Supplementary materials

Common variation in *COL4A1/COL4A2* is associated with sporadic cerebral small vessel disease

Materials included.

- Supplementary table e-1: Cohort characteristics and summary statistics.
- Supplementary table e-2: Original measures of association and imputation quality
    A. Original measures of association (β) and the derived odds ratios (ORs).
    B. Imputed and directly genotyped SNPs.
- Supplementary table e-3: Functional annotations.

    A. Functional annotation of SNPs in moderate LD (r^2>0.3) with rs9521732, rs9521733 and rs9515199.

    B. Functional annotation of SNPs in high LD (r^2>0.9) with rs9521732, rs9521733 and rs9515199: Regulome db database results.

- Reference list

Supplementary table e-1. Cohort characteristics and summary statistics

A. Cohorts contributing data for the ICH phenotype			
Cohort name	**GOCHA**	**ISGC**	**GERFHS**
	STUDY INFORMATION		
Study design	Prospective hospital-based case-control	3 separate prospective hospital-based case-control cohorts	Prospective population-based case-control
Study inclusion and exclusion criteria	Included: primary acute ICH cases presenting to participating centers, aged >55 years, confirmation of primary ICH on CT. Excluded: warfarin-, trauma-, brain tumor-, vascular malformation related ICH, hemorrhagic transformation of ischemic stroke, other secondary causes.		Included: spontaneous ICH in the Greater Cincinnati/ Northern Kentucky region; age ≥18 years. Excluded: trauma-, brain tumour-, vascular malformation related ICH.
ICH diagnosis criteria	Non-traumatic abrupt onset of severe headache, altered level of consciousness, and/or focal neurological deficit that is associated with a focal collection of blood within the brain parenchyma as observed on CT or at autopsy and is not due to hemorrhagic transformation of an infarction.		
Deep ICH diagnosis criteria	Involving predominantly the basal ganglia, periventricular white matter, or internal capsule, and infratentorial ICH		
Lobar ICH diagnosis criteria	Involving predominantly the cortex and underlying white matter		
Key study references	Falcone et al. Stroke 2013(1)	Woo et al. Am J Hum Genet 2014(2)	Woo et al. Stroke 2002(3)
	SUBJECTS' CHARACTERISTICS		
Ancestry of study sample	European descent		
Total number of ICH cases	298	450	797
Age of all cases with ICH (mean ± SD)	74 ±10	72 ±12	67 ±15
Number of deep ICH cases	125	279	470
Age of cases with deep ICH (mean ± SD)	71 ±13	69 ±14	65 ±16
Number of lobar ICH cases	173	171	327
Age of cases with lobar ICH (mean ± SD)	76 ±11	74 ±13	71 ±12
Number of controls	457	489	539

A. Cohorts contributing data for the ICH phenotype			
Cohort name	**GOCHA**	**ISGC**	**GERFHS**
Age of controls (mean ± SD)	72 ±8	73 ±12	66 ±15
	GENOTYPING INFORMATION		
GWAS panel used	Illumina HumanHap 610 Quad		Affymetrix 6.0
QC steps done before imputation	QC steps before imputation done in every centre, including removal of population outliers, missing data and HWE departures		
Imputation software	IMPUTE2 (https://mathgen.stats.ox.ac.uk/impute/impute_v2.html)		
Which reference imputation done to	1000-genomes (June 2011)		
Imputation quality metric reported	IMPUTE info score & PLINK info score (http://pngu.mgh.harvard.edu/~purcell/plink/)		
QC filtering applied to imputed variants	Removed SNPs with PLINK info score of <0.7 and minimum allele frequency (MAF) <1%		
Software and statistical model used for association analyzed	PLINK ICH = bo + b1*SNP + b2*age + b3*sex + b4*PC1 ... PC4		
Handling of population stratification	Principal component (PC) analysis. Remotion of population outliers by visual inspection of PC1*PC2 plot, followed by inclusion of PC1-PC4 in regression models		
Hardy Weinberg Equilibrium (HWE)	Excluded if HWE $p<1x10^{-6}$		
Adjusting for cryptic or overt relatedness	Excluded individuals with an inferred first- or second-degree relative in the sample identified on the basis of pairwise allele sharing estimates (estimated genome proportion shared identical by descent) $\pi>0.1875$		
Strand and build of the human genome on which results are provided	+ strand, build 37/hg19		

GOCHA: Genetics of Cerebral Hemorrhage with Anticoagulation; ISGC: International Stroke Genetics Consortium; GERFHS: Genetic and Environmental Risk Factors for Hemorrhagic Stroke

B. Cohorts contributing data for the ischemic stroke phenotype															
Cohort name	WTCCC 2_UK	WTCCC 2_D	ISGS & SWISS	Rotter-dam	Milano	ARIC	ASGC	Brains	CHS	deCODE	FHS	GASROS	GEOS	HPS	HVH
	STUDY INFORMATION														
Study design[1]	CS	CS	CS	PB	CS	PB	CS	CS	PB	CS	PB	CS	CS	CS	CS
Inclusion and exclusion criteria	Included: Patients with ischemic stroke who were of European ancestry from Europe, North America, and Australia, together with controls of matched ancestry. All studies used a case-control methodology.														
IS definition	Stroke was defined as a typical clinical syndrome with radiological confirmation. Stroke subtyping was done with the TOAST classification system.														
Reference	Traylor et al, Lancet Neurology 2012(4)														
	SUBJECTS' CHARACTERISTICS														
Ancestry of study sample	European ancestry (from Europe, North-America, Australia)														
Total no.cases/controls	2374/ 5175	1174/ 797	1070/ 2329	367/ 5396	372/ 407	385/ 8803	1162/ 1244	361/ 444	454/ 2817	2391/ 26970	171/ 4164	516/ 1202	448/ 498	578/ 468	566/ 1290
No. CE cases	460	330	247	-	65	93	240	29	147	399	48	169	90	-	88
No. LVD cases	498	346	229	-	74	31	421	120	-	255	-	95	37	-	61
No. SVD cases	474	106	201	-	25	63	310	97	73	240	-	38	54	-	173
Age of cases and controls (mean ± SD)	72±13/ 52[2]	67 ±13/ 63±11	67±14/ 65±13	71±8/ 69±9	57±16/ 51±8	57±5/ 54±6	73±13/ 70±12	74±14/ ≥65	82±6/ 86±6	73±12/ 57±21	80±11/ 72±12	67±15/ 48±9	41±7/ 40±7	65±8/ 59±9	69±9 / 67±9
	GENOTYPING INFORMATION														
GWAS panel used	Illumina 660		Illumina 550/610/ 660	Illumina 550	Illumina 610/660	Affy-metrix 6.0	Illumina 610		Illumina 370	Illumina 317/370	Affy-metrix 550	Affymetrix 6.0	Illumina H Omni	Illu-mina 610	Illu-mina 370
QC filters applied before imputation	QC steps before imputation done in every center, including removal of ancestry outliers defined by principal component analysis and poorly typed individuals.														
Imputation software	MaCH http://www.sph.umich.		MaCH v1.0.16	MaCH		MaCH v1.0.16			BIM-BAM(5)	IMPUTE	MaCH v1.0.15	MaCH v1.0.16	Not impute	MaCH	BIM-BAM

B. Cohorts contributing data for the ischemic stroke phenotype															
Cohort name	WTCCC 2_UK	WTCCC 2_D	ISGS & SWISS	Rotter-dam	Milano	ARIC	ASGC	Brains	CHS	deCODE	FHS	GASROS	GEOS	HPS	HVH
	edu/csg/abecasis/MACH/download/												d		
Reference imputed to	HapMap2 CEU		1000G (Aug 2010)	HapMap#22	HapMap2 CEU	HapMap2 CEU	HapMapII #24 CEU		NCBI b35	HapMap2 CEU	HapMap #22	HapMap 3 CEU+TSI training set	Not imputed	HapMap2 #22	
Imputation quality metric reported	MaCH oevar[3]								O/E ratio	IMPUTE info	MaCH oevar		n/a	MaCH oevar	O/E ratio
QC filtering applied to imputed variants	Included SNPs with imputation quality ≥0.3 (O/E variance or IMPUTE info score) and MAF ≥1%														
Software and statistical model used for association analyzed	Logistic regression for all cohorts with a cross-sectional study design to model the multiplicative SNP effects on risk of the dichotomous outcome of stroke against ancestry-matched controls. Cox proportional-hazards models for prospective studies to assess time to first stroke, fitting an additive model relating genotype dose to the stroke outcome.														
Handling of population stratification	Four cohorts used ancestry-informative principal components as covariates to correct for population stratification (ISGS/SWISS, GEOS, ASGC, Brains). Age and sex were included as covariates in two centers (ISGS/SWISS and Brains), sex was used as a covariate in one center (GASROS) and one center used recruitment phase (1 or 2) as a covariate (GEOS). In all other centers no covariates were included.														
HWE	All centers screened SNPs for HWE errors prior to imputation or analysis														
Adjusting for cryptic or overt relatedness	All centers screened individuals for relatedness prior to imputation or analysis														

[1]CS: cross sectional study; PB:population-based study.
[2]The approximate age of genotyping of the 2738 controls from the 1958 Birth Cohort. Age was not available for the remaining controls.
[3]MaCH oevar: the measure is the ratio of the empirically observed variance of the allele dosage to the expected binomial variance at Hardy–Weinberg equilibrium

WTCCC2: Wellcome Trust Case Control Consortium 2; UK: The United Kingdom of Great Britain and Northern Ireland; D: Germany; ISGS: Ischemic Stroke Genetics Study; SWISS: Sibling with Ischaemic Stroke Study; ARIC: Atherosclerosis Risk in Communities study; ASGC: Australian Stroke Genetics Collaborative; BRAINS: Bio-Repository of DNA in stroke; CHS: Cardiovascular Health Study; FHS: Framingham Heart Study; GASROS: Genes Affecting Stroke Risk and Outcome Study; GEOS: Genetics of Early-Onset Stroke: HPS: Heart Protection Study; HVH: Heart and Vascular Health Study

C. Cohorts contributing data for the WMH in ischemic stroke phenotype											
Cohort name	**ASGC**	**GASROS**	**SWISS**	**ISGS**	**Edinburgh**	**Milano**	**SGUL**	**Oxford FLAIR**	**Oxford T2**	**Munich FLAIR**	**Munich T2**
	STUDY INFORMATION										
Study design[1]	HB	HB	HB	HB	HB	HB	HB	PB	PB	HB	HB
Study inclusion and exclusion criteria	Included: age >18 years; diagnosis of ischemic stroke of any subtype Excluded: cases of CADASIL, vasculitis, demyelinating and mitochondrial disorders										
WMH quantification method used	WMH assessed in hemisphere contralateral to acute ischemic stroke. All supratentorial WM lesions were included in WMH volume (WMHV) with the exception of WMH corresponding to lacunar infarcts; sequential axial FLAIR images analyzed using a validated, semi-automated protocol (MRIcro, Un. of Nottingham)				WMH assessed in the hemisphere contralateral to acute ischemic stroke. All supratentorial white matter lesions were included in WMH volume (WMHV) with the exception of WMH corresponding to lacunar infarcts; Analyzed using DISPunc semi-automated lesion drawing software.						
MRI scanner used	1.5 T Siemens Magnetom Avanto	1.5 T GE Medical Signa			1.5T GE Medical Signa, 1.5 T Siemens	1.5T Siemens, 0.5T Philips	1.5 T Philips, 1.5T GE Signa LX	1.5T GE Medical Signa, 1.5 T Philips		1.5T Siemens Magnetom, 1T Siemens, 1.5 T GE Medical Signa	1.5T Siemens Magnetom, 3T and 1.5T GE Medical Signa, 1T Siemens
MRI sequence used for assessing WMH	Axial FLAIR				Axial FLAIR	Coronal FLAIR or Axial FLAIR	Axial FLAIR	Coronal FLAIR	Axial T2	Axial FLAIR	Axial T2
Key study references	Adib-Samii et al, Stroke 2013; Rost et al Neurology 2010;(6, 7)										
	SUBJECTS' CHARACTERISTICS										
Ancestry of study sample	European ancestry										

C. Cohorts contributing data for the WMH in ischemic stroke phenotype											
Cohort name	**ASGC**	**GASROS**	**SWISS**	**ISGS**	**Edinburgh**	**Milano**	**SGUL**	**Oxford FLAIR**	**Oxford T2**	**Munich FLAIR**	**Munich T2**
Number of cases assessed for WMH	104	975	115	209	65	152	323	65	75	447	203
Age of cases assessed for WMH (mean ± SD)	65±13	66±14	66±11	68±14	69 ± 13	58 ± 14	71 ± 13	65 ± 15	68 ± 13	66 ± 12	67 ± 12
GENOTYPING INFORMATION											
GWAS panel used	Human610-Quad	Affymetrix 6.0, Illumina Human610-Quad or Illumina OmniExpress beadchips	Illumina650K-Quad		Illumina Human660W-Quad	Illumina Human610-Quad or Human660W-Quad	Illumina Human660W-Quad				
QC filters applied to genotype data before imputation	Individuals removed if inferred sex discordant with recorded sex; if >5% missing genotype data;				Individuals removed if inferred sex discordant with recorded sex; if >5% missing genotype data; SNPs were excluded for minor allele frequency <1% or >5% missing data;						
Imputation software	IMPUTE2										
Which reference imputation done to	HapMap3 and 1000 Genomes Project Phase pilot (June 2010)	HapMap3 and 1000 Genomes Project Phase pilot (June 2010) or 1000 Genomes Integrated Release (June 2011)	HapMap3 and 1000 Genomes Project Phase pilot (June 2010)		1000 Genomes Phase 1 integrated variant set (March 2012)						

C. Cohorts contributing data for the WMH in ischemic stroke phenotype											
Cohort name	**ASGC**	**GASROS**	**SWISS**	**ISGS**	**Edinburgh**	**Milano**	**SGUL**	**Oxford FLAIR**	**Oxford T2**	**Munich FLAIR**	**Munich T2**
Imputation quality metric reported	IMPUTE info score & PLINK info score										
QC filtering applied to imputed variants	Removed SNPs with IMPUTE info score <0.3 and SNPs with MAF <1%										
Software and statistical model used for association analyzed	WMHV from the hemisphere contralateral to acute ischemic stroke was doubled to obtain whole brain values and adjusted for normal inter-individual variation in head size. Values were natural log transformed to a normal distribution. Within each group, rank- transformed residuals were derived from a linear regression model predicting WMHV with age, sex, and the first 2 ancestry principle components as covariates in GenABEL. Thus, the phenotype was adjusted for age because WMHV is highly age dependent. Principal components, derived using EIGENSTRAT, were included to correct for potential population stratification. Association analysis was undertaken in PLINK using pseudodosages, a fractional count of 0 to 1 alleles for each genotype weighted by imputation probability, within a linear regression (additive) model.										
Handling of population stratification	Principal component analysis. Remotion of population outliers by visual inspection of PC1*PC2 plot, followed by inclusion of PC1-PC4 into regression models.				2 ancestry informative principal components covariates were included in the model used to derive the WMH phenotype						
HWE	Excluded if HWE $p<1x10^{-6}$										
Adjusting for cryptic or overt relatedness	Excluded individuals with an inferred first- or second-degree relative in the sample identified on the basis of pairwise allele sharing estimates (estimated genome proportion shared identical by descent) $\pi>0.2$				To obtain a set of putatively unrelated individuals, a hidden Markov model (HMM) was used to infer identity by descent and then individuals were removed iteratively to obtain a set with pairwise identity by descent of <5%						
Strand and build of the human genome on which results are provided	+ strand, build 37/hg19										

[1]HB: hospital-based stroke study; PB: population-based stroke study

SGUL: ST George's University London

D. Cohorts contributing data for the WMH in the population phenotype							
Cohort name	**ARIC**	**CHS**	**FHS**	**Rotterdam Study I**	**Rotterdam Study II**	**AGES**	**ASPS**
	STUDY INFORMATION						
Study design	Prospective population-based cohort studies						
Study inclusion and exclusion criteria	In all cohorts, participants were excluded if they lacked information on MRI, genotypes, or both, or if they suffered a stroke or transient ischemic attack prior to MRI. In addition, CHS did not genotype participants with clinical cardiovascular disease at baseline. ASPS and RS did not perform MRI scans in participants with dementia, and FHS analyses excluded participants who had dementia at the time of MRI.						
WMH quantification method used	In AGES-Reykjavik, ASPS, FHS, and Rotterdam, WMH volume was estimated on a quantitative scale using custom-written computer programs based on an automatic segmentation algorithm or a semiautomatic segmentation analysis involving operator-guided removal of nonbrain elements. In ARIC and CHS, WMH volume was estimated on a semiquantitative 10-point scale by visual comparison with 8 templates that successively increased from barely detectable white matter lesions to extensive, confluent abnormalities. Study participants' brain images were compared with the reference standards after aligning them to approximately the same apparent size. Hence, visual grades are inherently normalized for brain size.						
MRI scanner used	General Electric or Picker 1.5 Tesla scanners	1.5T General Electric or Picker or 0.35 T Toshiba	1 or 1.5 T Siemens Magnetom scanner	1.5 T Siemens Vision scanner	1.5 T GE Healthcare scanner	1.5 T Signa TwinSpeed system	1.5 T Gyroscan S15 and ACS
MRI sequence used for assessing WMH	T1- and T2-weighted scans in the axial plane were obtained for all participants. These were complemented by either scans obtained with fluid attenuation inversion recovery or proton density sequences to allow better separation of white matter hyperintensities and cerebrospinal fluid.						
Key study references	Fornage M et al, Annals of Neurology 2011(8)						
	SUBJECTS' CHARACTERISTICS						
Ancestry of study sample	European ancestry						
Number of cases in the study assessed for WMH	798	2184	2200	380	567	2467	765

D. Cohorts contributing data for the WMH in the population phenotype							
Cohort name	**ARIC**	**CHS**	**FHS**	**Rotterdam Study I**	**Rotterdam Study II**	**AGES**	**ASPS**
Age of cases assessed for WMH (mean ± SD)	63±4	72±5	64± 11	73±8	67±5	76±5	65±8
	GENOTYPING INFORMATION						
GWAS panel used	Affymetrix GeneChip SNP Array 6.0	Illumina Human 370-Duo BeadChip	Affymetrix GeneChip Human mapping 500K Array Set and 50K Human gene Focused panel	Illumina HumanHap550 Duo BeadChip	Illumina HumanHap550 Duo BeadChip	Illumina Human 370-Duo BeadChip	Illumina 610-Quad BeadChip
QC filters applied to genotype data before imputation	Participant-specific quality control filters were applied based on missing call rate, cryptic relatedness, and number of Mendelian errors per individual. SNP-specific quality controls included filters for call rate, minor allele frequency, Hardy-Weinberg equilibrium, differential missingness by outcome or genotype.						
Imputation software	MaCH (v1.0.15 or 1.0.16)	BIM-BAM 15	MaCH (v1.0.15 or 1.0.16)				
Which reference imputation done to	HapMap2 CEU #22; +strand of NCBI build 36	+strand of NCBI build 35	HapMap2 CEU #22; +strand of NCBI build 36				
Imputation quality metric reported	O/E ratio and oevar						
QC filtering applied to imputed variants	Excluded SNPs with O/E ratio & oevar <0.3 and MAF<1%						
Software and statistical model used for association analyzed	Within each study, a linear regression model was used to evaluate the association of the natural log-transformed volume of WMHs (log[WMH + 1]) with the number of minor alleles (0 to 2) at each SNP. Analyses were adjusted only for age, sex, and total intracranial volume (except in ASPS, ARIC, and CHS). ARIC and CHS also adjusted for study site, and FHS adjusted for familial structure.						
HWE	Excluded if Hardy-Weinberg $p<1x10^{-5}$ in CHS and $p<1x10^{-6}$ in AGES-Reykjavik, ARIC, ASPS, FHS, and Rotterdam						

D. Cohorts contributing data for the WMH in the population phenotype							
Cohort name	**ARIC**	**CHS**	**FHS**	**Rotterdam Study I**	**Rotterdam Study II**	**AGES**	**ASPS**
Adjusting for cryptic or overt relatedness and population stratification	Studies were screened for latent population substructure, including cryptic relatedness, using suitable programs: EIGENSTRAT in ARIC, FHS and AGES-Reykjavik, an IBD matrix in ASPS and Rotterdam, and using principal component analysis in CHS.						
Strand and build of the human genome	+ strand; hg18						

AGES: Aging Gene-Environment Susceptibility-Reykjavik Study
ASPS: Austrian Stroke Prevention Study

Supplementary table e-2.

A. Original measures of association (β) and the derived odds ratios (ORs).

	Phenotype	β (minor allele)	SE of β	OR (minor allele)	95% CI of OR
rs 9521732 (minor allele A)	**All ICH**	0.1548	0.0519	1.17	1.05 to 1.29
	Deep ICH	0.2451	0.0614	1.28	1.13 to 1.44
	Lobar ICH	0.0489	0.068	1.05	0.92 to 1.20
	All IS	0.0261	0.0161	1.03	0.99 to 1.06
	CE IS	0.022	0.0321	1.02	0.96 to 1.09
	LVD IS	-0.0465	0.0354	0.95	0.89 to 1.02
	Lacunar IS	0.0858	0.0357	1.09	1.02 to 1.17
	WMH in IS*	0.0707	0.0289	1.07	1.01 to 1.14
	WMH in population *	0.0088	0.008	1.01	0.99 to 1.04
rs 9521733 (minor allele C)	**All ICH**	0.1706	0.0521	1.19	1.07 to 1.31
	Deep ICH	0.2547	0.0614	1.29	1.14 to 1.46
	Lobar ICH	0.0674	0.068	1.07	0.94 to 1.22
	All IS	0.0294	0.0164	1.03	1.00 to 1.06
	CE IS	0.0151	0.0326	1.02	0.95 to 1.08
	LVD IS	-0.0466	0.0353	0.95	0.89 to 1.02
	Lacunar IS	0.0979	0.0365	1.10	1.03 to 1.18
	WMH in IS*	0.0691	0.0288	1.07	1.01 to 1.13
rs 9515199 (minor allele C)	**All ICH**	0.1554	0.0519	1.17	1.06 to 1.29
	Deep ICH	0.2465	0.0614	1.28	1.14 to 1.44
	Lobar ICH	0.0487	0.068	1.05	0.92 to 1.20
	All IS	0.0308	0.0165	1.03	1.00 to 1.07
	CE IS	0.02	0.0332	1.02	0.96 to 1.09
	LVD IS	-0.0406	0.036	0.96	0.89 to 1.03
	Lacunar IS	0.0891	0.0365	1.09	1.02 to 1.17
	WMH in IS*	0.0707	0.0289	1.07	1.01 to 1.14
	WMH in population *	0.0088	0.008	1.01	0.99 to 1.04

We calculated the ORs from the β-coefficients by using the EXP function in Excel (inverse of the natural logarithm). We calculated the 95% CIs of the ORs by first calculating the error factor (EF=EXP(1.96*SE)) and then dividing and multiplying the OR with the EF to calculate the lower and upper 95% CIs respectively.

* For WMH cohorts the ORs represent the OR per 1 unit standard deviation (SD) change in WMH volume. For WMH in ischemic stroke cohorts, we calculated this from the β-coefficient by simply using the EXP function in Excel, as the WMH values had a SD of 1. For WMH in population cohorts, we divided the ß-coefficients and SEs by their respective pooled SD across cohorts (0.684) first and then used the transformed β-coefficient and SE to calculate the respective ORs and their 95% CIs.

B. Imputed and directly genotyped SNPs.

Phenotype	Study	Imputation quality measures		
		rs9521732	**rs9521733**	**rs9515199**
ICH	ISGC	1[a]	G	1[a]
	GOCHA	1[a]	G	0.99[a]
	GERFHS	G	G	1[a]
ISCHEMIC STROKE	ARIC	G	n/a	1[b]
	ASGC	0.96[b]	G	0.97[b]
	Brains	0.99[b]	G	0.98[b]
	CHS	0.92[c]	G	0.92[c]
	deCODE	0.99[d]	G	0.99[d]
	GASROS	1[b]	0.99[b]	n/a
	GEOS	G	n/a	n/a
	HVH	0.92[c]	G	0.91[c]
	ISGS & SWISS	0.98[b]	G	0.98[b]
	Milano	0.97[b]	1[b]	0.97[b]
	WTCCC2_D	0.98[b]	G	0.97[b]
	WTCCC2_UK	0.99[b]	G	0.99[b]
	FHS	0.94[b]	0.94[b]	0.94[b]
	HPS	0.97[b]	G	0.97[b]
	Rotterdam	0.99[b]	1[b]	0.99[b]
WMH in symptomatic stroke	Edinburgh	0.97[d]	G	0.97[d]
	Munich_T2	0.90[d]	G	0.90[d]
	Munich_FLAIR	1[d]	G	1[d]
	Oxford_T2	0.94[d]	G	0.94[d]
	Oxford_FLAIR	0.77[d]	G	0.77[d]
	SGUL	0.98[d]	G	0.98[d]
	Milano	0.80[d]	0.84[d]	0.80[d]
	ASGC	1[d]	1[d]	1[d]
	GASROS_affy	G	G	1[d]
	GASROS_illumina	1[d]	G	1[d]
	GASROS_omni	G	1[d]	1[d]
	ISGC-SWISS	0.93[d]	G	0.93[d]
WMH in population	AGES	0.99[b]	n/a	0.98[b]
	ARIC	G	n/a	1[b]
	CHS	0.92[c]	n/a	0.92[c]
	FHS	0.94[b]	n/a	0.94[b]
	Rotterdam Study I	0.99[b]	n/a	0.99[b]
	Rotterdam Study II	0.99[b]	n/a	0.99[b]
	ASPS	0.98[b]	n/a	0.98[b]

G: directly genotyped; n/a: SNP not available for the study.

Imputation quality numbers provided reflect the minimum quality measure across sub-phenotypes, rounded to 2 decimal places.

[a] PLINK info [b] MaCH oevar [c] BIMBAM O/E ratio [d] IMPUTE info

Supplementary table e-3.

A. Functional annotation of SNPs in moderate LD (r^2>0.3) with rs9521732, rs9521733 and rs9515199*.

Gene	hg 19	Consequence Ensembl	Consequence SeattleSeq	Haploreg: dbSNP functional annotation
COL4A2	rs111269249	-	-	intronic
COL4A2	rs115945569	-	-	intronic
COL4A2	rs11619425	intron_variant	intron-variant	intronic
COL4A2	rs11619427	intron_variant	intron-variant	intronic
COL4A2	rs11619430	intron_variant	intron-variant	intronic
COL4A2	rs11838637	intron_variant	intron-variant	intronic
COL4A2	rs11838776	intron_variant	intron-variant	intronic
COL4A2	rs12853693	intron_variant	intron-variant	intronic
COL4A2	rs1888004	intron_variant	intron-variant	intronic
COL4A2	rs1888005	intron_variant	intron-variant	intronic
COL4A2	rs1927342	intron_variant	intron-variant	intronic
COL4A2	rs1927343	-	intron-variant	intronic
COL4A2	rs1927344	intron_variant	intron-variant	intronic
COL4A2	rs1927345	intron_variant	intron-variant	intronic
COL4A2	rs1927346	intron_variant	intron-variant	intronic
COL4A2	rs1927347	intron_variant	intron-variant	intronic
COL4A2	rs1927349	intron_variant	intron-variant	intronic
COL4A2	rs1927355	intronic downstream_ gene_variant†	-	intronic
COL4A2	rs1999013	intron_variant	intron-variant	intronic
COL4A2	rs2149067	intron_variant	intron-variant	intronic
COL4A2	rs2391825	intron_variant	intron-variant	intronic
COL4A2	rs34402154	intron_variant	intron-variant	intronic
COL4A2	rs34992019	intron_variant	intron-variant	intronic
COL4A2	rs3899318		intron-variant	intronic
COL4A2	rs4283091	intron_variant	intron-variant	intronic
COL4A2	rs4492912	intron_variant	intron-variant	intronic
COL4A2	rs4547215	intron_variant	intron-variant	intronic
COL4A2	rs4586292	intron_variant	intron-variant	intronic
COL4A2	rs4771674	intron_variant	intron-variant	intronic
COL4A2	rs4771675	intron_variant	intron-variant	intronic
COL4A2	rs4771676	intron_variant	intron-variant	intronic
COL4A2	rs4773157	intron_variant	intron-variant	intronic
COL4A2	rs4773169	intron_variant	intron-variant	intronic
COL4A2	rs4773170	intron_variant	intron-variant	intronic
COL4A2	rs4773171	-	intron-variant	intronic
COL4A2	rs4773173	intron_variant	intron-variant	intronic
COL4A2	rs4773174	intron_variant	intron-variant	intronic
COL4A2	rs4773177	intron_variant	intron-variant	intronic
COL4A2	rs55940034	intron_variant	intron-variant	intronic
COL4A2	rs61963197	intron_variant	intron-variant	intronic

Gene	hg 19	Consequence Ensembl	Consequence SeattleSeq	Haploreg: dbSNP functional annotation
COL4A2	rs7318424	-	intron-variant	intronic
COL4A2	rs7318742	intron_variant	intron-variant	intronic
COL4A2	rs7321362	intron_variant	intron-variant	intronic
COL4A2	rs7323228	intron_variant	intron-variant	intronic
COL4A2	rs7326145	-	intron-variant	intronic
COL4A2	rs7328731	intron_variant	intron-variant	intronic
COL4A2	rs7333596	intron_variant	intron-variant	intronic
COL4A2	rs7333748	-	intron-variant	intronic
COL4A2	rs7334022	intron_variant	intron-variant	intronic
COL4A2	rs750598	intron_variant	intron-variant	intronic
COL4A2	rs7982993	-	intron-variant	intronic
COL4A2	rs7983374	intron_variant	intron-variant	intronic
COL4A2	rs7990844	intron_variant	intron-variant	intronic
COL4A2	rs7991842	intron_variant	intron-variant	intronic
COL4A2	rs7999034	intron_variant	intron-variant	intronic
COL4A2	rs872587	-	intron-variant	intronic
COL4A2	rs872588	intron_variant	intron-variant	intronic
COL4A2	rs872589	intron_variant	intron-variant	intronic
COL4A2	rs913746	-	intron-variant	intronic
COL4A2	rs9284253	intron_variant	intron-variant	intronic
COL4A2	rs9301454	intron_variant	intron-variant	intronic
COL4A2	rs9515195	-	intron-variant	intronic
COL4A2	rs9515196	-	intron-variant	intronic
COL4A2	rs9515197	intron_variant	intron-variant	intronic
COL4A2	**rs9515199**	**intron_variant**	**intron-variant**	**intronic**
COL4A2	rs9515201	-	intron-variant	intronic
COL4A2	rs9515204	intron_variant	intron-variant	intronic
COL4A2	rs9521717	-	intron-variant	intronic
COL4A2	rs9521718	-	intron-variant	intronic
COL4A2	rs9521719	intron_variant	intron-variant	intronic
COL4A2	rs9521720	intron_variant	intron-variant	intronic
COL4A2	rs9521721	-	intron-variant	intronic
COL4A2	rs9521729	intron_variant	intron-variant	intronic
COL4A2	rs9521730	-	intron-variant	intronic
COL4A2	**rs9521732**	**intron_variant**	**intron-variant**	**intronic**
COL4A2	**rs9521733**	**intron_variant**	**intron-variant**	**intronic**
COL4A2	rs9521734	intron_variant	intron-variant	intronic
COL4A2	rs9521735	intron_variant	intron-variant	intronic
COL4A2	rs9521739	intron_variant	intron-variant	intronic
COL4A2	rs9521740	intron_variant	intron-variant	intronic
COL4A2	rs9521742	-	intron-variant	intronic
COL4A2	rs9521743	intron_variant	intron-variant	intronic
COL4A2	rs9521744	intron_variant	intron-variant	intronic
COL4A2	rs9521746	intron_variant	intron-variant	intronic
COL4A2	rs9521747	intron_variant	intron-variant	intronic
COL4A2	rs9521748	intron_variant	intron-variant	intronic

Gene	hg 19	Consequence Ensembl	Consequence SeattleSeq	Haploreg: dbSNP functional annotation
COL4A2	rs952359	intron_variant	intron-variant	intronic
COL4A2	rs9555692	-	intron-variant	intronic
COL4A2	rs9555694	-	intron-variant	intronic
COL4A2	rs9555695	intron_variant	intron-variant	intronic
COL4A2	rs9559780	-	intron-variant	intronic
COL4A2	rs9559781	-	intron-variant	intronic
COL4A2	rs9559788	-	intron-variant	intronic
COL4A2	rs9583489	intron_variant	intron-variant	intronic
COL4A2	rs9588151	intron_variant	intron-variant	Intronic

-:SNP not included in database

* the three SNPs significantly associated with deep ICH are shown in bold.

† A sequence variant located 3' of a gene

B. Functional annotation of SNPs in high LD (r^2>0.9) with rs9521732, rs9521733 and rs9515199: Regulome db database results.

HG19	Regulome db score	Motifs[1]	Histone modification[2]	Protein binding[3]	Chromatin structure[4]
rs9521732	no data	n/a	n/a	n/a	n/a
rs9515199	no data	n/a	n/a	n/a	n/a
rs9521735	no data	n/a	n/a	n/a	n/a
rs4771674	no data	n/a	n/a	n/a	n/a
rs9521733	5	-	✓	-	✓
rs9521734	6	✓	✓	-	-
rs1999013	5	✓	✓	-	✓
rs9555695	4	-	✓	✓	✓

n/a: not applicable; - : no evidence; ✓: evidence
Gray boxes indicate that there is some evidence for this functional annotation based on experiments done on more relevant tissues: brain, spinal cord or blood vessel tissue.

Regulome db score: The scoring system represents with increasing confidence that a variant lies in a functional location and is likely to result in a functional consequence (lower scores indicate increasing evidence for a variant to be located in a functional region.)
1: Likely to affect binding and linked to expression of a gene target
2: Likely to affect binding
3: Less likely to affect binding
4-6: Minimal binding evidence (lack evidence of the variant actually disrupting the site of binding)

[1]Motifs: the SNP is located in an area of short recurring patterns in DNA (motif) thought to have a regulatory function and hence may predict transcription factor binding sites.
[2]Histone modification: the SNP is located in an area of histone modification. Histones are proteins that associate with DNA in the nucleus and help condense it into chromatin. Histone modifications are a range of post-translational modifications to the N-terminal tails of the histones in chromatin, which include a series of methylations and acetylations at defined lysine and arginine residues. Histone modification profiles are associated with differences in gene transcription and hence can be used as a generic tool to identify functional elements in the genome.
[3]Protein binding: the SNP is located in an area binding a transcription factor.
[4]Chromatin structure: the SNP is located in an area of possible chromatin accessibility, suggesting the area has a regulatory function. Chromatin is a complex of DNA and proteins that forms chromosomes within the nucleus of eukaryotic cells.

Reference List

(1) Falcone GJ, Biffi A, Devan WJ, et al. Burden of Blood Pressure-Related Alleles Is Associated With Larger Hematoma Volume and Worse Outcome in Intracerebral Hemorrhage. *Stroke* 2013;44:321-326.

(2) Woo D, Falcone GJ, Devan WJ, et al. Meta-analysis of genome-wide association studies identifies 1q22 as a novel susceptibility locus for intracerebral hemorrhage. *American Journal of Human Genetics* 2014;pii: S0002-9297(14)00070-6. doi: 10.1016/j.ajhg.2014.02.012. [Epub ahead of print].

(3) Woo D, Sauerbeck LR, Kissela BM, et al. Genetic and Environmental Risk Factors for Intracerebral Hemorrhage: Preliminary Results of a Population-Based Study. *Stroke* 2002;33:1190-1196.

(4) 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 Neurology* 2012;11:951-962

(5) Servin B, Stephens M. Imputation-Based Analysis of Association Studies:Candidate Regions and Quantitative Traits. *PLoS Genetics* 2007;3:e114.

(6) Adib-Samii P, Rost N, Traylor M, et al. 17q25 Locus Is Associated With White Matter Hyperintensity Volume in Ischaemic Stroke, But Not With Lacunar Stroke Status. *Stroke* 2013;44:1609-1615.

(7) Rost NS, Rahman RM, Biffi A, et al. White matter hyperintensity volume is increased in small vessel stroke subtypes. *Neurology* 2010;75:1670-1677.

(8) Fornage M, Debette S, Bis JC, et al. Genome-Wide Association Studies of Cerebral White Matter Lesion Burden: The CHARGE Consortium. *Annals of Neurology* 2011;69:928-939.