. The second previously unreported association lies in an intergenic region, the nearest gene to which is *HTRA1* (HtrA Serine Peptidase 1), a gene in which rare homozygous variation leads to Cerebral Autosomal Recessive Arteriopathy with Subcortical Infarcts and Leucoencephalopathy (CARASIL). [64] *HTRA1*, through processing of LTBP-1 (latent transforming growth factor beta binding protein 1), promotes transforming growth factor beta (TGF-beta) signaling in the vascular extracellular matrix (ECM). [65] The presence of both rare and common risk variants in *HTRA1* points to it being a key molecule in lacunar stroke pathogenesis, and is a feature shared with another gene identified in this study, *COL4A2*, in which rare variants also cause monogenic forms of cerebral small vessel disease.[7] Candidate gene studies have previously shown associations not reaching genome-wide significance in *COL4A2* with lacunar stroke and the same region has also previously been associated with intracerebral haemorrhage in multi-trait analysis, and coronary artery disease. [46,66,67] Four other loci identified (*SH3PXD2A, LOX-ZNF474-LOC100505841, SLC25A44-PMF1-BGLAP, FOXF2-FOXQ1*) were associated with broad stroke in MEGASTROKE (see appendix pp 11-18 for associations of all SNPs in MEGASTROKE) or a previous meta-analysis, [5,62] although this is the first study to confirm their association specifically with lacunar stroke. At the *SLC25A44-PMF1-BGLAP* locus, the TWAS results point to an association of genetically elevated *SLC25A44* with lacunar stroke, which was validated in colocalization analysis. *SLC25A44* plays a key role in catabolism of branched-chain amino acids in brown adipose tissue by transporting them into mitochondria, [68] and thus has potential as a mediating factor in the relationship between metabolic disease and lacunar stroke. However, variants in close LD have also been associated with mosaic Y chromosome loss, [69] highlighting mosaicism as an alternative mechanism. Further functional studies will be required to untangle these relationships with lacunar stroke. The strength of association of all

associated variants was moderate to large in the context of GWAS (OR ranging from 1.10 to 1.25 in Europeans) and notably larger than effects previously reported for variants associated with broad stroke phenotypes. [5] This is consistent with the variants acting specifically on the lacunar stroke subtype rather than on stroke as a whole.

We also found that 11 of the 12 lead SNPs influence DNA methylation at genome-wide significance, pointing to epigenetic changes being one source through which risk of lacunar stroke is conferred. Whether this genetically-altered DNA methylation influences transcription of nearby genes – and which genes are affected – should be the focus of further study. A pathway analysis implicated several biological processes in lacunar stroke pathophysiology. Two pathways involved the ECM, the network of extracellular molecules that provide scaffolding and biochemical support to surrounding tissues. Disruption of the vascular ECM has been hypothesized to be a key component in pathogenesis of CSVD, particularly in monogenic forms, and several of the genes implicated in this study (*COL4A2, LOX, SH3PXD2A, GPR126, HTRA1*) play a key role in the ECM. [70] This finding lends support to this hypothesis and suggests ECM dysfunction also has a key role in sporadic CSVD.

We performed Mendelian randomization to assess whether cardiovascular risk factors showed evidence of causal association with lacunar stroke. We found support for genetically predicted elevated blood pressure (systolic, diastolic, and pulse pressure), type 2 diabetes and smoking being associated with lacunar stroke. The results are consistent with those from observational studies, and suggest that targeting these factors would reduce risk of lacunar stroke. [71] There was evidence not reaching Bonferroni corrected significance for a protective effect of increased high-density lipoprotein on risk of lacunar stroke, and no association with low-density lipoprotein, replicating findings in previous studies. [72,73] Overall these finding show that the impact of the direct effects of low-density-lipoprotein lowering medications such as statins on incidence of lacunar stroke is likely to be minimal.

Our study emphasises the benefit of accurate phenotyping using MRI. Using this approach, the heritability of lacunar stroke using GREML was substantial, and larger than previous estimates based on TOAST subtyping. [74] Using LD Score, the heritability was larger in the MRI confirmed group, but estimates were considerably lower than for GREML. The use of MRI subtyping also increased the strength of association of the lacunar stroke associated variants although this increase was not quite

significant. These results suggest that further genetic risk factor studies in lacunar stroke are likely to be more successful if MRI subtyping is used.

Our study has limitations. The analysis was performed in a predominantly European ancestry population. Large studies including diverse ancestries should be performed to assess the generalizability of findings to all ethnic groups. The MTAG approach relies on the relatively strong assumption that associated variants act on both traits, which may not always be the case for WMH and lacunar stroke, as they reflect downstream effects of a shared common ancestor, SVD. To control for this, we only considered SNPs showing association with both traits and showing greater significance in MTAG analysis than with WMH or lacunar stroke alone, as being significant. However independent replication will remain the gold standard for confirming these and all other reported associations in this article. Recent studies have suggested that a more conservative threshold of $p<1x10^{-8}$ should be considered in GWAS using larger imputation panels such as here. [75] If using this threshold one locus (*ZBTB14-EPB41L3*) would no longer be significant. Additional caution should therefore be applied when interpreting this finding, particularly as it was not significant in MTAG analysis. To increase sample size and study power, we used publicly available controls in analyses. As such it was not possible to determine whether these individuals had a history of lacunar stroke. Our analyses did not adjust for age and sex; there is an ongoing debate about the importance of including such covariates in genetic studies. [76] In analyses with substantial differences between case and control populations it is possible that this could result in subtle biases.

In summary, these findings represent substantial progress in identifying the genetic mechanisms underlying lacunar stroke, a disease for which there remain significant deficits in our understanding of the molecular causes. Our findings highlight diverse mechanisms contributing to the disease, implicating disruption of the vascular ECM *(COL4A2, LOX, SH3PXD2A, GPR126, HTRA1*), pericyte differentiation (*FOXF2, GPR126*), TGF-beta signaling *(HTRA1),* and myelination (*ULK4, GPR126*) in disease risk. This provides novel insights into the pathogenesis of lacunar stroke, and highlights multiple candidates to take forward into functional experiments to identify specific mechanisms conferring risk of lacunar stroke which could be targeted therapeutically.

Acknowledgements

The UK Household Longitudinal Study is led by the Institute for Social and Economic Research at the University of Essex and funded by the Economic and Social Research Council. The survey was conducted by NatCen and the genome-wide scan data were analysed and deposited by the Wellcome Trust Sanger Institute. Information on how to access the data can be found on the Understanding Society website https://www.understandingsociety.ac.uk/. This research made use of the UK Biobank Resource under application number 36509. Ethical approval for UK Biobank was received from the research ethics committee (REC reference 11/NW/0382). We are grateful to deCODE genetics for providing data for this analysis. We acknowledge the contribution of Giorgio Boncoraglio.

Funding

This work, including collection and genotyping of the UK Young Lacunar Stroke DNA Study 2 (DNA Lacunar 2), was supported by a British Heart Foundation Programme Grant (RG/16/4/32218). Matthew Traylor was supported by The Barts Charity and The National Institute of Health Research Barts Biomedical Research Centre. The National Institute of Neurological Disorders and Stroke – Stroke Genetics Network (NINDS-SIGN) study was funded by the US National Institute of Neurological Disorders and Stroke, National Institutes of Health (U01 NS069208 and R01 NS100178). Collection of the UK Young Lacunar Stroke DNA Study 1 (DNA Lacunar) was primarily supported by the Wellcome Trust (WT072952) with additional support from the Stroke Association (TSA 2010/01). Genotyping of the DNA Lacunar samples was supported by a Stroke Association Grant (TSA 2013/01). The principal funding for the WTCCC2 stroke study was provided by the Wellcome Trust, as part of the Wellcome Trust Case Control Consortium 2 project (085475/B/08/Z and 085475/Z/08/Z and WT084724MA). Hugh Markus is supported by a National Institute for Health Research (NIHR) Senior Investigator award, and his work is supported by the Cambridge Universities NIHR Comprehensive Biomedical Research Centre. CML is supported by the NIHR Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London. Dr Anderson is supported by NIH R01NS103924 and K23NS086873. Professor Rothwell and the Oxford Vascular Study are funded by the NIHR Oxford Biomedical Research Centre and by the Wellcome Trust. Collection and genotyping

of the Sahlgrenska Academy Study on Ischemic Stroke were primarily supported by the Swedish Research Council (grant #2018-02543), the Swedish Heart and Lung Foundation (20190203), and the Swedish State under the agreement between the Swedish government and the county councils (the ALF-agreement, ALFGBG-720081). The genetic data from Geisinger was made available through the collaboration with Regeneron Genetic Centre. Dr Lemmens is a senior clinical investigator of FWO Flanders. Edinburgh Mild Stroke Study (MSS2) was funded by the Wellcome Trust (WT088134/Z/09/A)

Contributors

HSM, MT, CML designed the experiment. MT performed the meta-analysis and downstream analyses. HSM provided oversight. EP analysed the TWAS. YZ, NT, MKB, VA, SK, LT, SB performed statistical analysis. HSM, JW, CDA, PMR, RL, NR, YR, RZ, IF-C, DS, CJ, MD, JR, SJK, BDM, SM, KC, CK, AH, VS, JFM, AL, RPG, TR, JJ-C, PS, AP, LL, CJG, LRJ, DJT were responsible for recruitment and phenotyping of cohorts. All authors contributed to, reviewed, and approved the final draft of the paper.

Declaration of Interests

CA reports grants from National Institutes of Health of the U.S., grants from American Heart Association, grants from Massachusetts General Hospital, grants from Bayer AG, personal fees from ApoPharma, Inc., outside the submitted work; Dr. Bell reports grants from British Heart Foundation, during the conduct of the study; AH reports grants from Academy of Finland, outside the submitted work; AL reports grants from Swedish Heart and Lung Foundation, grants from Region Skåne, grants from Skåne University Hospital, grants from Freemasons Lodge of Instruction Eos in Lund, grants from Lund University, grants from The Foundation of Färs & Frosta - one of Sparbanken Skåne's ownership Foundations, grants from Swedish Research Council, during the conduct of the study; personal fees from Astra Zeneca, personal fees from BMS/Pfizer, personal fees from Portola, personal fees from Bayer, outside the submitted work; HSM reports grants from British Heart Foundation, during the conduct of the study; personal fees from BIBA, outside the submitted work; JR reports grants from National Institutes of Health, grants from OneMind, during the conduct of the study; personal fees from Boehringer Ingelheim, personal fees from Pfizer, personal fees from New Beta Innovation, outside the submitted work; PR reports personal fees from Bayer, personal fees from BMS, outside the submitted

work; VS reports other from Novo Nordisk, other from Sanofi, grants from Bayer Ltd, outside the submitted work; JW reports grants from Wellcome Trust, during the conduct of the study. All other authors declare no competing interesting. This information was verified by MT and HSM.

Data Sharing

GWAS Summary statistics from these analyses are available at GWAS Catalog (https://www.ebi.ac.uk/gwas/summary-statistics) and on the Cerebrovascular Portal (http://www.cerebrovascularportal.org). Individual level data from the NINDS-SIGN Stroke study are available to researchers through dbGAP: https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000615.v1.p1.

Research in context

Evidence before this study

Using the terms “stroke”, “small vessel stroke”, “lacunar stroke”, “small vessel disease”, “white matter hyperintensities”, “genetics”, “GWAS”, we searched PubMed (https://pubmed.ncbi.nlm.nih.gov) and GWAS Catalog (https://www.ebi.ac.uk/gwas/) for relevant reports. We only considered peer-reviewed reports in English. Only a single locus on chromosome 16q24 has been robustly associated specifically with lacunar stroke compared to over 30 with broad stroke phenotypes.

Added value of this study

The present findings substantially expand the number of genetic associations with lacunar stroke, with 5 loci now associated directly and a further 7 associated with lacunar stroke jointly with white matter hyperintensities. These loci highlight several key mechanisms in lacunar stroke pathogenesis - including extracellular matrix dysfunction, myelination, and pericyte differentiation. The current findings also show that individuals with increased genetic predisposition to elevated blood pressure, history of smoking, and type 2 diabetes are at increased risk of lacunar stroke, pointing to the causal role of these factors in disease etiology.

Implications of all the available evidence

There currently are no treatments that prevent lacunar stroke aside from management of vascular risk factors such as blood pressure. This is due in part to lack of understanding of mechanisms underlying the disease. The present findings highlight novel mechanisms underlying lacunar stroke pathogenesis, and therefore point to pathways which have potential to be targeted by therapeutics. Improved treatment of elevated blood pressure and type 2 diabetes, as well as prevention of smoking in the population will likely reduce the burden of lacunar stroke.

Table 1. Genome-wide Significant Loci for Lacunar Stroke in (A.) Univariate or (B.) Multi-trait Analysis.

							Lacunar Stroke (European: 6030 Cases, 219,389 Controls)			Lacunar Stroke (Transethnic: 7338 Cases, 225,258 Controls)		White Matter Hyperintensities (N=42,310)		MTAG
Nearest Genes	Chr	BP	Genomic Context	rsid	RA/ OA	RAF	OR(SE)	P value	N Studies	OR(SE)	P value	Beta(SE)	P value	P value
A.														
ICA1L*-WDR12-****CARF****-****NBEAL1***	2	203,968,973	Intronic	rs72934535	T/C	0.89	1.25(0.04)	$3.7x10^{-9}$	12	1.22(0.04)	$5.2x10^{-8}$	0.070(0.01)	$2.8x10^{-10}$	$5.3x10^{-16}$
ULK4	3	41,839,370	Intronic	rs4621303	T/A	0.83	1.15(0.03)	$1.7x10^{-7}$	14	1.16(0.03)	$6.4x10^{-9}$	0.015(0.01)	0.12	$2.2 x10^{-7*}$
SPI1-SLC39A13-PSMC3-RAPSN	11	47,434,986	Exonic	rs2293576	G/A	0.67	1.14(0.02)	$7.2x10^{-10}$	14	1.14(0.02)	$6.0x10^{-10}$	0.030(0.01)	$3.1x10^{-5}$	$6.4x10^{-13}$
ZCCHC14	16	87,575,332	Intergenic	rs12445022	A/G	0.34	1.13(0.02)	$2.5x10^{-8}$	13	1.12(0.02)	$9.0x10^{-8}$	0.019(0.01)	0.0078	$3.1x10^{-9}$
ZBTB14-EPB41L3	18	5,389,832	Intergenic	rs9958650	G/A	0.10	1.18(0.03)	$9.9x10^{-7}$	12	1.19(0.03)	$2.4x10^{-8}$	-0.011(0.01)	0.33	0.0005
B.														
SLC25A44*-PMF1-BGLAP*	1	156,197,380	Intronic	rs2984613	C/T	0.64	1.10(0.02)	$2.5x10^{-5}$	13	1.09(0.02)	$1.4x10^{-5}$	0.037(0.01)	$2.3x10^{-7}$	$8.2x10^{-10}$
LOX-ZNF474-LOC100505841	5	121,518,378	Downstream	rs2303655	T/C	0.81	1.14(0.03)	$3.6x10^{-5}$	11	1.12(0.03)	0.00014	0.050(0.01)	$1.4x10^{-8}$	$1.9x10^{-10}$
FOXF2-FOXQ1	6	1,366,718	Intergenic	rs7766042	C/T	0.11	1.17(0.03)	$3.7x10^{-6}$	11	1.18(0.03)	$1.2x10^{-6}$	0.045(0.01)	$7.1x10^{-5}$	$5.2x10^{-9}$
VTA1-GPR126	6	142,562,417	Intergenic	rs225744	C/T	0.77	1.11(0.03)	$3.5x10^{-5}$	12	1.09(0.02)	0.00050	0.037(0.01)	$5.8x10^{-6}$	$9.2x10^{-9}$
SH3PXD2A	10	105,447,838	Intronic	rs61000833	T/C	0.60	1.10(0.02)	$1.7x10^{-5}$	12	1.07(0.02)	0.0024	0.049(0.01)	$2.0x10^{-12}$	$6.0x10^{-13}$
HTRA1-ARMS2	10	124,233,181	Intronic	rs79043147	T/C	0.07	1.21(0.04)	$3.2x10^{-6}$	11	1.22(0.04)	$1.1x10^{-6}$	0.057(0.01)	$1.8x10^{-5}$	$1.6x10^{-9}$
COL4A2	13	111,040,681	Intronic	rs11838776	A/G	0.29	1.11(0.02)	$4.3x10^{-6}$	12	1.11(0.02)	$1.6x10^{-6}$	0.050(0.01)	$7.9x10^{-11}$	$7.9x10^{-13}$

A, Associations reaching genome-wide significance for Lacunar Stroke; B, Associations reaching genome-wide significance in multi-trait analysis. Chr, chromosome; BP, base position (hg19); RA, risk allele; OA, other allele; RAF, risk allele frequency; MTAG, multi-trait analysis of GWAS; logBF, log (Bayes factor); *, As A/T and C/G SNPs are removed by MTAG, results are presented for SNP in highest LD (rs9842261); Genes in bold type were associated in TWAS analysis and confirmed by colocalization.

Figure 1. Analysis pipeline

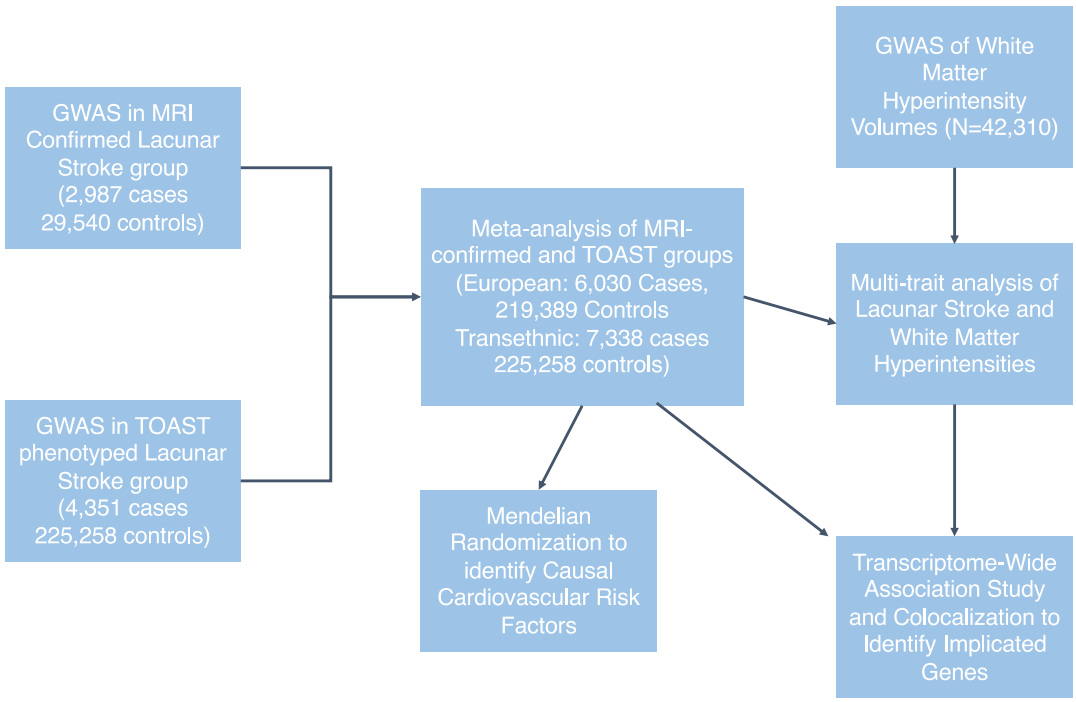


Figure 2. Manhattan plot of -log10(p-values) for genomewide SNP associations with A) lacunar stroke (transethnic analysis) and B) lacunar stroke multi-trait analysis, by genomic position.

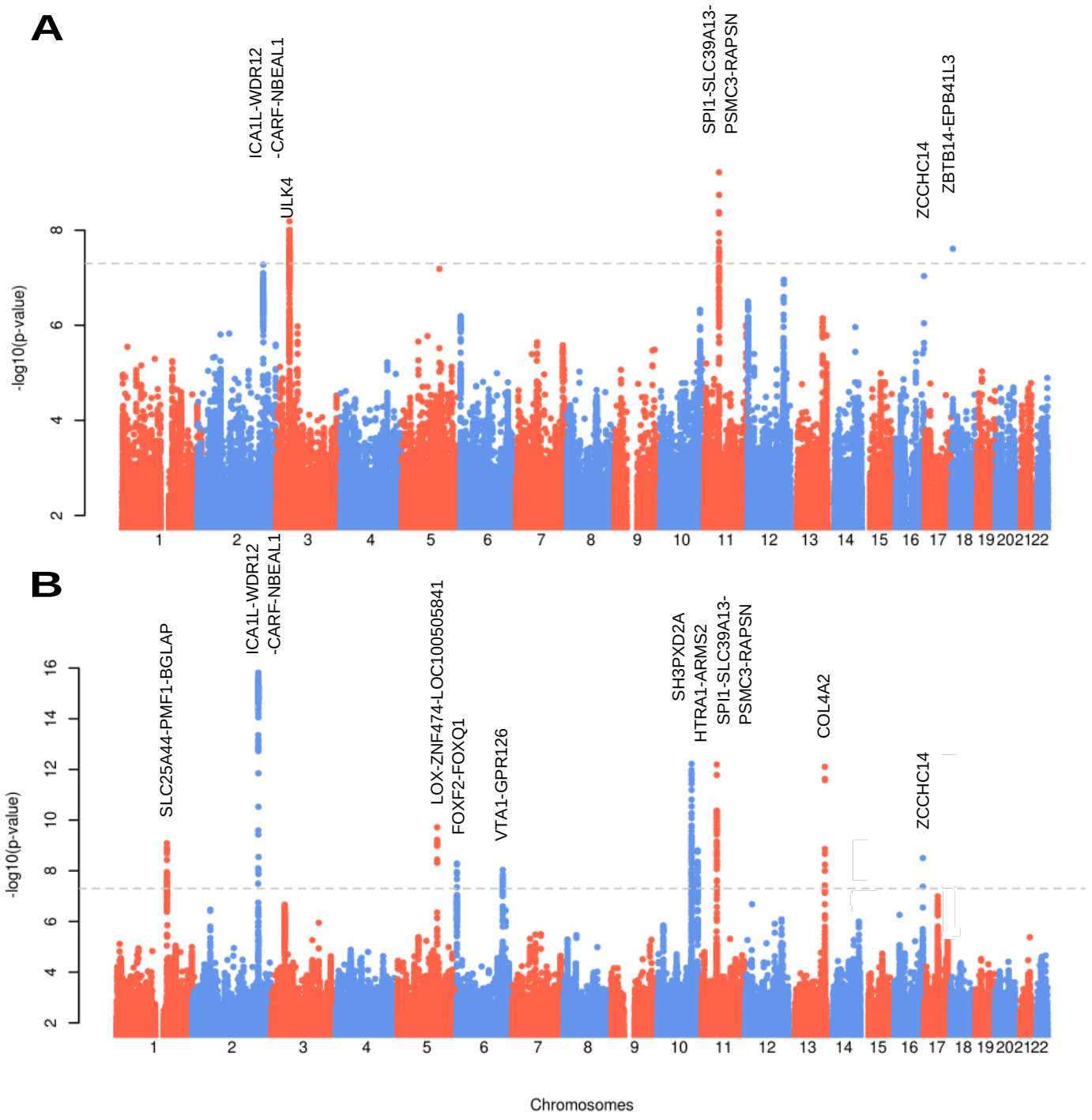
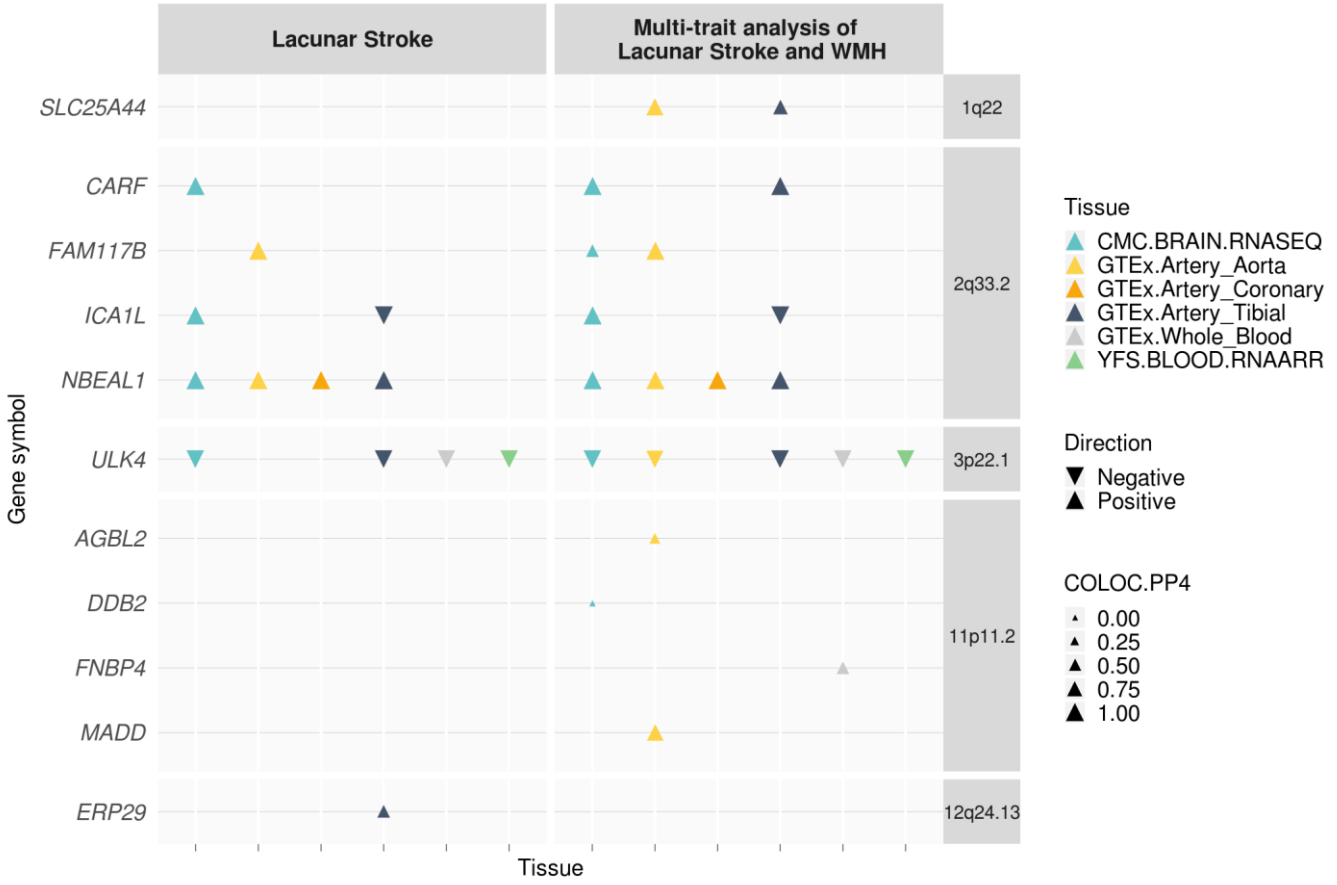
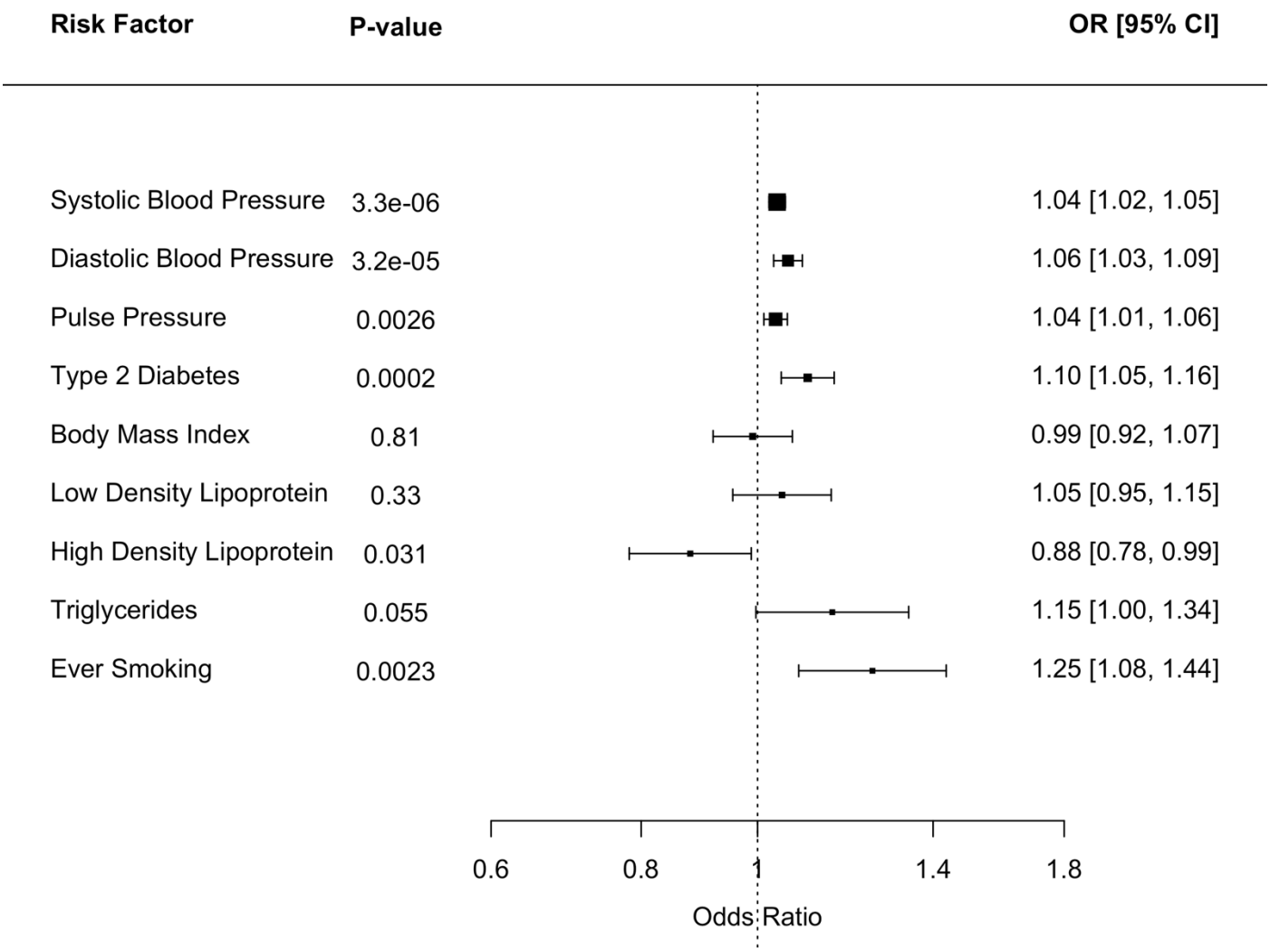


Figure 3. Genes for which expression is associated with lacunar stroke in 6 tissues from Transcriptome-Wide Association Analysis, with evidence of colocalization of gene expression and lacunar stroke signals given by triangle size



COLOC.PP4, the posterior probability of hypothesis 4 in colocalization analysis, that there is a consistent association between lacunar stroke and expression of the given gene.

Figure 4. Odds ratios for associations between genetically proxied cardiovascular risk factors and lacunar stroke from Mendelian Randomization analysis using the inverse variance weighted method



Estimates are presented as odds ratios per genetically proxied increase in each risk factor (original scale); OR, odds ratio.

References

1. Pantoni L. Cerebral small vessel disease: from pathogenesis and clinical characteristics to therapeutic challenges. *Lancet Neurol* 2010; **9**(7): 689-701.
2. Wardlaw JM, Smith C, Dichgans M. Small vessel disease: mechanisms and clinical implications. *Lancet Neurol* 2019; **18**(7): 684-96.
3. 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; **12**(8): 822-38.
4. Malik R, Rannikmae K, Traylor M, et al. Genome-wide meta-analysis identifies 3 novel loci associated with stroke. *Ann Neurol* 2018; **84**(6): 934-9.
5. Malik R, Chauhan G, Traylor M, et al. Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nat Genet* 2018; **50**(4): 524-37.
6. Traylor M, Malik R, Nalls MA, et al. Genetic variation at 16q24.2 is associated with small vessel stroke. *Ann Neurol* 2017; **81**(3): 383-94.
7. Tan R, Traylor M, Rutten-Jacobs L, Markus H. New insights into mechanisms of small vessel disease stroke from genetics. *Clinical science (London, England : 1979)* 2017; **131**(7): 515-31.
8. Jerrard-Dunne P, Cloud G, Hassan A, Markus HS. Evaluating the genetic component of ischemic stroke subtypes: a family history study. *Stroke* 2003; **34**(6): 1364-9.
9. Persyn E. The genetic basis of MRI markers of cerebral small vessel disease; a Genome Wide Association Study in 42,310 participants. *Under Review*.
10. Adams HP, Jr., Bendixen BH, Kappelle LJ, et al. Classification of subtype of acute ischemic stroke. Definitions for use in a multicenter clinical trial. TOAST. Trial of Org 10172 in Acute Stroke Treatment. *Stroke* 1993; **24**(1): 35-41.
11. 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; **44**(3): 328-33.
12. Loci associated with ischaemic stroke and its subtypes (SiGN): a genome-wide association study. *Lancet Neurol* 2015.
13. McCarthy S, Das S, Kretzschmar W, et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet* 2016; **48**(10): 1279-83.
14. Rajajee V, Kidwell C, Starkman S, et al. Diagnosis of lacunar infarcts within 6 hours of onset by clinical and CT criteria versus MRI. *J Neuroimaging* 2008; **18**(1): 66-72.
15. 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; **116**(19): 2157-64.
16. Bamford J, Sandercock P, Dennis M, Burn J, Warlow C. Classification and natural history of clinically identifiable subtypes of cerebral infarction. *Lancet* 1991; **337**(8756): 1521-6.
17. Auton A, Brooks LD, Durbin RM, et al. A global reference for human genetic variation. *Nature* 2015; **526**(7571): 68-74.
18. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 2010; **26**(17): 2190-1.
19. Winkler TW, Day FR, Croteau-Chonka DC, et al. Quality control and conduct of genome-wide association meta-analyses. *Nature protocols* 2014; **9**(5): 1192-212.
20. Devlin B, Roeder K. Genomic control for association studies. *Biometrics* 1999; **55**(4): 997-1004.
21. Bulik-Sullivan BK, Loh PR, Finucane HK, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* 2015; **47**(3): 291-5.
22. Yang J, Ferreira T, Morris AP, et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet* 2012; **44**(4): 369-75, s1-3.
23. Nagelkerke NJ. A note on a general definition of the coefficient of determination. *Biometrika* 1991; **78**(3): 691-2.
24. Lee SH, DeCandia TR, Ripke S, et al. Estimating the proportion of variation in susceptibility to schizophrenia captured by common SNPs. *Nat Genet* 2012; **44**(3): 247-50.
25. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* 2011; **88**(1): 76-82.
26. Turley P, Walters RK, Maghzian O, et al. Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nat Genet* 2018; **50**(2): 229-37.
27. Turchin MC, Stephens M. Bayesian multivariate reanalysis of large genetic studies identifies many new associations. *PLoS Genet* 2019; **15**(10): e1008431.
28. Gusev A, Ko A, Shi H, et al. Integrative approaches for large-scale transcriptome-wide association studies. *Nat Genet* 2016; **48**(3): 245-52.

29. Battle A, Brown CD, Engelhardt BE, Montgomery SB. Genetic effects on gene expression across human tissues. *Nature* 2017; **550**(7675): 204-13.
30. Fromer M, Roussos P, Sieberts SK, et al. Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nat Neurosci* 2016; **19**(11): 1442-53.
31. Raitakari OT, Juonala M, Ronnemaa T, et al. Cohort profile: the cardiovascular risk in Young Finns Study. *Int J Epidemiol* 2008; **37**(6): 1220-6.
32. Giambartolomei C, Vukcevic D, Schadt EE, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet* 2014; **10**(5): e1004383.
33. Staley JR, Blackshaw J, Kamat MA, et al. PhenoScanner: a database of human genotype-phenotype associations. *Bioinformatics* 2016; **32**(20): 3207-9.
34. Kamat MA, Blackshaw JA, Young R, et al. PhenoScanner V2: an expanded tool for searching human genotype-phenotype associations. *Bioinformatics* 2019; **35**(22): 4851-3.
35. Wishart DS, Feunang YD, Guo AC, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res* 2018; **46**(D1): D1074-d82.
36. Cotto KC, Wagner AH, Feng YY, et al. DGIdb 3.0: a redesign and expansion of the drug-gene interaction database. *Nucleic Acids Res* 2018; **46**(D1): D1068-d73.
37. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol* 2015; **11**(4): e1004219.
38. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005; **102**(43): 15545-50.
39. Teslovich TM, Musunuru K, Smith AV, et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* 2010; **466**(7307): 707-13.
40. Hoffmann TJ, Ehret GB, Nandakumar P, et al. Genome-wide association analyses using electronic health records identify new loci influencing blood pressure variation. *Nat Genet* 2017; **49**(1): 54-64.
41. Yengo L, Sidorenko J, Kemper KE, et al. Meta-analysis of genome-wide association studies for height and body mass index in ~700,000 individuals of European ancestry. *bioRxiv* 2018: 274654.
42. Mahajan A, Taliun D, Thurner M, et al. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat Genet* 2018; **50**(11): 1505-13.
43. Liu M, Jiang Y, Wedow R, et al. Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nat Genet* 2019; **51**(2): 237-44.
44. Haycock PC, Burgess S, Wade KH, Bowden J, Relton C, Davey Smith G. Best (but oft-forgotten) practices: the design, analysis, and interpretation of Mendelian randomization studies. *Am J Clin Nutr* 2016.
45. Yavorska OO, Burgess S. MendelianRandomization: an R package for performing Mendelian randomization analyses using summarized data. *Int J Epidemiol* 2017.
46. Chung J, Marini S, Pera J, et al. Genome-wide association study of cerebral small vessel disease reveals established and novel loci. *Brain* 2019.
47. Verhaaren BF, Debette S, Bis JC, et al. Multi-Ethnic Genome-Wide Association Study of Cerebral White Matter Hyperintensities on MRI. *Circ Cardiovasc Genet* 2015; **8**(2): 398-409.
48. Persyn E, Hanscombe KB, Howson JMM, Lewis CM, Traylor M, Markus HS. Genome-wide association study of MRI markers of cerebral small vessel disease in 42,310 participants. *Nature communications* 2020; **11**(1): 2175.
49. Bonder MJ, Luijk R, Zhernakova DV, et al. Disease variants alter transcription factor levels and methylation of their binding sites. *Nat Genet* 2017; **49**(1): 131-8.
50. Gaunt TR, Shihab HA, Hemani G, et al. Systematic identification of genetic influences on methylation across the human life course. *Genome Biol* 2016; **17**: 61.
51. Chen L, Ge B, Casale FP, et al. Genetic Drivers of Epigenetic and Transcriptional Variation in Human Immune Cells. *Cell* 2016; **167**(5): 1398-414.e24.
52. Bin BH, Hojyo S, Hosaka T, et al. Molecular pathogenesis of spondylocheirodysplastic Ehlers-Danlos syndrome caused by mutant ZIP13 proteins. *EMBO Mol Med* 2014; **6**(8): 1028-42.
53. Hirose T, Shimazaki T, Takahashi N, et al. Morphometric analysis of thoracic aorta in Slc39a13/Zip13-KO mice. *Cell Tissue Res* 2019; **376**(1): 137-41.
54. Ehret GB, Munroe PB, Rice KM, et al. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* 2011; **478**(7367): 103-9.
55. Guo DC, Grove ML, Prakash SK, et al. Genetic Variants in LRP1 and ULK4 Are Associated with Acute Aortic Dissections. *Am J Hum Genet* 2016; **99**(3): 762-9.
56. Liu M, Xu P, Guan Z, et al. Ulk4 deficiency leads to hypomyelination in mice. *Glia* 2018; **66**(1): 175-90.

57. Liu M, Xu P, O'Brien T, Shen S. Multiple roles of Ulk4 in neurogenesis and brain function. *Neurogenesis (Austin)* 2017; **4**(1): e1313646.
58. Paavola KJ, Sidik H, Zuchero JB, Eckart M, Talbot WS. Type IV collagen is an activating ligand for the adhesion G protein-coupled receptor GPR126. *Science signaling* 2014; **7**(338): ra76.
59. Mehta P, Piao X. Adhesion G-protein coupled receptors and extracellular matrix proteins: Roles in myelination and glial cell development. *Dev Dyn* 2017; **246**(4): 275-84.
60. Menezes MJ, McClenahan FK, Leiton CV, Aranmolate A, Shan X, Colognato H. The extracellular matrix protein laminin alpha2 regulates the maturation and function of the blood-brain barrier. *J Neurosci* 2014; **34**(46): 15260-80.
61. Yao Y, Chen ZL, Norris EH, Strickland S. Astrocytic laminin regulates pericyte differentiation and maintains blood brain barrier integrity. *Nature communications* 2014; **5**: 3413.
62. Identification of additional risk loci for stroke and small vessel disease: a meta-analysis of genome-wide association studies. *Lancet Neurol* 2016.
63. Rajani RM, Quick S, Ruigrok SR, et al. Reversal of endothelial dysfunction reduces white matter vulnerability in cerebral small vessel disease in rats. *Science translational medicine* 2018; **10**(448).
64. Hara K, Shiga A, Fukutake T, et al. Association of HTRA1 mutations and familial ischemic cerebral small-vessel disease. *N Engl J Med* 2009; **360**(17): 1729-39.
65. Beaufort N, Scharrer E, Kremmer E, et al. Cerebral small vessel disease-related protease HtrA1 processes latent TGF-beta binding protein 1 and facilitates TGF-beta signaling. *Proc Natl Acad Sci U S A* 2014; **111**(46): 16496-501.
66. Rannikmae K, Sivakumaran V, Millar H, et al. COL4A2 is associated with lacunar ischemic stroke and deep ICH: Meta-analyses among 21,500 cases and 40,600 controls. *Neurology* 2017; **89**(17): 1829-39.
67. Nikpay M, Goel A, Won HH, et al. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat Genet* 2015; **47**(10): 1121-30.
68. Yoneshiro T, Wang Q, Tajima K, et al. BCAA catabolism in brown fat controls energy homeostasis through SLC25A44. *Nature* 2019; **572**(7771): 614-9.
69. Wright DJ, Day FR, Kerrison ND, et al. Genetic variants associated with mosaic Y chromosome loss highlight cell cycle genes and overlap with cancer susceptibility. *Nat Genet* 2017; **49**(5): 674-9.
70. Joutel A, Haddad I, Ratelade J, Nelson MT. Perturbations of the cerebrovascular matrisome: A convergent mechanism in small vessel disease of the brain? *J Cereb Blood Flow Metab* 2016; **36**(1): 143-57.
71. Rutten-Jacobs LCA, Markus HS. Vascular Risk Factor Profiles Differ Between Magnetic Resonance Imaging-Defined Subtypes of Younger-Onset Lacunar Stroke. *Stroke* 2017; **48**(9): 2405-11.
72. Hindy G, Engstrom G, Larsson SC, et al. Role of Blood Lipids in the Development of Ischemic Stroke and its Subtypes: A Mendelian Randomization Study. *Stroke* 2018; **49**(4): 820-7.
73. Georgakis MK, Malik R, Anderson CD, Parhofer KG, Hopewell JC, Dichgans M. Genetic determinants of blood lipids and cerebral small vessel disease: role of high-density lipoprotein cholesterol. *Brain* 2020; **143**(2): 597-610.
74. Bevan S, Traylor M, Adib-Samii P, et al. Genetic heritability of ischemic stroke and the contribution of previously reported candidate gene and genomewide associations. *Stroke* 2012; **43**(12): 3161-7.
75. Wu Y, Zheng Z, Visscher PM, Yang J. Quantifying the mapping precision of genome-wide association studies using whole-genome sequencing data. *Genome Biol* 2017; **18**(1): 86.
76. Pirinen M, Donnelly P, Spencer CC. Including known covariates can reduce power to detect genetic effects in case-control studies. *Nat Genet* 2012; **44**(8): 848-51.