Meta-Analysis of Genome-Wide Association Studies Identifies Genetic Risk Factors for Stroke in African Americans

Cara L. Carty, PhD*; Keith L. Keene, PhD*; Yu-Ching Cheng, PhD; James F. Meschia, MD; Wei-Min Chen, PhD; Mike Nalls, PhD; Joshua C. Bis, PhD; Steven J. Kittner, MD, MPH; Stephen S. Rich, PhD; Salman Tajuddin, MD, PhD; Alan B. Zonderman, PhD; Michele K. Evans, MD; Carl D. Langefeld, PhD; Rebecca Gottesman, MD, PhD; Thomas H. Mosley, PhD; Eyal Shahar, MD, MPH; Daniel Woo, MD; Kristine Yaffe, MD; Yongmei Liu, MD, PhD; Michèle M. Sale, PhD; Martin Dichgans, MD; Rainer Malik, PhD; W.T. Longstreth Jr, MD, MPH; Braxton D. Mitchell, PhD, MPH; Bruce M. Psaty, MD, PhD; Charles Kooperberg, PhD; Alexander Reiner, MD, MS; Bradford B. Worrall, MD, MSc; Myriam Fornage, PhD; on behalf of the COMPASS and METASTROKE Consortia

Background and Purpose—The majority of genome-wide association studies (GWAS) of stroke have focused on European-ancestry populations; however, none has been conducted in African Americans, despite the disproportionately high burden of stroke in this population. The Consortium of Minority Population Genome-Wide Association Studies of Stroke (COMPASS) was established to identify stroke susceptibility loci in minority populations.

Methods—Using METAL, we conducted meta-analyses of GWAS in 14 746 African Americans (1365 ischemic and 1592 total stroke cases) from COMPASS, and tested genetic variants with $P<10^{-6}$ for validation in METASTROKE, a consortium of ischemic stroke genetic studies in European-ancestry populations. We also evaluated stroke loci previously identified in European-ancestry populations.

Results—The 15q21.3 locus linked with lipid levels and hypertension was associated with total stroke (rs4471613; $P=3.9\times10^{-8}$) in African Americans. Nominal associations ($P<10^{-6}$) for total or ischemic stroke were observed for 18 variants in or near genes implicated in cell cycle/mRNA presplicing (*PTPRG*, *CDC5L)*, platelet function (*HPS4*), blood–brain barrier permeability (*CLDN17*), immune response (*ELTD1, WDFY4,* and *IL1F10-IL1RN*), and histone modification (*HDAC9)*. Two of these loci achieved nominal significance in METASTROKE: 5q35.2 ($P=0.03$), and 1p31.1 ($P=0.018$). Four of 7 previously reported ischemic stroke loci (*PITX2, HDAC9, CDKN2A/CDKN2B,* and *ZFHX3)* were nominally associated ($P<0.05$) with stroke in COMPASS.

Conclusions—We identified a novel genetic variant associated with total stroke in African Americans and found that ischemic stroke loci identified in European-ancestry populations may also be relevant for African Americans. Our findings support investigation of diverse populations to identify and characterize genetic risk factors, and the importance of shared genetic risk across populations. **(*Stroke*. 2015;46:2063-2068. DOI: 10.1161/STROKEAHA.115.009044.)**

Key Words: African Americans ■ genetic association studies ■ genome-wide association study ■ meta-analysis ■ stroke

As the fourth leading cause of death and a leading cause of long-term disability, stroke causes a substantial burden of mortality and morbidity in the United States.[1] Stroke is a heterogeneous disease consisting of multiple subtypes, each with unique pathogeneses[2] and risk factors.[3] Familial aggregation studies suggest that stroke has a substantial

Received March 15, 2015; final revision received May 20, 2015; accepted May 22, 2015.

Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA (C.L.C., C.K., A.R.); Center for Health Disparities, East Carolina University, Greenville, NC (K.L.K.); Baltimore Veterans Administration Medical Center, MD (Y.-C.C., S.J.K.); University of Maryland School of Medicine, Baltimore (Y.-C.C., S.J.K., B.D.M.); Mayo Clinic Florida, Jacksonville, FL (J.F.M.); Public Health Sciences, University of Virginia, Charlottesville (W.-M.C., S.S.R., M.M.S., B.B.W.); National Institute of Aging, NIH, Bethesda, MD (M.N., S.T., A.B.Z., M.K.E.); University of Washington, Seattle (J.C.B., W.T.L., B.M.P., A.R.); Wake Forest University, Winston-Salem, NC (C.D.L., Y.L.); Division of Cerebrovascular Neurology, Johns Hopkins University School of Medicine, Baltimore, MD (R.G.); Department of Neurology, University of Mississippi Medical Center, Jackson (T.H.M.); Mel and Enid Zuckerman College of Public Health, University of Arizona, Tucson (E.S.); Department of Neurology and Rehabilitation Medicine, University of Cincinnati, OH (D.W.); Department of Psychiatry, University of California, San Francisco (K.Y.); Institute for Stroke and Dementia Research, Munich Cluster for Systems Neurology (Synergy), Klinikum der Universität München, Ludwig-Maximilians-Universität; Munich, Germany (M.D., R.M.); Group Health Research Unit, Group Health Cooperative, Seattle, WA (B.M.P.); and Center for Human Genetics, University of Texas, Houston (M.F.).

*Drs Carty and Keene contributed equally.

Guest Editor for this article was Vladimir Hachinski, MD.

The online-only Data Supplement is available with this article at http://stroke.ahajournals.org/lookup/suppl/doi:10.1161/STROKEAHA.115.009044/-/DC1.

Correspondence to Cara L. Carty, PhD, Center for Translational Science, George Washington University, Children's National Medical Center, 111 Michigan Ave NW, Washington, DC 20010. E-mail ccarty@cnmc.org

genetic component.[4,5] Recent genome-wide association studies (GWAS), conducted almost exclusively in individuals of European ancestry, have identified stroke susceptibility loci on chromosomes 4q25,[6] 7p21.1,[7] 6p21,[8] 9p21,[6] 11q22,[9] 12p13,[10] 12q24,[11] and 16q22.[12] However, stroke incidence and mortality in African Americans are both nearly twice that in European-Americans.[13,14] Moreover, African Americans have strokes at younger ages on average, and more frequently endure poststroke disability.[13]

Using data obtained from the newly formed Consortium of Minority Population Genome-Wide Association Studies of Stroke (COMPASS), we conducted the first discovery GWAS meta-analysis of stroke in African Americans, validated our findings in the large METASTROKE consortium of ischemic stroke genetic studies in European-ancestry populations,[12] and determined if GWAS findings robustly associated with ischemic stroke in European-ancestry populations were also associated with stroke in African Americans.

Methods

Study Population

African American individuals with physician-adjudicated stroke and genome-wide single nucleotide polymorphism (SNP) data were included in these analyses. COMPASS includes cohort studies—Atherosclerosis Risk in Communities (ARIC) Study,[15] Cardiovascular Health Study (CHS),[16] Dynamics of Health, Aging, and Body Composition (HABC) Study,[17] and the Women's Health Initiative (WHI),[18]—and case-control studies—Genetics of Early Onset Stroke (GEOS),[19] Ischemic Stroke Genetics Study (ISGS),[20] Vitamin Intervention for Stroke Prevention (VISP),[21] and Healthy Aging in Neighborhoods of Diversity across the Life Span (HANDLS),[22]—as well as an affected sibpair study—Siblings with Ischemic Stroke Study (SWISS).[23]

Outcomes

Adjudicated ischemic or total strokes from all studies excluding HANDLS were analyzed. Race/ethnicity- and sex-matched controls were randomly selected from HANDLS and used as controls in the analyses of SWISS-ISGS and VISP, which lacked genotyped controls. All studies provided data for the total stroke analysis, which included ischemic and hemorrhagic strokes but excluded subarachnoid hemorrhage. All studies contributed to the ischemic stroke analyses with the exception of Dynamics of Health, Aging, and Body Composition Study, which lacked stroke subtype information. In the cohort studies, only first (incident) clinically validated strokes were considered and individuals with baseline history of stroke were excluded.

Genotype Data

All studies imputed SNPs using HapMap II reference populations; SNPs with low imputation quality (r^2<0.3) were excluded. We analyzed SNPs available in ≥2 studies with minor allele frequency ≥0.01 and passing stringent quality control criteria, for a total of ≈2.6 million SNPs. The online-only Data Supplement contains additional details about study designs, stroke definition, adjudication procedures, and genotyping.

Analysis

We used additive genetic models with a count of variant alleles (0, 1, or 2) for each genotyped SNP or allelic dose for imputed SNPs. Cohort studies used Cox proportional hazard models to evaluate associations between SNPs and the time to incident stroke. Case-control studies used logistic regression models (additional details in the online-only Data Supplement). To control for potential population stratification, principal components of global ancestry were estimated in each study and included as covariates. Models were additionally adjusted, as appropriate, for age, sex, and site. In each study, the distribution of test statistics was reviewed using Q–Q plots to detect potential inflation because of population stratification; no large deviations were noted. We combined study-specific results in fixed effects meta-analyses with inverse variance weighting using METAL[24]; *P*-value meta-analyses were also conducted. The genome-wide significance threshold was $P<5\times10^{-8}$ for the GWAS discovery but we investigated all SNPs with $P<10^{-6}$.

Validation of COMPASS Findings

Because of the absence of another large African American sample with GWAS and adjudicated stroke data, we performed a look-up of COMPASS SNPs with $P<10^{-6}$ in the METASTROKE ischemic stroke results. Given the known differences in linkage disequilibrium (LD) patterns between populations of European ancestry and African ancestry, we expanded the region of interest for each locus to include available SNPs±500 kb of the index SNPs. We applied Bonferroni correction to account for the number of loci tested.

As a secondary aim, we used COMPASS data to evaluate ischemic stroke and ischemic stroke subtype variants identified in European-ancestry populations.[6–8,10–12] In COMPASS, we tested the reported European-ancestry GWAS SNP (ie, index SNP) as well as SNPs in moderate LD (r^2≥0.50) with the index SNP (±500 kb) based on HapMap CEU to capture more broadly the European-ancestry index signal. Again, we applied a Bonferroni-corrected significance threshold.

Power

Using Quanto (v1.2.4),[25] we estimated reasonable power (≥80%) to detect low frequency common variants (0.09≥minor allele frequency≤0.15) associated with ≥50% increased (or decreased) risk of stroke in our discovery aim. As minor allele frequency increased to 0.50, we had reasonable power to detect increasingly smaller SNP-stroke associations, down to ≈25% increased (or decreased) risk (ie, effect size≈±0.22–0.29). Power to detect rarer variants, or common variants with more modest associations, was limited.[26]

Results

Discovery of Stroke-Associated Loci

COMPASS comprises 14745 African Americans, including 1365 ischemic and 1592 total stroke cases (Table I in the online-only Data Supplement). We identified an association between rs4471613 (15q21.3) and total stroke in African Americans, (β [SE]=0.82 [0.15]; $P=3.9\times10^{-8}$] with suggestive evidence for ischemic stroke (β [SE]=0.90 [0.18]; $P=4.6\times10^{-7}$]. An additional 18 SNPs had suggestive evidence of association ($P<1\times10^{-6}$; Table II in the online-only Data Supplement). Five were for ischemic stroke—including *PTPRG*, *CDC5L,* and *HPS4-ASPHD2* loci SNPs and 1 intergenic SNP (chr14q31)—and 15 SNPs were for total stroke—including SNPs in the *PTPRG, HDAC9,* and *WDFY4* loci and intergenic regions. Two SNPs, rs704341 (*PTPRG*/intron) and rs4471613 (15q21.3/intergenic), were suggestively associated with both ischemic and total stroke. In sensitivity analyses including only the cohort studies, associations for both SNPs remained below the $P<10^{-6}$ threshold (Table III in the online-only Data Supplement).

Validation of COMPASS SNPs in METASTROKE

None of the 19 SNPs from COMPASS (or SNPs in LD with them) met the Bonferroni threshold for significance in METASTROKE, *P*=0.05/19=0.003. We did observe nominal

Table 1. Validation of COMPASS Top Loci in METASTROKE

			COMPASS		METASTROKE				
Locus	Annotation	SNP	Outcome	*P* Value	Replication *P* Value	No. of SNPs*	Top SNP in Locus	*P* Value	LD†
1p31.1	Intergenic/*ELTD1*	rs1937787	Total	7.33×10^{-7}	0.195‡	75	rs35020936	0.018§	0.35
2q13	Intergenic/*IL1RN, IL1F10, IL36RN*	rs11681884	Total	6.13×10^{-7}	0.342‡	56	rs17042905	0.086	0.30
3p14.2	Intron/*PTPRG*	rs704341	Isc, total	7.11×10^{-7}	0.738	10	rs1871394	0.078	0.31
5q35	Intergenic/*MSX2, NKX2-5*	rs7705819	Total	9.47×10^{-7}	0.598	18	rs11747282	0.297	0.59
5q35	Intergenic/*SUMO2P6, GAPDHP71*	rs4867766	Total	5.22×10^{-7}	0.031	4	rs4867766	0.031§	1
6p21.1	UTR3/*CDC5L*	rs11572061	Isc	9.99×10^{-7}	n/a	8	rs11571943	0.140	0.48
6q16.1	Intergenic/*TSG1*	rs9345396	Total	1.03×10^{-7}	0.252‡	n/a	…	…	…
7p21	Intron/*HDAC9*	rs17347800	Total	3.59×10^{-7}	0.940‡	12	rs17138751	0.208	0.31
10p14	Intergenic/*LOC100507163*	rs768606	Total	1.36×10^{-7}	0.221	1	rs768606	0.221	1
10q11.2	Intron/*WDFY4*	rs17771318	Total	8.94×10^{-8}	0.941‡	n/a			
11q24	Intergenic/*TRNAK27, GLULP3, UBASH3B*	rs2084637	Total	9.12×10^{-7}	0.980‡	37	rs34614177	0.084	1
12q23	Intergenic/*RNU6-36, MIR135A2*	rs248812	Total	9.15×10^{-7}	0.171‡	24	rs34552	0.056	0.51
14q31	Intergenic	rs10400694	Isc	8.96×10^{-7}	0.922‡	70	rs12890538	0.203	0.32
14q31	Intergenic/*FLRT2*	rs7156510	Total	9.82×10^{-7}	0.876‡	77	rs12890538	0.203	0.34
14q31	Intergenic	rs1564060	Total	2.67×10^{-7}	0.798‡	46	rs4904162	0.225	0.40
15q21.3	Intergenic (near *ALDH1A2, AQP9, LIPC*)	rs4471613	Total, isc	3.94×10^{-8}	0.142	11	rs12591835	0.104	0.49
21q11.2	Intergenic (near *LIP1, ABCC13*)	rs2822388	Total	4.96×10^{-7}	0.224	2	rs2822388	0.224	1
21q22.1	Intergenic (*CLDN17*)	rs7283054	Total	9.75×10^{-7}	N/a	9	rs9974937	0.246	0.35
22q12.1	Intron/*HPS4*	rs5752326	Isc	8.84×10^{-7}	0.683‡	52	rs4822727	0.172	0.89

COMPASS indicates Consortium of Minority Population Genome-Wide Association Studies of Stroke; LD, linkage disequilibrium; n/a=not available; and SNP, single nucleotide polymorphism.

*SNPs±500 kb of index SNP and available in METASTROKE.

†Based on LD with the COMPASS SNP in ASW.

‡Direction of effect consistent with COMPASS SNP.

§For the METASTROKE validation, *P* values are nominally significant (not adjusted for multiple testing).

associations for 5q35.2 (rs4867766; *P*=0.031) and 1p31.1 (rs35020936; *P*=0.018; Table 1). In 5q35.2, the intergenic SNP rs4867766 was the top SNP in both race/ethnicities, despite different minor allele frequency, 0.08 in African American and 0.21 in European-ancestry populations, whereas in 1p31.1, the top SNP differed in the populations (Table 1).

Testing of Ischemic Stroke GWAS Loci Identified in European-Ancestry Populations

None of the European-ancestry index SNPs were associated with ischemic stroke in African Americans at *P*=0.007 (*P*=0.05/7 loci tested; Table 2). When we also investigated SNPs in modest LD (Table IV in the online-only Data Supplement), we found suggestive evidence of association (*P*<0.05) for the *PITX2, HDAC9, CDKN2A/CDKN2B,* and *ZFHX3* loci.

Discussion

The COMPASS collaboration represents the first large-scale GWAS meta-analysis of stroke in African Americans. We report a novel genome-wide association for total stroke at the 15q21.3 locus and report 14 additional loci suggestively associated with total or ischemic stroke in African Americans. In addition, in our African American population, we found suggestive evidence of replication for the *PITX2, HDAC9, CDKN2A/CDKN2B,* and *ZFHX3* loci previously associated with stroke in European-ancestry populations, pointing to potential shared mechanisms for stroke susceptibility.

The top SNP, rs4471613, is located near the 3′ region of the aquaporin 9 gene (*AQP9*), the aldehyde dehydrogenase 1 family, member A2 (*ALDH1A2*), and hepatic lipase (*LIPC*) genes. Previous work suggests a role for *AQP9* in cerebral energy metabolism as well as in brain ischemia, development of cerebral edema, and postischemic reuptake of glycerol and lactate.[27,28] Intergenic SNPs in this region also are associated with blood lipids in populations of diverse ancestry[29] and hypertension in African Americans.[30] Although this region is mechanistically appealing, rs4471613 is in low LD with these reported SNPs. The location of rs4471613 in a H3-lysine-27-acetylation histone mark in 7 different cell types reflecting 5 tissues (ENCODE data from UCSC genome browser [https://

Table 2. COMPASS Results for Ischemic Stroke SNPs Identified in European-Ancestry Populations

European-Ancestry SNP				COMPASS Ischemic Stroke Association					
Index SNP	Annotation	*P* Value	Outcome	Index SNP *P* Value	No. of SNPs in LD*	Top SNP	Allele/AF	*P* Value	LD With Index SNP
[12]rs13407662	2p16/intergenic	5.2×10^{-8}	SVD	0.184‡	3	rs2111856	T/0.08	0.141	0.83
[7]rs1906599	4q25/*PITX2*	1.4×10^{-9}	CE	0.104‡	23	rs2634073	T/0.41	0.014§	0.85
[12]rs6843082	4q25/*PITX2*	2.8×10^{-16}	CE	0.223‡	20	rs2634071	T/0.41	0.016§	0.84
[6]rs2200733	4q25/*PITX2*	2.2×10^{-10}	IS/CE	0.195‡	40	rs2634073	T/0.41	0.014§	0.55
[8]rs556621	6p21.1/intergenic	4.7×10^{-8}	LAA	0.136	18	rs658726	T/0.67	0.062	0.74
[7]rs11984041	7p21.1/*HDAC9*	1.9×10^{-11}	LVD	0.120‡	11	rs2526619	A/0.59	0.093	0.81
[7]rs2107595	7p21.1/*HDAC9*	2.0×10^{-16}	LVD	0.170‡	14	rs28688791	T/0.73	0.049§	0.67
[26]rs2383207†	9p21.3/*CDKN2A/B*	2.9×10^{-5}	LVD	0.182‡	40	rs1333040	T/0.62	0.049§	0.55
[10]rs11833579	12p13.33/*NINJ2*	2.3×10^{-8}	IS	0.642	1	rs11833579	A/0.22	0.642	1.0
[10]rs12425791	12p13.33/*NINJ2*	1.1×10^{-9}	IS	0.754‡	1	rs11833579	A/0.22	0.642	0.78
[12]rs879324	16q22.3/*ZFHX3*	2.3×10^{-8}	CE	0.919	12	rs16971456	C/0.85	0.048§	0.51

AF indicates average allele frequency in COMPASS African Americans; CE, cardioembolism; COMPASS, Consortium of Minority Population Genome-Wide Association Studies of Stroke; IS, all ischemic stroke; LAA, large artery atherosclerosis; LD, linkage disequilibrium; LVD, large vessel disease; SNP, single nucleotide polymorphism; and SVD, small vessel disease.

*No. of SNPs (±500 kb) in moderate LD ($r^2\geq0.50$) with the index SNP in HapMap CEU, and available in COMPASS.

†Significant replication in METASTROKE; originally reported in another study.

‡Direction of effect in COMPASS consistent with original report.

§*P* values are nominally significant (not adjusted for multiple testing).

genome.ucsc.edu]) suggests a possible regulatory function for this SNP, or neighboring SNPs in this region. However, this SNP was not significantly associated with ischemic stroke in METASTROKE ($P=0.13$). Further evaluation of genetic factors influencing stroke across the *AQP9-LIPC* region, particularly in populations of African descent, is warranted.

Two intergenic loci (5q35 and 1p31.1) identified in COMPASS were modestly associated ($P<0.05$) with stroke in METASTROKE. Overall, the lack of strong replication in the larger METASTROKE European-ancestry population urges caution when interpreting COMPASS associations. However, as has been reported for other phenotypes,[31] COMPASS loci could be unique to African-ancestry populations. Of note, the *HDAC9* variants associated with large vessel stroke in European-ancestry populations[7,12] are located >500 kb from the novel intronic *HDAC9* variant identified in COMPASS.

Replication of stroke GWAS variants across different ancestry/ethnicity groups are lacking but important for prioritizing genetic variants for translational research and understanding population differences in stroke burden. In our secondary aim to replicate previous GWAS findings from European-ancestry populations, we found suggestive evidence of replication ($P<0.05$) for 4 loci. Interestingly, 4 of the 6 nominally significant SNPs in these loci were originally associated with cardioembolic stroke in European-ancestry populations. None were significant after Bonferroni correction; nonetheless, these trends suggest that variants in these loci may influence stroke risk independent of race/ethnicity.

Some limitations of our study deserve mention. COMPASS case-control studies are limited to nonfatal and less severe strokes than longitudinal cohort studies, and are more likely to have potential selection bias. However, findings from analyses of the cohort studies were generally similar to the overall meta-analysis, though a rare intronic variant in *CHD3*, rs9899375, reached significance, $P=5.2\times10^{-9}$, in the cohort analyses only.

The limited number of African American stroke samples in COMPASS restricts power to discover new loci and also to replicate previous associations of modest effect size. In addition, SNP imputation in admixed populations, such as African Americans, is challenging because of differing LD patterns and greater genetic diversity than in the standard reference panels.[32] Given the use of YRI/CEU HapMap II reference panels rather than African-American-specific panels for imputation, we may have slightly reduced power to detect associations (resulting in greater chance of false negatives), and reduced ability to localize stroke-associated variants in our population.

Total and ischemic stroke are heterogeneous outcomes. Recent GWAS of ischemic stroke have detected heterogeneity of associations across stroke subtypes.[7] Although the largest GWAS of stroke in African Americans, COMPASS still lacks adequate numbers to stratify by stroke subtype. In COMPASS studies with subtype information, ischemic stroke is predominant and lacunar (or small vessel) stroke is the most common ischemic stroke subtype, consistent with data indicating that intracranial atherosclerotic and lacunar (or small vessel) strokes predominate in African Americans.[8,14] Furthermore, small vessel ischemic stroke incidence in African Americans is more than double that of US whites.[33] Thus, some of the observed differences in risk loci may reflect differences in distributions of major ischemic stroke subtypes across diverse populations. However, these differences offer an opportunity to identify shared or distinct risk factors and mechanisms for specific ischemic stroke subtypes across diverse populations, such as the nominally significant replication of cardioembolic-associated loci (*PITX2* and *ZFHX3*) in COMPASS. Replication of African American GWAS SNPs in a European-ancestry

population is not ideal because of differences in population LD, with African-ancestry populations having higher genetic diversity on average than European populations.[34] Yet, when we broadly interrogated the index SNP signals, we observed modest replication. Such data may help to localize the actual causal variants within GWAS signals.

Summary

Despite its limitations, our study presents results for a unique, lesser-studied population with a substantial stroke burden. Published stroke GWAS have identified only a handful of replicable associations, perhaps reflecting stroke phenotypic complexity and potential gene-by-environment influences in different populations.[35] Thus, genetic association studies in populations of diverse ancestry, such as this one, are critical to understanding the genetic basis of stroke, an increasingly important global disease.

Acknowledgments

We thank the staff and participants of the Atherosclerosis Risk in Communities study for their important contributions, the Women's Health Initiative (WHI) investigators and staff for their dedication, and the study participants for making the program possible. For a list of investigators who have contributed to WHI science, please visit: https://www.whi.org/researchers/Documents%20%20Write%20a%20Paper/WHI%20Investigator%20Long%20List.pdf.

Sources of Funding

Atherosclerosis Risk in Communities Study was supported by National Heart, Lung, and Blood Institute contracts (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C), R01HL087641, R01HL59367, and R01HL086694; National Human Genome Research Institute contract U01HG004402; and National Institutes of Health (NIH) contract HHSN268200625226C. Infrastructure was partly supported by Grant no. UL1RR025005, a component of the NIH and NIH Roadmap for Medical Research. Cardiovascular Health Study (CHS) was supported by National Heart, Lung, and Blood Institute contracts HHSN268201200036C, HHSN268200800007C, N01HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, N01HC85086; and grants U01HL080295, R01HL087652, R01HL105756, R01HL103612, and R01HL120393 with contributions from the National Institute of Neurological Disorders and Stroke. Additional support was provided through R01AG023629 from the National Institute on Aging. A full list of principal CHS investigators and institutions can be found at CHS-NHLBI.org. The provision of genotyping data were supported, in part, by the National Center for Advancing Translational Sciences, grant UL1TR000124, and the National Institute of Diabetes and Digestive and Kidney Disease Research Center grant DK063491 to the Southern California Diabetes Endocrinology Research Center. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. Genetics of Early Onset Stroke was supported by the NIH Genes, Environment and Health Initiative Grant U01 HG004436, as part of the GENEVA consortium, with additional support provided by the Mid-Atlantic Nutrition and Obesity Research Center (P30-DK072488); the Office of Research and Development, Medical Research Service, and the Baltimore Geriatrics Research, Education, and Clinical Center of the Department of Veterans' Affairs. Study recruitment and datasets assembly were supported by a Cooperative Agreement with the Division of Adult and Community Health, Centers for Disease Control and grants from the National Institute of Neurological Disorders and Stroke and the NIH Office of Research on Women's Health (R01-NS45012, U01-NS069208-01). Dr Cheng was supported by the Department of Veterans' Affairs career development award (1IK2BX001823). Healthy Aging in Neighborhoods of Diversity across the Life Span was supported by the Intramural Research Program of the NIH, National Institute of Aging and the National Center on Minority Health and Health Disparities (project no. Z01-AG000513 and human subjects protocol no. 2009-149). Ischemic Stroke Genetics Study (ISGS) and Siblings with Ischemic Stroke Study (SWISS) were supported by the National Institute of Neurological Disorders and Stroke grants (R01 NS42733; PI Meschia) and (R01NS39987; PI Meschia), respectively with additional support, in part, from the Intramural Research Program of the National Institute of Aging (Z01 AG000954-06; PI Singleton). Both studies used samples and clinical data from the NIH-NINDS Human Genetics Resource Center DNA and Cell Line Repository, human subjects protocol no. 2003-081 and 2004-147. Vitamin Intervention for Stroke Prevention (VISP) was funded by the National Institute of Neurological Disorders and Stroke (R01-NS34447). Genome-wide association study data for a subset of VISP participants supported by the National Human Genome Research Institute (U01-HG005160), as part of the Genomics and Randomized Trials Network (PI: Drs Sale and Worrall). Women's Health Initiative was supported by the National Heart, Lung, and Blood Institute, NIH, and Department of Health and Human Services through contracts N01WH22110, 24152, 32100-2, 32105-6, 32108-9, 32111-13, 32115, 32118-32119, 32122, 42107-26, 42129-32, and 44221. Funders had no role in study design, data collection and analysis, decision to publish, or article preparation.

Disclosures

Dr Worrall is an associate editor for the journal *Neurology*. The other authors report no conflicts.

References

1. Roger VL, Go AS, Lloyd-Jones DM, Benjamin EJ, Berry JD, Borden WB, et al; American Heart Association Statistics Committee and Stroke Statistics Subcommittee. Heart disease and stroke statistics–2012 update: a report from the American Heart Association. *Circulation*. 2012;125:e2–e220. doi: 10.1161/CIR.0b013e31823ac046.
2. Saeed M. Editorial comment–Unraveling the pagodian architecture of stroke as a complex disorder. *Stroke*. 2004;35:824–825.
3. Hajat C, Dundas R, Stewart JA, Lawrence E, Rudd AG, Howard R, et al. Cerebrovascular risk factors and stroke subtypes: differences between ethnic groups. *Stroke*. 2001;32:37–42.
4. Liao D, Myers R, Hunt S, Shahar E, Paton C, Burke G, et al. Familial history of stroke and stroke risk. The Family Heart Study. *Stroke*. 1997;28:1908–1912.
5. Bak S, Gaist D, Sindrup SH, Skytthe A, Christensen K. Genetic liability in stroke: a long-term follow-up study of Danish twins. *Stroke*. 2002;33:769–774.
6. Gretarsdottir S, Thorleifsson G, Manolescu A, Styrkarsdottir U, Helgadottir A, Gschwendtner A, et al. Risk variants for atrial fibrillation on chromosome 4q25 associate with ischemic stroke. *Ann Neurol*. 2008;64:402–409. doi: 10.1002/ana.21480.
7. Bellenguez C, Bevan S, Gschwendtner A, Spencer CC, Burgess AI, Pirinen M, et al. Genome-wide association study identifies a variant in HDAC9 associated with large vessel ischemic stroke. *Nat Genet*. 2012;44:328–333.
8. Holliday EG, Maguire JM, Evans TJ, Koblar SA, Jannes J, Sturm JW, et al; Australian Stroke Genetics Collaborative; International Stroke Genetics Consortium; Wellcome Trust Case Control Consortium 2. Common variants at 6p21.1 are associated with large artery atherosclerotic stroke. *Nat Genet*. 2012;44:1147–1151. doi: 10.1038/ng.2397.
9. Dichgans M, Malik R, König IR, Rosand J, Clarke R, Gretarsdottir S, et al; METASTROKE Consortium; CARDIoGRAM Consortium; C4D Consortium; International Stroke Genetics Consortium. Shared genetic susceptibility to ischemic stroke and coronary artery disease: a genome-wide analysis of common variants. *Stroke*. 2014;45:24–36. doi: 10.1161/STROKEAHA.113.002707.
10. Ikram MA, Seshadri S, Bis JC, Fornage M, DeStefano AL, Aulchenko YS, et al. Genomewide association studies of stroke. *N Engl J Med*. 2009;360:1718–1728. doi: 10.1056/NEJMoa0900094.
11. Kilarski LL, Achterberg S, Devan WJ, Traylor M, Malik R, Lindgren A, et al; GARNET Collaborative Research Group, Wellcome Trust Case Control Consortium 2, Australian Stroke Genetic Collaborative,

the METASTROKE Consortium, and the International Stroke Genetics Consortium. Meta-analysis in more than 17,900 cases of ischemic stroke reveals a novel association at 12q24.12. *Neurology*. 2014;83:678–685. doi: 10.1212/WNL.0000000000000707.
12. Traylor M, Farrall M, Holliday EG, Sudlow C, Hopewell JC, Cheng YC, et al; Australian Stroke Genetics Collaborative, Wellcome Trust Case Control Consortium 2 (WTCCC2); International Stroke Genetics Consortium. Genetic risk factors for ischaemic stroke and its subtypes (the METASTROKE collaboration): a meta-analysis of genome-wide association studies. *Lancet Neurol*. 2012;11:951–962. doi: 10.1016/S1474-4422(12)70234-X.
13. Lloyd-Jones D, Adams RJ, Brown TM, Carnethon M, Dai S, De Simone G, et al. Heart disease and stroke statistics–2010 update: a report from the American Heart Association. *Circulation*. 2010;121:e46–e215.
14. White H, Boden-Albala B, Wang C, Elkind MS, Rundek T, Wright CB, et al. Ischemic stroke subtype incidence among whites, blacks, and Hispanics: the Northern Manhattan Study. *Circulation*. 2005;111:1327–1331. doi: 10.1161/01.CIR.0000157736.19739.D0.
15. The ARIC Investigators. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. *Am J Epidemiol*. 1989;129:687–702.
16. Fried LP, Borhani NO, Enright P, Furberg CD, Gardin JM, Kronmal RA, et al. The Cardiovascular Health Study: design and rationale. *Ann Epidemiol*. 1991;1:263–276.
17. Wassel Fyr CL, Kanaya AM, Cummings SR, Reich D, Hsueh WC, Reiner AP, et al; Health, Aging, Body Composition Study. Genetic admixture, adipocytokines, and adiposity in Black Americans: the Health, Aging, and Body Composition study. *Hum Genet*. 2007;121:615–624. doi: 10.1007/s00439-007-0353-z.
18. Women's Health Initiative Study Group. Design of the Women's Health Initiative clinical trial and observational study. *Control Clin Trials*. 1998;19:61–109.
19. Cheng YC, O'Connell JR, Cole JW, Stine OC, Dueker N, McArdle PF, et al. Genome-wide association analysis of ischemic stroke in young adults. *G3 (Bethesda)*. 2011;1:505–514. doi: 10.1534/g3.111.001164.
20. Meschia JF, Brott TG, Brown RD Jr, Crook RJ, Frankel M, Hardy J, et al; Ischemic Stroke Genetics Study. The Ischemic Stroke Genetics Study (ISGS) Protocol. *BMC Neurol*. 2003;3:4. doi: 10.1186/1471-2377-3-4.
21. Spence JD, Howard VJ, Chambless LE, Malinow MR, Pettigrew LC, Stampfer M, et al. Vitamin Intervention for Stroke Prevention (VISP) trial: rationale and design. *Neuroepidemiology*. 2001;20:16–25.
22. Evans MK, Lepkowski JM, Powe NR, LaVeist T, Kuczmarski MF, Zonderman AB. Healthy aging in neighborhoods of diversity across the life span (HANDLS): overcoming barriers to implementing a longitudinal, epidemiologic, urban study of health, race, and socioeconomic status. *Ethn Dis*. 2010;20:267–275.
23. Meschia JF, Brown RD Jr, Brott TG, Chukwudelunzu FE, Hardy J, Rich SS. The Siblings With Ischemic Stroke Study (SWISS) protocol. *BMC Med Genet*. 2002;3:1.
24. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*. 2010;26:2190–2191. doi: 10.1093/bioinformatics/btq340.
25. Gauderman WJ, Morrison JM. QUANTO 1.1: a computer program for power and sample size calculations for genetic-epidemiology studies 2006. http://hydra.usc.edu/gxe. Accessed April 14, 2009.
26. Gschwendtner A, Bevan S, Cole JW, Plourde A, Matarin M, Ross-Adams H, et al; International Stroke Genetics Consortium. Sequence variants on chromosome 9p21.3 confer risk for atherosclerotic stroke. *Ann Neurol*. 2009;65:531–539. doi: 10.1002/ana.21590.
27. Lettre G, Palmer CD, Young T, Ejebe KG, Allayee H, Benjamin EJ, et al. Genome-wide association study of coronary heart disease and its risk factors in 8,090 African Americans: the NHLBI CARe Project. *PLoS Genet*. 2011;7:e1001300. doi: 10.1371/journal.pgen.1001300.
28. Badaut J. Aquaglyceroporin 9 in brain pathologies. *Neuroscience*. 2010;168:1047–1057. doi: 10.1016/j.neuroscience.2009.10.030.
29. Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M, et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature*. 2010;466:707–713. doi: 10.1038/nature09270.
30. Adeyemo A, Gerry N, Chen G, Herbert A, Doumatey A, Huang H, et al. A genome-wide association study of hypertension and blood pressure in African Americans. *PLoS Genet*. 2009;5:e1000564. doi: 10.1371/journal.pgen.1000564.
31. Li J, Glessner JT, Zhang H, Hou C, Wei Z, Bradfield JP, et al. GWAS of blood cell traits identifies novel associated loci and epistatic interactions in Caucasian and African-American children. *Hum Mol Genet*. 2013;22:1457–1464. doi: 10.1093/hmg/dds534.
32. Hao K, Chudin E, McElwee J, Schadt EE. Accuracy of genome-wide imputation of untyped markers and impacts on statistical power for association studies. *BMC Genet*. 2009;10:27. doi: 10.1186/1471-2156-10-27.
33. Woo D, Gebel J, Miller R, Kothari R, Brott T, Khoury J, et al. Incidence rates of first-ever ischemic stroke subtypes among blacks: a population-based study. *Stroke*. 1999;30:2517–2522.
34. Altshuler DM, Gibbs RA, Peltonen L, Dermitzakis E, Schaffner SF, Yu F, et al. Integrating common and rare genetic variation in diverse human populations. *Nature*. 2010;467:52–58.
35. Dichgans M, Markus HS. Genetic association studies in stroke: methodological issues and proposed standard criteria. *Stroke*. 2005;36:2027–2031. doi: 10.1161/01.STR.0000177498.21594.9e.

SUPPLEMENTAL MATERIAL

Meta-analysis of genome-wide association studies identifies genetic risk factors for stroke in African Americans

Supplemental Methods

Meta-Analysis

Additional checks for evidence of population stratification or structure were performed during the meta-analysis. The genomic inflation factors were estimated and determined to be 1.01 for the total stroke analysis, indicating minimal evidence of inflation of the test statistics due to stratification. As expected, the GC-adjusted test statistics were virtually identical to the unadjusted values (i.e. the top SNPs were consistent between both approaches), though p-values increased slightly.
In sensitivity analyses, we performed a meta-analysis of p-values using the weighted Z-score method implemented in METAL. This approach has reduced power compared with the effect size meta-analysis,[1] but does not necessarily require that effect estimates be measured on the same scale in all studies. Although the p-values from this analysis increased, in general, the ranking of SNPs was similar, with the exception of the GWAS-significant SNP, rs4471613. This SNP, which was only available in the cohort studies (ARIC, CHS, HABC and WHI) and attained GWAS-significance in the cohort only analysis, was no longer significant at the GWAS significance threshold in the p-value meta-analysis. Given the focus on discovery in this aim, interest in avoiding false negatives, and the subsequent replication effort, we present results from the effect estimate meta-analysis in the manuscript.

Study Descriptions

Atherosclerosis Risk in Communities (ARIC) Study
Population. The ARIC study is a prospective population-based study of atherosclerosis and clinical atherosclerotic diseases in 15,792 men and women, including 11,478 non-Hispanic white participants, drawn from 4 U.S. communities (Suburban Minneapolis, Minnesota; Washington County, Maryland; Forsyth County, North Carolina, and Jackson, Mississippi).[2] In the first three communities, the sample reflects the demographic composition of the community. In Jackson, only black residents were enrolled. Ancestry was self-reported during an interview. Participants were handed a card and asked to tell the interviewer which best described his or her race. Choices offered were: White, Black, American Indian/ Alaskan Native, Asian/Pacific Islander, Other: specify. Over 99% identified as either white or black. Only self-identified blacks were included in this study. Participants were between age 45 and 64 years at their baseline examination in 1987-1989 when blood was drawn for DNA extraction and participants consented to genetic testing.[2] Only individuals free of stroke or TIA at baseline were included in the analysis.

Genotyping. Single-nucleotide polymorphisms (SNPs) were genotyped on the Affymetrix 6.0 chip and were imputed to ≈2.5 million SNPs based on a panel of cosmopolitan reference haplotypes from HapMap CEU and YRI (HapMap II CEU and YRI (build 35, release 21)). MACH v1.0.16 was used to perform genotype imputations and allele dosage information was summarized in the imputation results. SNPs were excluded if they had no chromosomal location, were monomorphic, had a call rate<95%, or had a Hardy–Weinberg equilibrium P-value<10–5. For each SNP, the ratio of the observed versus expected variance of the dosage served as the measure of imputation quality. SNPs with MAF<1% and imputation quality <0.3 were filtered out prior to meta-analysis. Individuals with and without genotype data did not significantly differ with regards to baseline CVD risk factors (not shown). We excluded individuals with a sex mismatch, those who were discordant for more than 5% of genotypes among 47 previously genotyped overlapping SNPs, persons who were 1st degree relatives,

those who were outliers based on average identity by state or based on Eigenstrat clustering, or persons who had an incident subarachnoid hemorrhage.

Stroke Ascertainment. Hospitalized strokes that occurred by December 31, 2007 were included in the present study. During annual telephone contacts, trained interviewers asked each ARIC participant to list all hospitalizations during the past year. Hospital records for any hospitalizations identified were then obtained. In addition, all local hospitals annually provided lists of stroke discharges (International Classification of Diseases, Ninth Revision, Clinical Modification codes 430 to 438), which were scrutinized for ARIC participant discharges. Details on quality assurance for ascertainment and classification of stroke are described elsewhere.[3] Briefly, the stroke diagnosis was assigned according to criteria adapted from the National Survey of Stroke.[4] Strokes secondary to trauma, neoplasm, hematologic abnormality, infection, or vasculitis were excluded, and a focal deficit lasting <24 hours was not considered to be a stroke. Out-of-hospital stroke was not ascertained and validated; thus, these potential stroke events were not included. Strokes were classified into hemorrhagic stroke (subarachnoid and intracerebral hemorrhage) and ischemic stroke (thrombotic and embolic brain infarction). A stroke was classified as ischemic when a brain CT or MRI revealed acute infarction and showed no evidence of hemorrhage. All definite ischemic strokes were further classified as lacunar, nonlacunar thrombotic, or cardioembolic on the basis of the recorded neuroimaging results. For this analysis, the hemorrhagic strokes identified by ARIC were censored at the time of their occurrence.

Cardiovascular Health Study (CHS)

Population. CHS is a population-based cohort study of risk factors for coronary heart disease (CHD) and stroke in adults ≥ 65 years conducted across four field centers in the United States.[5] The original predominantly white cohort of 5,201 persons (4,964 whites) was recruited in 1989-1990 from a random sample of people on Medicare eligibility lists, and an additional 687 blacks were enrolled subsequently in 1992-93 for a total sample of 5,888. Race/ethnicity was determined by self-identification at interview.

Only self-identified blacks were included in these analyses. DNA was extracted from blood samples drawn on all participants who consented to genetic testing at their baseline examination.

Genotyping. In 2007-2008, genotyping was performed at the General Clinical Research Center's Phenotyping/Genotyping Laboratory at Cedars-Sinai using the Illumina 370CNV Duo®BeadChip system on participants who were free of CVD at baseline. Persons were also excluded for a subject-specific genotyping call rate ≤95%.

Stroke Ascertainment. Participants were examined annually from enrollment to 1999, and since then continue to be under surveillance for stroke.[6, 7] Since baseline, participants have also been contacted twice a year to identify potential cardiovascular events, including stroke. In addition, all hospitalizations were screened for potential stroke events. For suspected events, information was collected from the participant or next of kin, from medical records, and if needed, from the participant's physician. When available, CT scans, MRI scans and reports were reviewed centrally. At a consensus conference using all available information, vascular neurologists adjudicated all events and reached a final decision about the occurrence of stroke, stroke types and subtypes.

Genetics of Early Onset Stroke (GEOS)

Population. GEOS is a population-based case-control study designed to identify genes associated with early-onset ischemic stroke and to characterize interactions of identified stroke genes and/or SNPs with environmental risk factors.[8-10] Participants were recruited from the

greater Baltimore-Washington area in 4 different time periods: Stroke Prevention in Young Women-1 (SPYW-1) conducted from 1992-1996, Stroke Prevention in Young Women-2 (SPYW-2) conducted from 2001-2003, Stroke Prevention in Young Men (SPYM) conducted from 2003-2007, and Stroke Prevention in Young Adults (SPYA) conducted in 2008.

Stroke Ascertainment. Case participants were hospitalized with a first cerebral infarction identified by discharge surveillance from one of the 59 hospitals in the greater Baltimore-Washington area and direct referral from regional neurologists. The abstracted hospital records of cases were reviewed and adjudicated for ischemic stroke subtype by a pair of neurologists according to previously published procedures with disagreements resolved by a third neurologist. The ischemic stroke subtype classification system retains information on all probable and possible causes, and is reducible to the more widely used TOAST system that assigns each case to a single category. Control participants without a history of stroke were identified by random-digit dialing and were balanced to cases by age and region of residence in each recruitment periods.

Genotyping. Genomic DNA was isolated from a variety of sample types, including cell line, whole blood, mouth wash and buccal swab. Samples were genotyped at the Johns Hopkins Center for Inherited Disease Research (CIDR) using the Illumina HumanOmni1-Quad_v1-0_B BeadChip (Illumina, San Diego, CA, USA). Assistance with data cleaning was provided by the GENEVA Coordinating Center (U01-HG-004446; PI Bruce S Weir). Individuals were excluded if they were unexpected duplicates, gender discrepancy and unexpected relatedness.

The Healthy Aging in Neighborhoods of Diversity across the Life Span Study (HANDLS)
In the absence of non-stroke control samples from the VISP, ISGS, and SWISS studies, controls from the Healthy Aging in Neighborhoods of Diversity across the Life Span study (HANDLS) study were used for the VISP and SWISS-ISGS case-control analyses (with no overlap across studies). Controls were sex and race/ethnicity-matched and randomly selected from all HANDLS participants not reporting history of stroke at baseline or reporting adjudicated stroke during follow-up.

Population. HANDLS is an interdisciplinary, community-based, prospective longitudinal epidemiologic study examining the influences of race and socioeconomic status (SES) on the development of age-related health disparities among socioeconomically diverse African Americans and whites in Baltimore, MD, USA. This study assesses physical parameters over a 20-year period while evaluating genetic, biologic, demographic, and psychosocial influences. HANDLS recruited 3,722 participants (2200 African Americans (59%) and 1522 whites (41%)) from Baltimore, MD.

Stroke Ascertainment. Stroke status at baseline was determined through self-report while incident strokes, other vascular events, and deaths were determined using medical records and clinic visits during follow-up.

Genotyping. Genotyping was focused on a subset of participants self-reporting as African American and was performed at the Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health. Genotype data (for up to 907,763 SNPs) were generated for 1,024 participants using either Illumina 1M and 1M duo arrays (n=709) ,or a combination of 550K, 370K, 510S and 240S to equate the million SNP level of coverage. Inclusion criteria for genetic data in HANDLS includes concordance between self-reported sex and sex estimated from X chromosome heterogeneity, > 95% call rate per participant (across all equivalent arrays), concordance between self-reported African ancestry and ancestry confirmed by analyses of genotyped SNPs, and no cryptic relatedness to any other samples at a level of proportional

sharing of genotypes > 15% (effectively excluding 1st cousins and closer relatives from the set of probands used in analyses). In addition, SNPs included in the analysis were filtered for HWE p-value > 1e-7, missing by haplotype p-values > 1e-7, minor allele frequency > 0.01, and call rate > 95%. Data analyses utilized the high-performance computational capabilities of the Biowulf Linux cluster at the NIH, Bethesda, Md. (http://biowulf.nih.gov).

Dynamics of Health, Aging and Body Composition Study (Health ABC)

Population. The Dynamics of Health, Aging and Body Composition Study (Health ABC) enrolled a total of 3075 well-functioning men and women aged 70-79 between April 1997-June 1998.[11] This analysis includes African American individuals with available DNA from the baseline examination.

Stroke Ascertainment. Participants were questioned about any hospitalizations for stroke every 6 months. When an event was reported, hospital records were collected and verified by a Health ABC Disease Adjudicator at each site. Date and causes of death were taken from the death certificate. Follow-up time was defined as the time from the baseline visit until the first event date (for those who had an event) or was censored at the last contact date (for those who did not have any event or were lost to follow-up) or the day of death (for those who died of non-cardiovascular causes).

Genotyping. DNA samples were genotyped using the Illumina 1M array. Samples were excluded from the dataset for the reasons of sample failure, genotypic sex mismatch, and first-degree relative of an included individual based on genotype data, low call rate <97% or HWE p-value $<10^{-6}$.

Ischemic Stroke Genetic Study (ISGS)

Population. ISGS is a multicenter inception cohort study.[12] Cases were recruited from inpatient stroke services at five United States academic medical centers.

Stroke Ascertainment. Cases are adult men and women over the age of 18 years diagnosed with first-ever ischemic stroke confirmed by a study neurologist on the basis of history, physical examination and CT or MR imaging of the brain. Cases had to be enrolled within 30 days of onset of stroke symptoms. Cases were excluded if they had: a mechanical aortic or mitral valve at the time of the index ischemic stroke, central nervous system vasculitis, or bacterial endocarditis. They were also excluded if they were known to have: cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL), Fabry disease, homocystinuria, mitochondrial encephalopathy with lactic acidosis and stroke-like episodes (MELAS), or sickle cell anemia. Diagnostic evaluation included: head CT (95%) or MRI (83%), electrocardiography (92%), cervical arterial imaging (86%), and echocardiography (74%). Medical records from all cases were centrally reviewed by a vascular neurology committee and assigned ischemic stroke subtype diagnoses according to TOAST criteria,[13] the Oxfordshire Community Stroke Project[14], and the Baltimore-Washington Young Stroke Study[15]. DNA was donated to the NINDS DNA Repository (Coriell Institute, Camden, NJ) for eligible samples with appropriate written informed consent.

Genotyping. DNA samples were genotyped using the Illumina 610 array and data analyses were supported by the high-performance computational capabilities of the Biowulf Linux cluster at the NIH (http://biowulf.nih.gov).

Siblings with Ischemic Stroke Study (SWISS)

Population. SWISS is a prospective multicenter affected sibling pair study of first-ever or recurrent ischemic stroke.[16] Subjects were recruited from 54 enrolling hospitals across the US and Canada. Samples were collected between 1999-2011. Ischemic stroke probands were enrolled at 66 US medical centers and 4 Canadian medical centers.

Stroke Ascertainment. All recruits were extensively clinically phenotyped and have imaging-confirmed ischemic stroke using either CT or MRI brain scans. Probands are adult men and women over the age of 18 years diagnosed with ischemic stroke confirmed by a study neurologist on the basis of history, physical examination and CT or MR imaging of the brain who also have a history of at least one living sibling with a history of stroke. Probands were excluded if 1) they had a mechanical aortic or mitral valve at the time of the index ischemic stroke, central nervous system vasculitis, or bacterial endocarditis or 2) were known to have cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL), Fabry disease, homocystinuria, mitochondrial encephalopathy with lactic acidosis and stroke-like episodes (MELAS), or sickle cell anemia. Siblings were enrolled using proband-initiated contact[17] or direct contact when permitted by Institutional Review Boards. Concordant siblings had their diagnosis of ischemic stroke confirmed by review of medical records by a central vascular neurology committee. Concordant siblings had the same eligibility criteria as probands. Subtype diagnoses were assigned to the index strokes of probands and concordant siblings according to TOAST criteria. Discordant siblings of the proband were confirmed to be stroke-free using the Questionnaire for Verifying Stroke-free Status.[18]

Genotyping. DNA samples were genotyped using the Illumina 660 array and data analyses were supported by the high-performance computational capabilities of the Biowulf Linux cluster at the NIH (http://biowulf.nih.gov).

The Vitamin Intervention for Stroke Prevention (VISP) Study

Population. VISP is a multi-center, double-blind, randomized, controlled clinical trial that enrolled patients aged 35 or older with homocysteine levels above the 25th percentile at screening and a non-disabling cerebral infarction (NDCI) within 120 days of randomization.[19,20] The trial was designed to determine if daily intake of a multivitamin tablet with high dose folic acid, vitamin B6 and vitamin B12 reduced recurrent cerebral infarction (1° endpoint), and nonfatal myocardial infarction (MI) or mortality (2° endpoints). Subjects were randomly assigned to receive daily doses of the high-dose formulation (n=1,827), containing 25mg pyridoxine (B6), 0.4mg cobalamin (B12), and 2.5mg folic acid; or the low-dose formulation (n=1,853), containing 200µg pyridoxine, 6µg cobalamin and 20µg folic acid. Enrollment in VISP began in August 1997, and was completed in December 2001, with 3,680 participants enrolled, from 55 clinic sites across the US and Canada and one site in Scotland. A subset of VISP participants gave consent and were included in the GWAS component of VISP, supported by the National Human Genome Research Institute (NHGRI), Grant U01 HG005160, as part of the Genomics and Randomized Trials Network (GARNET)

Stroke Ascertainment. NDCI was defined as an ischemic brain infarction not due to embolism from a cardiac source, characterized by the sudden onset of a neurological deficit. The deficit must have persisted for at least 24 hours, or if not, an infarction in the part of the brain corresponding to the symptoms must have been demonstrated by CT or MRI imaging.

Genotyping. Samples were genotyped at the Center for Inherited Disease Research using the Illumina HumanOmni1-Quad_v1-0_B BeadChip (Illumina, San Diego, CA, USA).

Women's Health Initiative (WHI)

Population. The goal of the WHI is to investigate the etiology and prevention of chronic disease in post-menopausal women.[21] WHI recruited approximately 161, 0000 postmenopausal women 50–79 years of age from 40 clinical centers in the US between 1993 and 1998. WHI consists of an observational study (OS), and clinical trials (CT) of postmenopausal hormone therapy (estrogen alone or estrogen plus progestin), a calcium and vitamin D supplement trial, and a dietary modification trial. A subset of 8,515 AA women who provided consent for DNA analysis were randomly selected for genome-wide genotyping as part of the SNP Health Association Resource (SHARe) project.[22] Study protocols and consent forms were approved by the institutional review boards for all participating institutions.

Stroke Ascertainment. All incident strokes, other vascular events, and deaths were identified through self-report at annual (OS) and semi-annual (CT) participant contacts, and through third-party reports by family members and proxies. Medical records were obtained for potential strokes, and adjudication was performed by trained physician adjudicators who assigned a diagnosis. Stroke diagnosis requiring and/or occurring during hospitalization was based on rapid onset of a neurological deficit attributable to an obstruction or rupture of an arterial vessel system. The deficit was not known to be secondary to brain trauma, tumor, infection or other cause and must have lasted more than 24 hours unless death supervened or a lesion compatible with acute stroke was evident on computed tomography or magnetic resonance imaging scan.[23] Strokes were classified as ischemic, hemorrhagic or unknown/missing. Ischemic stroke subtypes were further classified using Trial of Org 10172 in Acute Stroke Treatment (TOAST) criteria. [13] For analyses, strokes subtypes judged as 'probable' or 'possible' were combined.

Genotyping. Genetic data were obtained from genome-wide scans using the Genome-wide Human SNP Array 6.0 (Affymetrix, Santa Clara, CA, www.affymetrix.com) of 909,622 single nucleotide polymorphisms (SNPs). Genotyping quality control included examination of concordance rates for blinded and un-blinded duplicates. Approximately 1% of SNPs failed genotyping and SNPs with call rates < 95% or concordance rates <98%, or minor allele frequency <1% were excluded. In addition to the genotyping, SNPs were imputed using MACH software[24] and the HapMapII CEU+YRI as reference populations. Out of 2,203,609 imputed SNPs, 2,192,719 SNPs had Rsq>0.3, the recommended cutoff for HapMap-based imputation, and were included in these analyses. Genotyping failed in 99 samples. Additional participants were excluded based on low call rates <95%, and sex or race/ethnicity discrepancy. First or second degree relatedness was assessed as described in Thornton and McPeek.[25] For the related individuals, the first or second degree relative with the highest call rate was retained in analyses and other family members were excluded. A principal-component (PC) analysis of all samples was performed using EIGENSTRAT[26]. The first 10 PCs were calculated for each individual and evaluated for their contribution to ancestral variation. Because most of the ancestral variation was explained by the first 4 PCs, only these were included as covariates in the analyses. We calculated lambda (λ), an indicator of over-dispersion due to potential population stratification by dividing the mean of the test statistics by the mean of the expected values from a Chi-squared distribution with 1 degree of freedom. Using λ, we investigated correction of p-values using genomic control to account for potential residual confounding by genetic ancestry.[27, 28]

AA women passing the above genotyping quality control criteria, with follow-up data, and without a history of stroke at baseline are included in these analyses.

METASTROKE Replication Cohort

To determine if stroke genetic variants identified in the COMPASS study would replicate in populations of European ancestry, we utilized the discovery sample population from the METASTROKE consortium.[29] The METASTROKE population consists of 15 cohorts with ischemic stroke cases and controls of European ancestry from Europe, North America, and Australia. Each case-control population was genotyped using either Affymetrix (Santa Clara, CA, USA) or Illumina (San Diego, CA, USA) GWAS SNP arrays. Genotyped data were imputed using the HapMap II reference populations. Logistic regression tests of association were performed for each individual case-control population, followed a meta-analysis using a fixed-effects inverse-variance weighted model as implemented in METAL.[30]

Supplemental Table I: Description of COMPASS discovery cohorts

	Cohort Studies				Case-Control Studies						
	ARIC	CHS	HABC	WHI	GEOS		ISGS	SWISS	HANDLS	VISP	HANDLS
Characteristic					**Cases**	**Controls**	**Cases**	**Cases**	**Controls***	**Cases**	**Controls***
Total N	2741	766	1139	7944	381	352	152	44	389	256	582
Female, N(%)	1757 (64%)	489 (64%)	651 (57%)	7944 (100%)	174 (46%)	156 (44%)	77 (51%)	22 (50%)	223 (57%)	116 (45.3%)	317 (55%)
Age in years, mean ± SD	53± 6	73±6	73±3	62±7	42±7	40±7	56±13	59±10	48 ± 9	63 ± 11	49 ± 9
BMI in kg/m^2, mean ± SD	30± 6	29±6	29±5	31± 6	32±8	29±7	N/A	N/A	30±9	30±7	30±8
Type 2 Diabetes, N (%)	444 (17%)	175 (24%)	291 (30%)	943 (12%)	87 (23%)	29 (8%)	48 (32%)	20 (45%)	66 (17%)	97 (38%)	115 (20%)
Hypertension, N (%)	1514 (55%)	554 (73%)	871 (77%)	4269 (54%)	215 (56%)	90 (26%)	126 (83%)	33 (75%)	207 (54%)	221 (86%)	295 (52%)
History of TIA, N (%)	0	11 (1%)	N/A	128 (2%)	N/A	N/A	21 (14%)	N/A	13 (4%)		20 (4%)
Lipid-lowering Medication Use, N (%)	33 (1%)	50 (7%)	129 (11%)	683 (9%)	N/A	N/A	N/A	N/A	44 (12%)	86 (34%)	77 (14%)
Recruitment Year(s)	1987-89	1989-90; 1992-3	1997-98	1993-98	1992-2008	1992-2007	2002-08	1999-2011	2004-present	Aug 1997-Dec 2001	2004-present
Cohort Studies											
Stroke Ascertainment (through year)	2007	2010	2008	2009							
Incident Ischemic Stroke Events, N	229	104	N/A	199							
Median Follow-up Time in years	10.3	11.6	N/A	11.9							
Total Stroke Events, N	265	124	128	242							
Median Follow-up Time in years	9.9	11.6	10.6	11.9							

*Samples from the HANDLS study were used as controls for the VISP and SWISS-ISGS case-control analyses (with no overlap across studies).

Supplemental Table II: SNPs nominally associated with ischemic or total stroke in the COMPASS discovery sample

SNP	Locus	Annotation	Alleles*	CAF (AA)	CAF (CEU)	Stroke Type	Beta† (SE)	P-value
rs1937787	1p31.1	intergenic *(ELTD1)*	T/C	0.84	0.67	total	-0.27 (0.05)	7.33×10^{-7}
rs11681884	2q13	intergenic *(IL1RN, IL1F10, IL36RN)*	T/C	0.43	0.11	total	-0.24 (0.05)	6.13×10^{-7}
rs704341	3p14.2	*PTPRG* (intron)	A/G	0.07	0.14	total	0.55 (0.11)	5.53×10^{-7}
						ischemic	0.62 (0.12)	7.11×10^{-7}
rs4867766	5q35	intergenic (*MSX2, NKX2-5*)	A/G	0.08	0.21	total	0.41 (0.08)	5.22×10^{-7}
rs7705819	5q35	intergenic (*SUMO2P6, GAPDHP71*)	T/C	0.64	0.75	total	0.23 (0.05)	9.47×10^{-7}
rs11572061	6p21.1	*CDC5L* (UTR3)	A/G	0.97	1.00	ischemic	-0.59 (0.12)	9.99×10^{-7}
rs9345396	6q16.1	intergenic (near *TSG1*)	T/C	0.01	0.01	total	1.75 (0.33)	1.03×10^{-7}
rs17347800	7p21	*HDAC9* (intron)	A/G	0.04	0.09	total	0.51 (0.10)	3.59×10^{-7}
rs768606	10p14	intergenic (near *LOC100507163*)	A/T	0.07	0.01	total	0.48 (0.09)	1.36×10^{-7}
rs17771318	10q11.2	*WDFY4* (intron)	A/G	0.01	0.04	total	1.46 (0.27)	8.94×10^{-8}
rs2084637	11q24	intergenic (*TRNAK27, GLULP3, UBASH3B)*	T/C	0.89	0.64	total	-0.32 (0.07)	9.12×10^{-7}
rs248812	12q23	intergenic (*RNU6-36, MIR135A2*)	A/C	0.83	0.76	total	-0.26 (0.05)	9.15×10^{-7}
rs1564060	14q31	intergenic	A/G	0.43	0.26	total	-0.24 (0.05)	2.67×10^{-7}
rs10400694	14q31	intergenic (near *FLRT2*)	C/G	0.52	0.75	ischemic	0.24 (0.05)	8.96×10^{-7}
rs7156510	14q31	intergenic	T/G	0.47	0.74	total	0.21 (0.04)	9.82×10^{-7}
rs4471613	15q21.3	intergenic (near *ALDH1A2, AQP9, LIPC*)	A/G	0.02	0.03	total	0.82 (0.15)	3.94×10^{-8}
					0.03	ischemic	0.90 (0.18)	4.62×10^{-7}
rs2822388	21q11.2	intergenic (near *LIP1, ABCC13*)	A/G	0.98	0.95	total	-0.86 (0.17)	4.96×10^{-7}
rs7283054	21q22.1	intergenic (*CLDN17)*	C/G	0.02	0.01	total	0.99 (0.20)	9.75×10^{-7}
rs5752326	22q12.1	*HPS4* (intron)	T/C	0.79	0.95	ischemic	-0.31 (0.06)	8.84×10^{-7}

CAF=coded allele frequency; AA=African Americans from COMPASS; CEU=HapMap II CEU population; SE=standard error

*coded allele/non-coded allele

†effect of each additional allele on risk of stroke (meta-analysis result)

Supplemental Table III: SNPs nominally associated with stroke in the COMPASS cohort studies

SNP	Locus	Annotation	Alleles*	CAF	Stroke Type	HR† (95%CI)	P-value
rs704341	3p14.2	*PTPRG* (intron)	A/G	0.07	total	1.73 (1.40 - 2.14)	5.53E-07
					isc	1.85 (1.45 - 2.36)	7.11E-07
rs781542	4q12	*SPINK2* (intron)	A/G	0.32	total	1.35 (1.20 - 1.53)	9.64E-07
rs6880837	5q31.2	*TGFBI* (intron)	T/C	0.28	total	1.33 (1.19 - 1.48)	3.88E-07
rs13168506	5q31.2	*TGFBI* (intron)	A/G	0.27	total	1.33 (1.19 - 1.49)	3.44E-07
rs9345396	6q16.1	3kb 3' of *TSG1*	T/C	0.01	total	5.78 (3.03 - 11.02)	1.03E-07
rs17145593	10p14	Intergenic (1.2Mb 3' of *GATA3)*	A/C	0.07	total	1.60 (1.34 - 1.92)	2.62E-07
rs768606	10p14	Intergenic (1.2Mb 3' of *GATA3)*	A/T	0.07	total	1.62 (1.35 - 1.94)	1.36E-07
rs17771318	10q11.22	*WDFY4* (intron)	A/G	0.01	total	4.29 (2.52 - 7.32)	8.94E-08
rs12291066	11p15.3	*MICAL2* (intron)	A/G	0.11	total	1.45 (1.25 - 1.69)	8.23E-07
rs4471613	15q21.3	Intergenic (near *ALDH1A2, AQP9, LIPC*)	**A/G**	**0.02**	**total**	**2.26 (1.69 - 3.03)**	**3.94E-08**
				0.02	isc	2.47 (1.74 - 3.50)	4.62E-07
rs12438353	15q25.3	*AGBL1* (intron)	T/C	0.50	isc	1.39 (1.23 - 1.57)	2.28E-07
rs9899375	**17p13.1**	***CHD3* (intron)**	**T/C**	**0.03**	**isc**	**3.30 (2.21 - 4.93)**	**5.23E-09**
					total	**2.57 (1.86 - 3.55)**	**1.20E-08**
rs2822388	21q11.2	55kb 5' of *C21orf81*	A/G	0.98	total	0.42 (0.30 - 0.59)	4.96E-07
rs7283054	21q21.3	19kb 3' of *CLDN17*	C/G	0.02	total	2.69 (1.81 - 4.00)	9.75E-07

CAF=coded allele frequency; 95%CI=95% confidence interval. Bolded SNPs are significant at GWAS discovery p<5e-8.

*coded allele/non-coded allele

† meta-analysis result of Cox proportional hazards models from cohort studies only, reflects the hazard of stroke for each coded allele

Supplementary Table IV: COMPASS replication of ischemic stroke SNPs reported in European-ancestry populations

Index SNP (annotation)	COMPASS SNP	LD (r^2) with Index SNP*	Allele/ AF†	Effect Size ± SE	P-value	N
rs13407662	**rs2111856**	**0.83**	**T/0.08**	**0.15 ± 0.10**	**0.1410**	**12,183**
(2p16/intergenic)	rs7594063	1.00	A/0.93	-0.07 ± 0.12	0.5551‡	11,450
	rs13421162	1.00	T/0.94	-0.06 ± 0.13	0.6394	11,450
rs1906599	**rs2634073**	**0.85**	**T/0.41**	**0.12 ± 0.05**	**0.0136**	**13,578**
(4q25/*PITX2)*	rs2634071	0.85	T/0.41	0.12 ± 0.05	0.0158	13,561
	rs2723334	0.97	T/0.48	0.1 ± 0.05	0.0520	13,582
	rs2466455	1.00	T/0.60	-0.12 ± 0.06	0.0618‡	11,450
	rs4611994	0.55	T/0.76	-0.08 ± 0.06	0.1383	13,580
	rs17042171	0.55	A/0.24	0.08 ± 0.06	0.1464	13,574
	rs2200733	0.55	T/0.23	0.08 ± 0.06	0.1950	13,561
	rs6843082	0.99	A/0.70	-0.07 ± 0.05	0.2227	13,571
	rs2634074	1.00	A/0.52	-0.07 ± 0.06	0.2792‡	11,450
	rs3853440	0.52	A/0.19	0.08 ± 0.08	0.2868‡	11,450
	rs2129983	1.00	A/0.56	-0.06 ± 0.06	0.3830‡	11,450
	rs2129982	0.99	A/0.71	-0.04 ± 0.05	0.3983	13,580
	rs12644625	0.54	T/0.11	0.08 ± 0.09	0.4002	10,806
	rs17042121	0.52	A/0.83	-0.05 ± 0.07	0.4942	13,582
	rs17042144	0.52	T/0.89	-0.03 ± 0.08	0.6682	13,577
	rs4032976	0.52	T/0.07	0.05 ± 0.13	0.6902‡	11,450
	rs13143308	1.00	T/0.30	0.02 ± 0.06	0.7033	12,183
	rs1906616	1.00	A/0.69	-0.03 ± 0.07	0.7047	11,450
	rs2220427	0.55	T/0.11	0.02 ± 0.08	0.7971	13,579
	rs4605724	0.52	A/0.11	0.02 ± 0.08	0.8085	13,582
	rs10516563	0.53	T/0.84	-0.02 ± 0.07	0.8252	12,178
	rs1906615	1.00	T/0.29	0.01 ± 0.07	0.9397	11,450
	rs12646447	0.55	T/0.89	-0.002 ± 0.08	0.9765	13,542
rs2200733	**rs2634073**	**0.55**	**T/0.41**	**0.12 ± 0.05**	**0.0136**	**13,578**
(4q25/*PITX2)*	rs2634071	0.56	T/0.41	0.12 ± 0.05	0.0158	13,561
	rs2723334	0.51	T/0.48	0.10 ± 0.05	0.0520	13,582
	rs1906599	0.55	T/0.43	0.08 ± 0.05	0.1041	13,582
	rs10019689	1.00	A/0.27	0.09 ± 0.06	0.1250	12,183
	rs4611994	1.00	T/0.76	-0.08 ± 0.06	0.1383	13,580
	rs17042171	1.00	A/0.24	0.08 ± 0.06	0.1464	13,574
	rs6533527	1.00	A/0.15	0.09 ± 0.07	0.2073	12,183
	rs1906617	1.00	A/0.85	-0.11 ± 0.09	0.2212	11,450
	rs6843082	0.55	A/0.70	-0.07 ± 0.05	0.2227	13,571
	rs11930528	0.78	T/0.27	0.05 ± 0.06	0.3278	13,578
	rs17042076	1.00	T/0.73	-0.06 ± 0.07	0.3590	11,450
	rs2129982	0.54	A/0.71	-0.04 ± 0.05	0.3983	13,580

	rs12644625	0.98	T/0.11	0.08 ± 0.09	0.4002	10,806
	rs2200732	0.91	T/0.76	-0.05 ± 0.07	0.4585	11,450
	rs6817105	1.00	T/0.76	-0.05 ± 0.07	0.4753	11,450
	rs1906593	1.00	T/0.24	0.05 ± 0.07	0.4804	11,450
	rs4540107	1.00	A/0.24	0.05 ± 0.07	0.4841	11,450
	rs17042121	0.96	A/0.83	-0.05 ± 0.07	0.4942	13,582
	rs2350269	1.00	T/0.13	0.06 ± 0.10	0.5009	11,450
	rs4626276	0.80	A/0.89	-0.05 ± 0.08	0.5498	13,583
	rs1906596	1.00	T/0.75	-0.04 ± 0.07	0.5938	11,450
	rs17042088	0.79	T/0.11	0.04 ± 0.08	0.5980	13,583
	rs1906592	1.00	T/0.77	-0.03 ± 0.07	0.6560	11,450
	rs17042144	0.96	T/0.89	-0.03 ± 0.08	0.6682	13,577
	rs12647316	0.79	T/0.14	0.03 ± 0.07	0.6967	13,583
	rs2129981	1.00	T/0.11	-0.04 ± 0.10	0.7168	11,450
	rs12639654	1.00	T/0.04	0.06 ± 0.17	0.7247	11,450
	rs17042059	1.00	A/0.13	0.02 ± 0.08	0.7601	12,183
	rs2220427	1.00	T/0.11	0.02 ± 0.08	0.7971	13,579
	rs4605724	0.96	A/0.11	0.02 ± 0.08	0.8085	13,582
	rs1906591	1.00	A/0.11	-0.02 ± 0.10	0.8104	11,450
	rs4543199	1.00	T/0.87	-0.02 ± 0.09	0.8233	11,450
	rs10516563	0.96	T/0.84	-0.02 ± 0.07	0.8252	12,178
	rs11098089	1.00	A/0.88	-0.02 ± 0.10	0.8504	11,450
	rs17042026	0.52	A/0.15	-0.02 ± 0.09	0.8541	11,450
	rs4529121	1.00	A/0.11	0.01 ± 0.10	0.9042	11,450
	rs12646754	1.00	T/0.13	-0.01 ± 0.11	0.9116	11,450
	rs17042098	1.00	A/0.12	0.01 ± 0.10	0.9190	11,450
	rs12646447	1.00	T/0.89	-0.002 ± 0.08	0.9765	13,542
rs6843082	**rs2634073**	**0.84**	**T/0.41**	**0.12 ± 0.05**	**0.0136**	**13,578**
(4q25/*PITX2)*	rs2634071	0.84	T/0.41	0.12 ± 0.05	0.0158	13,561
	rs2723334	0.96	T/0.48	0.1 ± 0.05	0.0520	13,582
	rs2466455	1.00	T/0.60	-0.12 ± 0.06	0.0618	11,450
	rs1906599	0.99	T/0.43	0.08 ± 0.05	0.1041	13,582
	rs4611994	0.55	T/0.76	-0.08 ± 0.06	0.1383	13,580
	rs17042171	0.55	A/0.24	0.08 ± 0.06	0.1464	13,574
	rs2200733	0.55	T/0.23	0.08 ± 0.06	0.1950	13,561
	rs2634074	1.00	A/0.52	-0.07 ± 0.06	0.2792	11,450
	rs2129983	1.00	A/0.56	-0.06 ± 0.06	0.3830	11,450
	rs2129982	1.00	A/0.71	-0.04 ± 0.05	0.3983	13,580
	rs12644625	0.53	T/0.11	0.08 ± 0.09	0.4002	10,806
	rs17042121	0.53	A/0.83	-0.05 ± 0.07	0.4942	13,582
	rs17042144	0.53	T/0.89	-0.03 ± 0.08	0.6682	13,577
	rs13143308	1.00	T/0.30	0.02 ± 0.06	0.7033	12,183
	rs1906616	1.00	A/0.69	-0.03 ± 0.07	0.7047	11,450

	rs2220427	0.54	T/0.11	0.02 ± 0.08	0.7971	13,579
	rs4605724	0.52	A/0.11	0.02 ± 0.08	0.8085	13,582
	rs10516563	0.52	T/0.84	-0.02 ± 0.07	0.8252	12,178
	rs1906615	1.00	T/0.29	0.01 ± 0.07	0.9397	11,450
	rs12646447	0.54	T/0.89	-0.002 ± 0.08	0.9765	13,542
rs556621	**rs658726**	**0.74**	**T/0.67**	**0.13 ± 0.07**	**0.0624**	**11,450**
(6p21.1/intergenic)	rs1767789	0.66	T/0.64	0.07 ± 0.05	0.1630	13,575
	rs646977	0.53	T/0.91	0.15 ± 0.12	0.2262	11,450
	rs504615	0.89	A/0.88	0.12 ± 0.11	0.2658	11,450
	rs1767788	0.89	T/0.88	0.12 ± 0.11	0.2689	11,450
	rs13202385	0.68	A/0.18	0.06 ± 0.06	0.3551	13,579
	rs1680900	0.89	A/0.10	-0.10 ± 0.11	0.3849	11,450
	rs12526438	0.68	A/0.18	0.05 ± 0.06	0.4002	13,579
	rs497177	0.84	T/0.10	-0.08 ± 0.10	0.4013	12,183
	rs9381341	0.88	C/0.81	-0.05 ± 0.08	0.5315‡	11,450
	rs4714797	0.88	C/0.82	-0.05 ± 0.08	0.5390	11,450
	rs9395035	0.88	A/0.82	-0.05 ± 0.08	0.5398	11,450
	rs9472313	0.96	A/0.10	-0.05 ± 0.11	0.6296	11,450
	rs4714796	0.88	C/0.15	-0.04 ± 0.09	0.6734‡	11,450
	rs632728	0.85	T/0.10	-0.03 ± 0.09	0.7257	13,581
	rs900403	0.65	A/0.33	-0.01 ± 0.05	0.8244	13,555
	rs4714801	0.91	A/0.91	0.01 ± 0.09	0.8927	13,574
	rs556512	1.00	A/0.23	0.003 ± 0.06	0.9607	13,569
rs11984041	**rs2526619**	**0.81**	**A/0.59**	**-0.08 ± 0.05**	**0.0935**	**13,542**
(7p21.1/*HDAC9*)	rs10241964	0.69	A/0.22	0.10 ± 0.07	0.1549	11,450
	rs10245779	0.85	C/0.23	0.08 ± 0.06	0.1573	13,554
	rs10255384	0.82	A/0.22	0.11 ± 0.08	0.1613	2,898
	rs2107595	0.52	A/0.21	0.09 ± 0.06	0.1701	12,183
	rs7783974	0.71	T/0.20	0.09 ± 0.08	0.2136	11,450
	rs2023937	0.84	A/0.78	-0.09 ± 0.08	0.2350	11,450
	rs7784712	0.85	T/0.16	0.08 ± 0.07	0.2457	13,579
	rs2023938	0.92	T/0.78	-0.07 ± 0.07	0.2553	12,182
	rs7792656	0.79	T/0.76	-0.06 ± 0.06	0.2982	12,183
rs2107595	**rs28688791**	**0.67**	**T/0.73**	**-0.12 ± 0.06**	**0.0486**	**10,830**
(7p21.1/*HDAC9*)	rs2526619	0.57	A/0.59	-0.08 ± 0.05	0.0935	13,542
	rs11984041	0.52	T/0.21	0.09 ± 0.06	0.1201	13,582
	rs2023936	0.69	C/0.67	-0.08 ± 0.05	0.1326	13,579
	rs10241964	0.60	A/0.22	0.10 ± 0.07	0.1549	11,450
	rs10245779	0.63	C/0.23	0.08 ± 0.06	0.1573	13,554
	rs10255384	0.57	A/0.22	0.11 ± 0.08	0.1613	2,898
	rs7798197	0.69	A/0.68	-0.09 ± 0.07	0.1714	11,450
	rs7783974	0.61	T/0.20	0.09 ± 0.08	0.2136	11,450
	rs7788833	0.66	T/0.69	-0.06 ± 0.05	0.2151	13,580

	rs7788972	0.75	A/0.30	0.08 ± 0.07	0.2266	11,450
	rs2023937	0.53	A/0.78	-0.09 ± 0.08	0.2350	11,450
	rs7784712	0.61	T/0.16	0.08 ± 0.07	0.2457	13,579
	rs2023938	0.52	T/0.78	-0.07 ± 0.07	0.2553	12,182
	rs2074633	0.66	T/0.73	-0.07 ± 0.06	0.2632	12,075
	rs7792656	0.56	T/0.76	-0.06 ± 0.06	0.2982	12,183
rs2383207	**rs1333040**	**0.55**	**T/0.62**	**0.10 ± 0.05**	**0.0487**	**13,582**
(9p21.3/*CDKN2A/B)*	rs6475606	0.85	T/0.88	0.12 ± 0.08	0.1227	13,578
	rs10116277	0.85	T/0.88	0.12 ± 0.08	0.1238	13,583
	rs1333047	0.87	A/0.12	-0.14 ± 0.10	0.1518	11,417
	rs1333049	0.85	C/0.24	0.08 ± 0.06	0.1520	13,580
	rs4977574	0.91	A/0.81	-0.08 ± 0.06	0.1987	13,569
	rs10757272	0.89	T/0.23	0.07 ± 0.06	0.2109	13,576
	rs1333048	0.93	A/0.70	-0.06 ± 0.05	0.2336	13,580
	rs10738610	0.97	A/0.79	0.09 ± 0.08	0.2379	11,450
	rs10757278	0.85	A/0.79	0.09 ± 0.08	0.2472	11,450
	rs10811650	0.60	A/0.79	-0.06 ± 0.06	0.2921	13,581
	rs2891168	0.91	A/0.79	-0.06 ± 0.06	0.3084	13,579
	rs1333046	0.97	A/0.26	-0.07 ± 0.07	0.3147	11,450
	rs944797	1.00	T/0.57	-0.05 ± 0.05	0.3340	13,582
	rs1537370	0.84	T/0.68	0.05 ± 0.05	0.3782	13,580
	rs1004638	1.00	A/0.12	0.08 ± 0.10	0.4212	11,450
	rs1537374	1.00	A/0.12	0.08 ± 0.10	0.4214	11,450
	rs10733376	1.00	C/0.88	-0.08 ± 0.10	0.4235	11,450
	rs1333043	0.97	A/0.88	-0.08 ± 0.10	0.4449	11,450
	rs7341786	0.97	A/0.12	0.08 ± 0.10	0.4487	11,450
	rs1412834	1.00	T/0.12	0.07 ± 0.10	0.4596	11,450
	rs7859362	1.00	T/0.12	0.07 ± 0.10	0.4613	11,450
	rs2184061	0.50	A/0.61	-0.05 ± 0.07	0.4684	11,450
	rs1333042	0.95	A/0.12	-0.06 ± 0.09	0.5017	12,177
	rs10738607	0.91	A/0.76	-0.04 ± 0.06	0.5121	13,583
	rs1537378	0.50	A/0.07	-0.13 ± 0.21	0.5359	2,741
	rs1333045	0.76	T/0.52	0.04 ± 0.07	0.5427	11,450
	rs10811647	0.59	C/0.80	0.05 ± 0.08	0.5607	11,450
	rs4977575	0.87	C/0.12	0.06 ± 0.10	0.5751	11,450
	rs9632884	0.74	C/0.89	-0.05 ± 0.11	0.6178	11,389
	rs1537371	0.90	A/0.88	-0.05 ± 0.10	0.6425	11,450
	rs1537375	0.98	T/0.33	-0.02 ± 0.05	0.6574	13,580
	rs7859727	0.90	T/0.74	-0.03 ± 0.08	0.6583	11,450
	rs1556516	0.90	C/0.88	-0.04 ± 0.10	0.6668	11,450
	rs1537373	0.90	T/0.12	0.04 ± 0.10	0.6792	11,450
	rs10738609	1.00	A/0.79	0.03 ± 0.07	0.6866	12,182
	rs2383206	1.00	A/0.57	-0.02 ± 0.06	0.6999	12,183

	rs10757269	0.83	A/0.16	-0.02 ± 0.09	0.8379	11,450
	rs10511701	0.97	T/0.29	0.002 ± 0.07	0.9813	11,450
rs11833579 (12p13.33/*NINJ2*)	**rs12425791**	**0.78**	**A/0.11**	**0.02 ± 0.07**	**0.7451**	**13,582**
rs12425791 (12p13.33/*NINJ2*)	**rs11833579**	**0.78**	**A/0.22**	**-0.03 ± 0.06**	**0.6419**	**13,583**
rs879324 (16q22.3/*ZFHX3)*	**rs16971456**	**0.51**	**C/0.85**	**0.13 ± 0.07**	**0.0482**	**13,582**
	rs9302644	0.73	C/0.21	-0.15 ± 0.09	0.0973	11,450
	rs7192350	0.64	A/0.77	0.08 ± 0.07	0.2543	11,450
	rs6416747	0.58	T/0.23	-0.08 ± 0.08	0.2589	11,450
	rs16971481	0.54	A/0.11	-0.08 ± 0.11	0.4465	2,899
	rs11643592	0.59	A/0.39	0.04 ± 0.06	0.5289	11,450
	rs7193343	0.71	T/0.24	-0.03 ± 0.06	0.5510	13,578
	rs719354	0.73	T/0.23	-0.02 ± 0.06	0.6796	13,583
	rs2106261	0.93	T/0.26	0.02 ± 0.05	0.7164	13,581
	rs8057081	0.70	T/0.28	-0.01 ± 0.05	0.8873	13,582
	rs12932445	1.00	T/0.86	-0.004 ± 0.07	0.9537	13,583
	rs4499262	0.95	A/0.14	0.001 ± 0.09	0.9931	2,899

AF=Allele frequency; N=number contributing to the analysis; the top SNPs in COMPASS for each locus are bolded and presented in the main text in Table 2.

*Reflects the numbers of SNPs (± 500kB) in moderate LD ($r^2 \geq 0.50$) with the index SNP in HapMap CEU, and available in COMPASS

†Coded allele and coded allele frequency in COMPASS studies contributing to the ischemic stroke analysis

‡Meta-analysis result for this SNP suggests heterogeneity of effect among contributing studies, (I^2>50%).

Supplemental References

1. Begum F, Ghosh D, Tseng GC, Feingold E. Comprehensive literature review and statistical considerations for GWAS meta-analysis. *Nucleic Acids Res*. 2012;40:3777-3784
2. The ARIC Investigators. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. *Am J Epidemiol*. 1989;129:687-702
3. Rosamond WD, Folsom AR, Chambless LE, Wang CH, McGovern PG, Howard G, et al. Stroke incidence and survival among middle-aged adults: 9-year follow-up of the Atherosclerosis Risk in Communities (ARIC) cohort. *Stroke*. 1999;30:736-743
4. The National Survey of Stroke. National Institute of Neurological and Communicative Disorders and Stroke. *Stroke*. 1981;12:I1-91
5. Fried LP, Borhani NO, Enright P, Furberg CD, Gardin JM, Kronmal RA, et al. The Cardiovascular Health Study: design and rationale. *Ann. Epidemiol.* 1991;1:263-276
6. Longstreth WT, Jr., Bernick C, Fitzpatrick A, Cushman M, Knepper L, Lima J, et al. Frequency and predictors of stroke death in 5,888 participants in the Cardiovascular Health Study. *Neurology*. 2001;56:368-375
7. Price TR, Psaty B, O'Leary D, Burke G, Gardin J. Assessment of cerebrovascular disease in the Cardiovascular Health Study. *Ann. Epidemiol.* 1993;3:504-507
8. MacClellan LR, Mitchell BD, Cole JW, Wozniak MA, Stern BJ, Giles WH, et al. Familial aggregation of ischemic stroke in young women: the Stroke Prevention in Young Women Study. *Genet. Epidemiol.* 2006;30:602-608
9. Kittner SJ, Stern BJ, Wozniak M, Buchholz DW, Earley CJ, Feeser BR, et al. Cerebral infarction in young adults: the Baltimore-Washington Cooperative Young Stroke Study. *Neurology*. 1998;50:890-894
10. Cheng YC, O'Connell JR, Cole JW, Stine OC, Dueker N, McArdle PF, et al. Genome-wide association analysis of ischemic stroke in young adults. *G3 (Bethesda)*. 2011;1:505-514
11. Wassel Fyr CL, Kanaya AM, Cummings SR, Reich D, Hsueh WC, Reiner AP, et al. Genetic admixture, adipocytokines, and adiposity in Black Americans: the Health, Aging, and Body Composition study. *Hum Genet*. 2007;121:615-624
12. Meschia JF, Brott TG, Brown RD, Jr., Crook RJ, Frankel M, Hardy J, et al. The Ischemic Stroke Genetics Study (ISGS) Protocol. *BMC neurology*. 2003;3:4
13. Adams HP, Jr., Bendixen BH, Kappelle LJ, Biller J, Love BB, Gordon DL, et al. Classification of subtype of acute ischemic stroke. Definitions for use in a multicenter clinical trial. TOAST. Trial of Org 10172 in Acute Stroke Treatment. *Stroke*. 1993;24:35-41
14. Bamford J, Sandercock P, Dennis M, Burn J, Warlow C. Classification and natural history of clinically identifiable subtypes of cerebral infarction. *Lancet*. 1991;337:1521-1526
15. Johnson CJ, Kittner SJ, McCarter RJ, Sloan MA, Stern BJ, Buchholz D, et al. Interrater reliability of an etiologic classification of ischemic stroke. *Stroke*. 1995;26:46-51
16. Meschia JF, Brown RD, Jr., Brott TG, Chukwudelunzu FE, Hardy J, Rich SS. The Siblings With Ischemic Stroke Study (SWISS) protocol. *BMC Med Genet*. 2002;3:1
17. Worrall BB, Chen DT, Meschia JF. Ethical and methodological issues in pedigree stroke research. *Stroke*. 2001;32:1242-1249
18. Meschia JF, Brott TG, Chukwudelunzu FE, Hardy J, Brown RD, Jr., Meissner I, et al. Verifying the stroke-free phenotype by structured telephone interview. *Stroke*. 2000;31:1076-1080
19. Spence JD, Howard VJ, Chambless LE, Malinow MR, Pettigrew LC, Stampfer M, et al. Vitamin Intervention for Stroke Prevention (VISP) trial: rationale and design. *Neuroepidemiology*. 2001;20:16-25
20. Toole JF. Vitamin intervention for stroke prevention. *J. Neurol. Sci.* 2002;203-204:121-124

21. Women's Health Initiative Study Group. Design of the Women's Health Initiative clinical trial and observational study. *Control Clin Trials*. 1998;19:61-109
22. Carty CL, Johnson NA, Hutter CM, Reiner AP, Peters U, Tang H, et al. Genome-wide association study of body height in African Americans: the Women's Health Initiative SNP Health Association Resource (SHARe). *Hum mol genet*. 2012;21:711-720
23. Wassertheil-Smoller S, Hendrix SL, Limacher M, Heiss G, Kooperberg C, Baird A, et al. Effect of estrogen plus progestin on stroke in postmenopausal women: the Women's Health Initiative: a randomized trial. *JAMA : the journal of the American Medical Association*. 2003;289:2673-2684
24. Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet. Epidemiol.* 2010;34:816-834
25. Thornton T, McPeek MS. ROADTRIPS: case-control association testing with partially or completely unknown population and pedigree structure. *Am. J. Hum. Genet.* 2010;86:172-184
26. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*. 2006;38:904-909
27. Devlin B, Roeder K. Genomic control for association studies. *Biometrics*. 1999;55:997-1004
28. Bacanu SA, Devlin B, Roeder K. Association studies for quantitative traits in structured populations. *Genet. Epidemiol.* 2002;22:78-93
29. Traylor M, Farrall M, Holliday EG, Sudlow C, Hopewell JC, Cheng YC, et al. Genetic risk factors for ischaemic stroke and its subtypes (the METASTROKE collaboration): a meta-analysis of genome-wide association studies. *Lancet neurol*. 2012;11:951-962
30. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*. 2010;26:2190-2191

Meta-Analysis of Genome-Wide Association Studies Identifies Genetic Risk Factors for Stroke in African Americans

Cara L. Carty, Keith L. Keene, Yu-Ching Cheng, James F. Meschia, Wei-Min Chen, Mike Nalls, Joshua C. Bis, Steven J. Kittner, Stephen S. Rich, Salman Tajuddin, Alan B. Zonderman, Michele K. Evans, Carl D. Langefeld, Rebecca Gottesman, Thomas H. Mosley, Eyal Shahar, Daniel Woo, Kristine Yaffe, Yongmei Liu, Michèle M. Sale, Martin Dichgans, Rainer Malik, W.T. Longstreth Jr, Braxton D. Mitchell, Bruce M. Psaty, Charles Kooperberg, Alexander Reiner, Bradford B. Worrall and Myriam Fornage
on behalf of the COMPASS and METASTROKE Consortia

Stroke. 2015;46:2063-2068; originally published online June 18, 2015;
doi: 10.1161/STROKEAHA.115.009044
Stroke is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231

Print ISSN: 0039-2499. Online ISSN: 1524-4628