Stroke

CLINICAL AND POPULATION SCIENCES

Genome-Wide Association Study Meta-Analysis of Stroke in 22 000 Individuals of African Descent Identifies Novel Associations With Stroke

Keith L. Keene, PhD*; Hyacinth I. Hyacinth, MD, PhD*; Joshua C. Bis, PhD; Steven J. Kittner, MD, MPH; Braxton D. Mitchell, PhD, MPH; Yu-Ching Cheng, PhD; Guillaume Pare, MD, MSc; Michael Chong, MSc; Martin O'Donnell, MD, PhD; James F. Meschia, MD; Wei-Min Chen, PhD; Michèle M. Sale, PhD; Stephen S. Rich, PhD; Mike A. Nalls, PhD; Alan B. Zonderman, PhD; Michele K. Evans, MD; James G. Wilson, MD; Adolfo Correa, MD, PhD; Hugh S. Markus, FMed Sci; Matthew Traylor, PhD; Cathryn M. Lewis, PhD; Cara L. Carty, PhD; Alexander Reiner, MD, MS; Jeff Haessler, MS; Carl D. Langefeld, PhD; Rebecca Gottesman, MD, PhD; Thomas H. Mosley, PhD; Daniel Woo, MD, MS; Kristine Yaffe, MD; YongMei Liu, MD, PhD; William T. Longstreth, MD, MPH; Bruce M. Psaty, MD, PhD; Charles Kooperberg, PhD; Leslie A. Lange, PhD; Ralph Sacco, MD, MS; Tatjana Rundek, MD, PhD; Jin-Moo Lee, MD, PhD; Carlos Cruchaga, PhD; Karen L. Furie, MD; Donna K. Arnett, PhD; Oscar R. Benavente, MD; Raji P. Grewal, MD; Leema Reddy Peddareddygari, MD; Martin Dichgans, MD; Rainer Malik, PhD; Bradford B. Worrall, MD, MSc; Myriam Fornage, PhD; on behalf of the COMPASS, SiGN, and METASTROKE Consortia

BACKGROUND AND PURPOSE: Stroke is a complex disease with multiple genetic and environmental risk factors. Blacks endure a nearly 2-fold greater risk of stroke and are 2× to 3× more likely to die from stroke than European Americans.

METHODS: The COMPASS (Consortium of Minority Population Genome-Wide Association Studies of Stroke) has conducted a genome-wide association meta-analysis of stroke in >22 000 individuals of African ancestry (3734 cases, 18 317 controls) from 13 cohorts.

RESULTS: In meta-analyses, we identified one single nucleotide polymorphism (rs55931441) near the *HNF1A* gene that reached genome-wide significance (P=4.62×10^{-8}) and an additional 29 variants with suggestive evidence of association (P<1×10^{-6}), representing 24 unique loci. For validation, a look-up analysis for a 100 kb region flanking the COMPASS single nucleotide polymorphism was performed in SiGN (Stroke Genetics Network) Europeans, SiGN Hispanics, and METASTROKE (Europeans). Using a stringent Bonferroni correction P value of 2.08×10^{-3} (0.05/24 unique loci), we were able to validate associations at the *HNF1A* locus in both SiGN (P=8.18×10^{-4}) and METASTROKE (P=1.72×10^{-3}) European populations. Overall, 16 of 24 loci showed evidence for validation across multiple populations. Previous studies have reported associations between variants in the *HNF1A* gene and lipids, C-reactive protein, and risk of coronary artery disease and stroke. Suggestive associations with variants in the *SFXN4* and *TMEM108* genes represent potential novel ischemic stroke loci.

CONCLUSIONS: These findings represent the most thorough investigation of genetic determinants of stroke in individuals of African descent, to date.

Key Words: brain ischemia ■ coronary artery disease ■ genome-wide association study ■ meta-analysis ■ phenotype ■ risk factors

Correspondence to: Keith L. Keene, PhD, East Carolina University, Brody School of Medicine Center for Health Disparities, Department of Biology 1800 W 5th St, Medical Pavilion Suite 6; Mai Stop 643, Greenville NC 27834. Email keenek@ecu.edu

This manuscript was sent to Jean-Claude Baron, Consulting Editor, for review by expert referees, editorial decision, and final disposition

*Drs Keene and Hyacinth contributed equally as joint first authors.

The Data Supplement is available with this article at https://www.ahajournals.org/doi/suppl/10.1161/STROKEAHA.120.029123.

For Sources of Funding and Disclosures, see page 2461.

Stroke is available at www.ahajournals.org/journal/str

Nonstandard Abbreviations and Acronyms

1000G	1000 genomes
ARIC	Atherosclerosis Risk in Communities
CHS	Cardiovascular Health Study
CIDR	Center for Inherited Disease Research
COMPASS	Consortium of Minority Population Genome-Wide Association Studies of Stroke
GEOS	Genetics of Early Onset Stroke
GWAS	genome-wide association study
HANDLS	Healthy Aging in Neighborhoods of Diversity across the Life Span
HNF1A	HNF1 homeobox A
ISGS	Ischemic Stroke Genetics Study
JHS	Jackson Heart Study
NINDS	National Institute of Neurological Disorders and Stroke
REGARDS	Reasons for Geographic and Racial Differences in Stroke
SiGN	Stroke Genetics Network
SIGNET	Sea Islands Genetics Network
SLESS	South London Ethnicity and Stroke Study
SNP	single nucleotide polymorphism
SWISS	Siblings with Ischemic Stroke Study
VISP	Vitamin Intervention for Stroke Prevention
WHI	Women's Health Initiative

Stroke is the second leading cause of death worldwide and a leading cause of long-term disability in the United States.[1] Stroke is a heterogeneous disease encompassing multiple subtypes with unique etiologies and risk factors.[2] Nearly 87% of the ≈795 000 strokes that occur each year in the United States are ischemic.[1] Epidemiological studies suggest a substantial genetic component for stroke with overall heritability estimates of 38% for all ischemic strokes, and subtype-specific estimates of 20% to 25% for small vessel disease[3] and up to 40% for large-vessel disease.[4] Compared with European Americans, blacks have a nearly 2-fold greater risk of incident stroke, >2-fold increased risk of fatal stroke, strokes at younger ages, and higher frequency of poststroke disability.[5,6] Despite this disproportionate burden, few attempts to map stroke susceptibility loci have focused on individuals of African ancestry.[7] Recent genome-wide association studies (GWAS) have identified several stroke susceptibility loci[8–14] primarily in individuals of European ancestry with little success replicating in non-European–ancestry populations[7,13,15,16] possibly due to differences in the genetic architecture of stroke among individuals of diverse ancestry.

This study represents a collective effort to investigate the genetic basis of stroke by mapping stroke susceptibility loci potentially unique to individuals of African ancestry. Using data obtained from the COMPASS (Consortium of Minority Population Genome-Wide Association Studies of Stroke), we expand upon our discovery GWAS meta-analysis of stroke in blacks[7] using 1000 genomes (1000G) imputed data in 22 000 individuals.

METHODS

To minimize the possibility of unintentionally sharing information that can be used to reidentify private information, a subset of the data generated for this study are available at the database of Genotypes and Phenotypes (dbGaP) and can be accessed at https://www.ncbi.nlm.nih.gov/gap/.

Study Population

COMPASS included a total of 22 051 individuals of African descent with either a physician-adjudicated stroke (n=3734) or no history of stroke (n=18 317; Table I in the Data Supplement) and genome-wide single nucleotide polymorphism (SNP) data. Participating studies include prospective cohorts (ARIC study [Atherosclerosis Risk in Communities],[17] CHS [Cardiovascular Health Study],[18] JHS [Jackson Heart Study],[19,20] the WHI [Women's Health Initiative],[21]; case-control studies (INTERSTROKE,[22] REGARDS [Reasons for Geographic and Racial Differences in Stroke],[23] ISGS [Ischemic Stroke Genetics Study],[24] VISP [Vitamin Intervention for Stroke Prevention],[25,26] SLESS [South London Ethnicity and Stroke Study],[27] the GEOS Study [Genetics of Early Onset Stroke],[28] the NINDS-SiGN [National Institute of Neurological Disorders and Stroke–Stroke Genetics Network],[29] HANDLS [Healthy Aging in Neighborhoods of Diversity Across the Life Span][30]); and an affected sib pair study—SWISS (Siblings With Ischemic Stroke Study).[31] Race/ethnicity-matched 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 participants provided written, informed consent, and institutional review boards approved each of the respective studies/institutions.

Outcomes

We defined stroke as a focal neurological deficit of presumed vascular cause with a sudden onset and lasting 24 hours or until death with clinical or radiological (computed tomography/magnetic resonance imaging) evidence with stroke diagnosis made when there is overwhelming clinical evidence in the absence of radiological evidence of a cerebral infarction. A lack of imaging data for all stroke cases does not increase the likelihood of false positives in our study. The cohort studies only considered first (incident) clinically validated ischemic strokes. Individuals with a baseline history of ischemic or hemorrhagic stroke were excluded.

Genotype Data

All studies imputed SNPs using 1000G Phase I Version 3 Haplotypes, except SLESS and WHI, which used 1000G Phase III data (1KGp3) reference populations. We excluded

Table 1. COMPASS Ischemic Stroke Suggestive and Genome-Wide Significant Inverse Variance Weighted Associations

Chr	Position*	Gene	SNP	Alleles (Coded/ Noncoded)	Beta	SE	Odds Ratio (CI)	Inverse Variance Weighted *P* Value	Direction	Het *P* Value	Sample Size	No. of Studies
1	112853017	*CTTNBP2NL* (nearest)	rs114947355	T/C	0.44	0.0902	1.56 (1.42–1.70)	9.05×10^{-07}	????–+?+???	0.1382	12610	3
1	112857084	*CTTNBP2NL* (nearest)	rs147779128	A/T	−0.46	0.0945	0.63 (0.57–0.69)	9.61×10^{-07}	?????–?–???	0.9293	9637	2
2	4083658	*NPM1P48* (nearest)	rs142655108	A/C	0.58	0.1089	1.79 (1.60–1.99)	9.52×10^{-08}	?????+?+???	0.2834	9637	2
2	198551159	*RFTN2* and *MARS2* (nearest)	rs115670077	T/G	0.35	0.072	1.43 (1.33–1.53)	8.48×10^{-07}	+?++++?+???	0.5735	16540	6
3	124048486	*KALRN*	rs72976591	A/C	0.17	0.0342	1.18 (1.14–1.22)	9.19×10^{-07}	++++++++++	0.5356	22018	11
3	133101791	*TMEM108*	rs113509723	–/AA	0.45	0.0841	1.58 (1.45–1.71)	6.46×10^{-08}	?????+?+???	0.2014	9637	2
3	153125290	*AK092619* (nearest)	rs184221467	A/G	0.62	0.1246	1.85 (1.63–2.10)	7.86×10^{-07}	?????+?+???	0.468	9637	2
4	99435032	*TSPAN5*	rs138134155	A/G	0.36	0.0705	1.43 (1.33–1.53)	3.94×10^{-07}	+?++++?++??	0.9442	18531	7
5	101123995	*OR7H2P* (nearest)	rs77460585	A/G	0.59	0.1165	1.80 (1.60–2.02)	4.36×10^{-07}	????–??+???	0.004981	10940	2
5	150981704	*FAT2* and *SPARC* (nearest)	rs114527838	A/G	−0.28	0.055	0.76 (0.72–0.80)	5.55×10^{-07}	–?–––––––??	0.7033	19032	8
6	97345991	*KLHL32* and *NDUFA4* (nearest)	rs146522546	–/CT	−0.45	0.0876	0.64 (0.58–0.69)	2.22×10^{-07}	???–––?–???	0.3829	13353	4
7	83432409	*SEMA3A*	rs6967981	T/G	0.15	0.0296	1.16 (1.12–1.19)	7.57×10^{-07}	+–+++++++––	0.1685	21970	11
8	1572874	*DLGAP2*	rs112455974	A/C	0.68	0.1336	1.97 (1.72–2.25)	3.77×10^{-07}	????+??+???	0.7366	10949	2
9	72475192	*C9orf135*	rs565295967	T/C	0.62	0.1199	1.86 (1.65–2.09)	2.41×10^{-07}	?????+?+???	0.1048	9637	2
10	53545098	*PRKG1*	rs140164788	T/C	0.52	0.1019	1.68 (1.52–1.86)	3.37×10^{-07}	????++?+???	0.7146	12618	3
10	53547264	*PRKG1*	rs74469072	T/G	0.52	0.1018	1.68 (1.52–1.86)	3.50×10^{-07}	????++?+???	0.7169	12618	3
10	120907173	*SFXN4*	rs150807690	–/G	−0.20	0.0378	0.82 (0.79–0.85)	9.67×10^{-08}	?–?–––?––––	0.3014	18180	8
11	11360296	*GALNT18*	rs115825287	T/C	0.35	0.0696	1.43 (1.33–1.53)	3.60×10^{-07}	??++++?+???	0.6076	15673	5
11	75683895	*UVRAG*	rs368167310	T/C	−0.55	0.1085	0.58 (0.52–0.65)	4.87×10^{-07}	?????–?–???	0.8172	9637	2
12	29288407	*FAR2* (nearest)	rs113025543	A/T	−0.27	0.0551	0.76 (0.72–0.81)	9.23×10^{-07}	–+–––––––?	0.7896	20224	10
12	29292793	*FAR2* (nearest)	rs142100833	C/G	0.24	0.0488	1.27 (1.21–1.34)	8.65×10^{-07}	+–++++++–?	0.4482	20119	10
12	29341407	*FAR2*	–	–/??	0.65	0.1272	1.91 (1.68–2.17)	3.79×10^{-07}	???++???+??	0.9784	5542	3
12	119502791	*SRRM4*	rs531465435	–/C	0.59	0.1162	1.81 (1.61–2.03)	3.39×10^{-07}	?????+?+???	0.5809	9637	2
12	119542751	*SRRM4*	rs192977447	A/T	0.43	0.0816	1.53 (1.41–1.66)	1.80×10^{-07}	???+++?++??	0.1962	15333	5
12†	121415209	*HNF1A* (nearest)	rs55931441	A/G	0.52	0.0947	1.68 (1.53–1.84)	4.62×10^{-08}	?????+?+???	0.4599	9637	2
14	93788855	*BTBD7*	rs113949028	–/G	0.20	0.0396	1.22 (1.17–1.27)	5.44×10^{-07}	?+?+++?++++	0.948	18255	8
18	68475060	*GTSCR1* (nearest)	rs181095590	A/G	0.58	0.1138	1.78 (1.59–2.00)	3.90×10^{-07}	?????+?+???	0.4538	9637	2

(*Continued*)

Table 1. Continued

Chr	Position*	Gene	SNP	Alleles (Coded/ Noncoded)	Beta	SE	Odds Ratio (CI)	Inverse Variance Weighted *P* Value	Direction	Het *P* Value	Sample Size	No. of Studies
19	29710081	*UQCRFS1* (nearest)	rs73923591	A/G	0.27	0.0548	1.31 (1.24–1.39)	6.18×10^{-07}	++++++++++?	0.8774	20246	10
21	36442465	*RUNX1*	rs116262092	A/T	−0.58	0.1174	0.56 (0.50–0.63)	7.04×10^{-07}	????−−?−???	0.9789	12581	3
21	36443919	*RUNX1*	rs147867382	C/G	−0.58	0.1174	0.56 (0.50–0.63)	7.95×10^{-07}	????−−?−???	0.9792	12579	3

Direction indicates the direction of the effect size: negative (−), neutral/unknown (./?), and positive (+) for each contributing cohort/population. Chr indicates chromosomes; COMPASS, Consortium of Minority Population Genome-Wide Association Studies of Stroke; Het, heterogeneity; and SNP, single nucleotide polymorphism.

*Chr position based on human genome (GRCh37/hg19).

†Genome-wide significance ($P<5\times10^{-8}$).

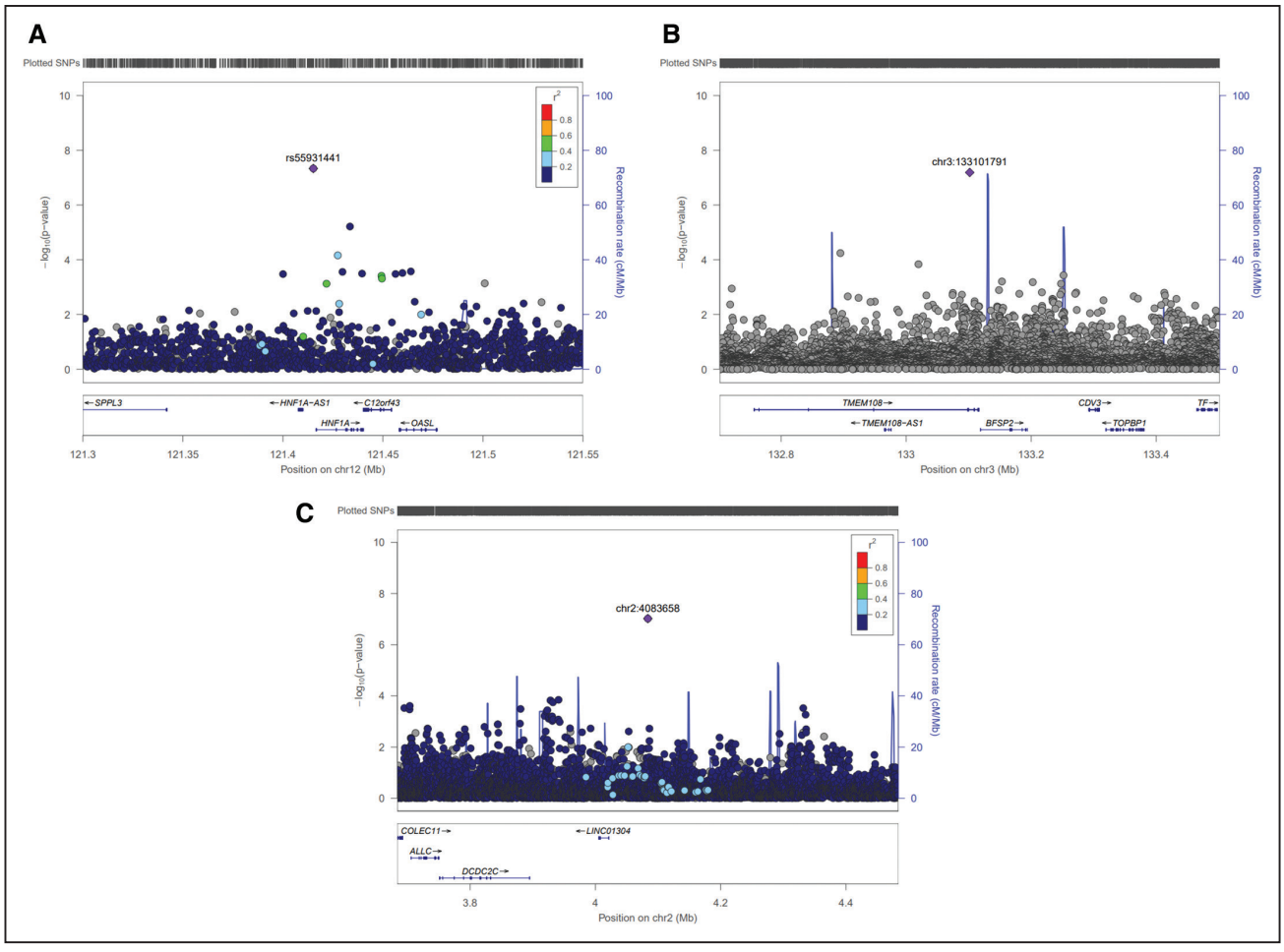


Figure. LocusZoom plots, with linkage disequilibrium based on hg19/1000 Genomes Nov 2014 AFR, depicting the top ($P=10^{-8}$) 3 associations with ischemic stroke in COMPASS (Consortium of Minority Population Genome-Wide Association Studies of Stroke) individuals of African descent.

A, *HNF1A* (rs55931441) chromosome (chr) 12 locus; (**B**) *TMEM108* (rs113509723) chr3 locus; (**C**) chr2 (rs142655108) locus nearest *NPM1P48*. SNP indicates single nucleotide polymorphism.

SNPs if they had invalid or missing alleles, *P* values, or β values; had minor allele frequencies <1%; imputation quality (r^2) <0.3; or were located on sex chromosomes. We analyzed SNPs available in ≥2 studies, for a total of ≈16.9 million SNPs. The Data Supplement contains study-specific details about design, stroke definition, adjudication procedures, and genotyping.

Analysis

We used logistic regression (additive genetic model) analyses with a count of variant alleles (0, 1, or 2) for each genotyped SNP or allelic dose for imputed SNPs. To control for potential population stratification, we included estimated study-specific principal components of global ancestry as covariates. As

Table 2. Genome-Wide and Suggestive COMPASS Associations With Look-Ups in European and Hispanic Populations From SiGN and METASTROKE

Chr	Unique Locus	Top SiGN European SNP	Alleles	*Z* Score	*P* Value	Direction	Top SiGN Hispanic SNP
1	*CTTNBP2NL* (nearest)	rs186896391	C/A	−3.28	0.0010*	−−−−−−+++−.	rs3121986
2	*NPM1P48* (nearest)	2-4077298 (rs527602504)	TC/T	2.56	0.0104	.....+....	rs60037207
2	*RFTN2* and *MARS2* (nearest)	2-198592085 (rs543821034)	C/T	2.98	0.0029	....+.....	rs150235598
3	*KALRN*	rs2034173	T/C	2.99	0.0027	+....+.....	rs185731506
3	*TMEM108*	rs13087036	C/A	−2.52	0.0116	−−+−−−+−−−	rs139695007
3	*AK092619* (nearest)	rs183598421	T/C	−2.36	0.0185	....−.....	rs200248409
4	*TSPAN5*	rs28392914	T/G	−3.16	0.0016*	+−−−−−−−−+	rs1045655
5	*OR7H2P* (nearest)	rs139061870	GT/G	2.80	0.0052	.....+..	rs73776672
5	*FAT2* and *SPARC* (nearest)	rs141575897	G/A	−3.03	0.0024	−−−−+..−−−−	rs80009114
6	*KLHL32* and *NDUFA4* (nearest)	rs200056339	C/CA	−2.68	0.0074	..−..−.−−..	rs78235656
7	*SEMA3A*	rs151172774	T/C	2.76	0.0058	+++++++−+++	rs6955094
8	*DLGAP2*	rs117175403	G/A	2.79	0.0053	−++++++−−++.	rs184526444
9	*C9orf135*	rs56179412	C/T	−2.13	0.0330	−−−−++−+−+−	rs77797545
10	*PRKG1*	rs10999787	C/A	−2.70	0.0069	−−−−++−−−−.	rs10998992
10	*SFXN4*	rs143931152	T/G	−3.64	0.0003*	−−−−−−−−−−.	rs56095167
11	*GALNT18*	rs117835740	C/T	−2.45	0.0142	.−.−.−.−.−−−	rs11021735
11	*UVRAG*	11-75761242 (rs565239444)	T/G	−2.76	0.0058	.....−....	rs138825035
12	*FAR2* (nearest)	rs151183596	T/A	−2.70	0.0070	−+−−−−.+−−−	rs141911197
12	*SRRM4*	rs61937966	C/T	3.37	0.0007*	+++++−+++++	rs4767761
12	*HNF1A* (nearest)*	rs182546302	T/A	−3.35	0.0008*	−+−−−+..−+.	rs80019595
14	*BTBD7*	rs112848587	C/T	−2.19	0.0284	....−−−.−.−	rs76789831
18	*GTSCR1* (nearest)	rs11151610	T/C	−3.27	0.0011*	−−−−−−−−−−	rs75968601
19	*UQCRFS1* (nearest)	rs148613358	T/C	3.22	0.0013*	.....++..+.	rs12608817
21	*RUNX1*	rs7280028	T/C	−3.42	0.0006*	−−−−−−−−−−	rs9981811

(*Continued*)

appropriate, we adjusted models for age, sex, and study site. We combined study-specific results in a fixed-effects meta-analyses with inverse variance weighting using METAL.[32] We also performed sample size weighted meta-analysis as an alternative approach to inverse variance weighting (Table II in the Data Supplement). We set a genome-wide significance (discovery) threshold of $P<5\times10^{-8}$ but investigated all SNPs with $P<10^{-6}$.

Validation of COMPASS Findings

Due to the absence of a comparable and adequately powered cohort of blacks with GWAS and adjudicated stroke data, we performed a look-up of COMPASS SNPs with $P<10^{-6}$ in the SiGN European and Hispanic ischemic stroke populations and METASTROKE total ischemic stroke populations (Table III in the Data Supplement). Additional METASTROKE subtype (cardioembolic, large-vessel, and small vessel) specific look-up analyses were performed to further validate these findings. Given the known differences in linkage disequilibrium patterns between populations of European and African ancestry, we expanded the region of interest for each locus to include available SNPs±100 kb of the index COMPASS SNPs as previously described[7] applying a Bonferroni correction to account for the number of loci tested.

RESULTS

Discovery of Stroke-Associated Loci

Using inverse variance weighting meta-analyses (Table 1), we identified one genome-wide significant association ($P<5\times10^{-8}$) and an additional 29 variants with suggestive evidence of association ($P<1\times10^{-6}$), representing 24 unique loci in total. The genome-wide significant association was detected upstream of the HNF1 homeobox A (*HNF1A*) gene on chromosome 12 (rs55931441; $P=4.62\times10^{-8}$, odds ratio, 1.68; Figure [A]).

Table 2. Continued

Alleles	*Z* Score	*P* Value	Direction	METASTROKE Top SNP	Alleles	Effect	*P* Value	Direction
A/G	−2.79	0.0052	−	rs10158830	C/G	0.073	0.0019*	++++++−+++++
T/C	−2.21	0.0268	−	rs114152357	A/T	−0.186	0.0048	−−+−++−−−−−−
G/A	−2.74	0.0061	−	rs191948652	A/T	0.513	0.005	+−+???+++?+?
C/G	−3.11	0.0019*	−	rs73188175	T/C	0.300	0.0019*	−−++++++++++
G/C	3.09	0.0020*	+	rs2699882	A/G	0.053	0.0096	+−+−−++−++++
GT/G	−2.86	0.0043	−	rs7427054	T/C	0.093	0.0015*	++−−−++++++−
G/C	−2.87	0.0041	−	rs12509107	A/G	−0.445	0.0168	−−???−??????
T/C	−3.43	0.0006*	−	rs62386289	T/C	−0.117	0.0039	−+−−−−−+−++
A/G	2.53	0.0113	+	rs6579892	A/T	0.075	0.00095*	++++++++++−
G/A	−2.77	0.0057	−	rs117804808	T/C	0.250	0.0099	+++−++++−++
A/G	3.18	0.0015*	+	rs150770834	A/G	0.494	0.0108	−?????++++++
A/T	−2.90	0.0037	−	rs11998452	A/G	−0.218	0.0021	+−+?−?−−−−−−
A/G	2.29	0.0220	+	rs143862820	T/C	0.289	0.0055	?++−?+++−+??
C/T	−2.81	0.0049	−	rs192204676	A/G	0.332	0.016	+−−−??++++++
G/A	−3.21	0.0013*	−	rs188855777	T/C	−0.653	0.0032	??????−?−??−
C/T	2.90	0.0037	+	rs4909989	A/G	−0.080	0.0033	−−−−−−−−−++−
A/G	−3.39	0.0007*	−	rs139079454	T/C	0.233	0.0043	+++?++++−+++
T/G	−3.50	0.0005*	−	rs12311115	A/G	−0.119	0.00031*	+−−−−−−−+−−+
A/G	−3.40	0.0007	−	rs78381318	A/G	0.194	0.000013*	++−+−+++++−+
C/T	−2.62	0.0087	−	rs117548270	A/G	−0.312	0.0017*	−−−+?−−−−−−−
C/G	−2.77	0.0057	−	rs111650311	T/C	0.072	0.0228	++++−++−++?+
C/T	2.98	0.0029	+	rs146227033	C/G	−0.245	0.00068*	−−−−?−−−+−−−
C/A	−3.12	0.0018*	−	rs2160742	A/G	0.074	0.0047	++−++++−+++−
G/A	2.92	0.0035	+	rs2247822	T/C	0.071	0.00055*	+−−−+++++++−

Chr indicates chromosomes; COMPASS, Consortium of Minority Population Genome-Wide Association Studies of Stroke; SiGN, Stroke Genetics Network; and SNP, single nucleotide polymorphism.

*Significance for replication $P<2.08\times10^{-3}$).

Validation of COMPASS SNPs in SiGN and METASTROKE

Expanding to the flanking regions and using a stringent Bonferroni correction of $\alpha=2.08\times10^{-3}$ for replication (0.05/24 unique loci), our most significant locus, *HNF1A*, was validated in both SiGN and METASTROKE European-ancestry cohorts and approached significance in SiGN Hispanics (Figure I in the Data Supplement). Overall, 16 of 24 loci showed evidence for validation across multiple populations (Table 2).

Likely due to the inclusion of ischemic stroke cases only, we were not able to replicate the novel association for rs4471613, which was associated with total (ischemic and hemorrhagic) stroke in our prior COMPASS HapMap imputation report (inverse variance weighting $P=0.85$).[7]

Additionally, we found no evidence of replication for loci previously associated with stroke in European-Ancestry populations (*P* ranging from 0.02 to 0.95; Tables IV and V in the Data Supplement).

DISCUSSION

This new COMPASS meta-analysis of ischemic stroke only identified 24 unique loci with suggestive (n=23) or genome-wide (n=1) evidence for association with ischemic stroke. The most significantly associated *HNF1A* variant, rs55931441 (G/A), is monomorphic in European populations (G allele present only), with a 2% minor allele frequency (allele A) reported in sub-Saharan and 1000G African populations, and 3.8% frequency in

COMPASS. This SNP was present in the only 2 studies imputed to 1000G Phase III (WHI and SLESS). Collectively, WHI and SLESS account for 9637 subjects (1147 stroke cases and 8490 controls). We were unable to assess the association for rs55931441 directly in our cross-ethnic look-up; however, SNPs in a 100 kb flanking region were significant (Figure I in the Data Supplement) in SiGN Europeans (top SNP rs182546302; $P=8.18\times10^{-4}$), METASTROKE ischemic stroke phenotype (top SNP rs117548270; $P=1.72\times10^{-3}$), and METASTROKE cardioembolic stroke phenotype (top SNP rs184865012; $P=9.98\times10^{-4}$), whereas SNP rs80019595 approached significance ($P=8.74\times10^{-3}$) in the SiGN Hispanic cohort. Previous studies have reported associations between variants in *HNF1A* and lipids,[33] C-reactive protein,[34,35] and risk of coronary artery disease and stroke.[33,35] Taken together, these findings may provide greater insight regarding subtype-specific influences and potential mechanism of *HNF1A* variants in stroke risk.

Three additional variants reached suggestive associations at the $P\leq10^{-8}$ level (rs113509723 in *TMEM108* (Figure [B]); rs142655108 near *NPM1P48* (Figure [C]); rs150807690 in *SFXN4*). The *NPM1P48* locus showed no evidence for replication in the cross-ethnic look-up, whereas *TMEM108* was replicated in SiGN Hispanics only (top SiGN Hispanic SNP rs139695007; $P=0.002$). The *SFXN4* SNP, rs150807690, is a G insertion (–/G) with a 22% minor allele frequency (G insertion) in the 1000G African population and 24% frequency in COMPASS. Variant rs150807690 did not replicate in SiGN Hispanic ($P=0.796$) or SiGN Europeans ($P=0.696$) analyses and was not present in the METASTROKE look-up; however, nearby SNPs with evidence of replication in a 100 kb flanking region were detected in SiGN Europeans (top SNP rs143931152; $P=2.68\times10^{-4}$) and SiGN Hispanics (top SNP rs56095167; $P=1.31\times10^{-3}$), located 35 540 bp and 97 388 bp from the indexed COMPASS variant, respectively. The *SFXN4* gene has not been previously implicated in stroke. The protein encoded by *SFXN4* is critical for mitochondrial respiration and erythropoiesis.[36,37] Recent clinical trials suggest that erythropoiesis-stimulating agents effectively treat anemia associated with chronic kidney disease but increase the risk of stroke possibly due to hyperviscosity.[38]

Of the 23 loci with suggestive association in COMPASS, 15 showed evidence for replication in ≥1 look-up analysis. One locus was replicated in SiGN Europeans only, four loci were replicated in SiGN Hispanics only, 2 loci were replicated in METASTROKE ischemic stroke only, whereas 8 loci had evidence for replication in ≥2 look-ups. Two loci, *SFXN4 and UQCRFS1*, were replicated in both the SiGN Europeans and Hispanics, 2 loci were replicated in SiGN Hispanics and METASTROKE ischemic stroke (*KALRN* and *FAR2*), and 3 loci were replicated in SiGN Europeans and METASTROKE ischemic stroke (*CTTNBP2L*, *GTSCR1*, and *RUNX1*). Most notably, one locus (*SRRM4*) was replicated in all 3 look-ups. Evidence for association across multiple ethnicities might indicate stroke susceptibility loci with a global impact. For example, the *KALRN* locus which was replicated in SiGN Hispanics and METASTROKE has been implicated in coronary artery disease risk across multiple populations[39–41] and was recently associated with ischemic stroke and lacunar stroke in a Han Chinese population.[42] Although the *SRRM4* locus, which was replicated in all 3 look-ups, has not previously been implicated in stroke, the gene is important for neurogenesis[43] and has shown associations with neurological conditions including Alzheimer disease[44] and epilepsy.[45]

Although this effort represents the largest stroke GWAS meta-analysis in individuals of African descent, the modest sample size of 3734 stroke cases limits our power to detect associations for variants with minor allele frequencies of ≤3%. Only 2 cohorts used the most recent imputation panel limiting our ability, and thus power, to detect novel variants only present in 1000G Phase III and not 1000G Phase I Version 3. Furthermore, individuals of African descent experience ischemic strokes of small vessel origin more frequently. Therefore, due to the increased genetic diversity of this COMPASS population combined with the greater prevalence of small vessel stroke, we are not surprised at a lack of validation of previous European-ancestry associations. Failure to replicate associations across ethnicities is a common occurrence in genetic studies of various diseases and, therefore, does not threaten the validity of our current study. Moreover, the lack of availability of an adequate replication cohort consisting of individuals of African descent suffering a stroke that have genome-wide SNP genotype data remains a substantial global challenge. Likewise, due to smaller linkage disequilibrium blocks and increased genetic diversity in populations of African descent, larger sample sizes would help alleviate limitations of statistical power, challenges associated with imputing genotypes, and allow for more detailed stroke subtype analyses. A recent analysis showed that although the number of GWAS conducted as of 2016 has increased >6-fold since 2009, African descent participants increased by only 2.5%.[46] Therefore, our study will help advance precision medicine applications by identifying genetic loci (and subsequent polygenic risk scores) for stroke prediction and risk stratification in diverse populations.

SUMMARY

Despite its limitations, genetic studies, such as COMPASS, that include minority populations have the huge potential to provide insight into the mechanisms underlying stroke disparities, such as the more than doubled

incidence and mortality rates and younger age of onset for stroke observed in blacks.[5,47] Our study identified novel associations for stroke that might not otherwise be detected in primarily European cohort studies. Collectively, this highlights the critical nature and importance of genetic studies in a more diverse population with a high stroke burden, such as was the case in this study.

ARTICLE INFORMATION

Received October 15, 2019; final revision received June 10, 2020; accepted June 18, 2020.

Affiliations

Department of Biology; Brody School of Medicine Center for Health Disparities, East Carolina University, Greenville, NC (K.L.K.). Aflac Cancer and Blood Disorder Center of Emory University and Children's Healthcare of Atlanta University, GA (H.I.H.). University of Washington, Seattle, WA (J.C.B., W.T.L.). Baltimore Veterans Administration Medical Center and University of Maryland School of Medicine, Baltimore, MD (S.J.K., B.D.M., Y.-C.C.). McMaster University and Population Health Research Institute, Hamilton Ontario (G.P., M.C.). National University of Ireland Galway, Ireland (M.O.). Mayo Clinic Florida, Jacksonville (J.F.M.). Center for Public Health Genomics, University of Virginia, Charlottesville (W.-M.C., M.M.S., S.S.R.). Laboratory of Neurogenetics, National Institute on Aging, Bethesda, MD (M.A.N.). Data Tecnica International, Glen Echo, MD (M.A.N.). Laboratory of Epidemiology and Population Science, National Institute on Aging, Baltimore, MD (A.B.Z., M.K.E.). University of Mississippi Medical Center, Jackson (J.G.W., A.C., T.H.M.). University of Cambridge, England, United Kingdom (H.S.M.). William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, United Kingdom (M.T.). Social, Genetic and Developmental Psychiatry Centre, King's College London, United Kingdom (C.M.L.). Initiative for Research and Education to Advance Community Health, Washington State University, Seattle (C.L.C.). Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA (A.R., J.H., C.K.). Division of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, NC (C.D.L., C.L.). Johns Hopkins University School of Medicine, Baltimore, MD (R.G.). Department of Neurology, University of Cincinnati, OH (D.W.). University of California, San Francisco (K.Y.). Cardiovascular Health Research Unit, Departments of Medicine, Epidemiology and Health Services, University of Washington, Seattle (B.M.P.). Kaiser Permanente Washington Health Research Institute, Seattle, WA (B.M.P.). University of Colorado Anschutz Medical Campus, Denver (L.A.L.). University of Miami, Miller School of Medicine, FL (R.S., T.R.). Washington University School of Medicine, St. Louis, MO (J.-M.L., C.C.). Brown University Warren Alpert Medical School, Providence, RI (K.L.F.). University of Kentucky, College of Public Health, Lexington (D.K.A.). University of British Columbia, Vancouver, (O.R.B.). Neuroscience Institute, Saint Francis Medical Center, Trenton, NJ (R.P.G., L.R.P.). Institute for Stroke and Dementia Research (ISD), University Hospital, LMU Munich, Germany (M.D., R.M.). Munich Cluster for Systems Neurology (SyNergy), Germany (M.D.). Department of Neurology, University of Virginia, Charlottesville (B.B.W.). Brown Foundation Institute of Molecular Medicine, University of Texas Health Science Center at Houston (M.F.).

Acknowledgments

We thank the staff and participants of the ARIC (Atherosclerosis Risk in Communities), CHS (Cardiovascular Health Study), VISP (Vitamin Intervention for Stroke Prevention), HANDLS (Healthy Aging in Neighborhoods of Diversity Across the Life Span), INTERSTROKE, ISGS (Ischemic Stroke Genetics Study), JHS (Jackson Heart Study), SIGNET-REGARDS (Sea Islands Genetics Network–Reasons for Geographic and Racial Differences in Stroke), GEOS (Genetics of Early Onset Stroke), SiGN (Stroke Genetics Network), SLESS (South London Ethnicity and Stroke Study), SWISS (Siblings With Ischemic Stroke Study), and WHI (Women's Health Initiative) studies for their dedication, and willingness to participate in the respective research studies, which made this work possible. Additional Information: Coauthor Michèle M. Sale died April 29, 2018.

Sources of Funding

ARIC Study (Atherosclerosis Risk in Communities): supported in whole or in part with Federal funds from the NHLBI (National Heart, Lung, and Blood Institute), National Institutes of Health, Department of Health and Human Services (contract numbers HHSN268201700001I, HHSN268201700002I, HHSN268201700003I, HHSN268201700004I, and HHSN268201700005I), R01HL087641, R01HL059367 and R01HL086694; National Human Genome Research Institute contract U01HG004402; and National Institutes of Health contract HHSN268200625226C. We thank the staff and participants of the ARIC study for their important contributions. Infrastructure was partly supported by Grant Number UL1RR025005, a component of the National Institutes of Health and National Institutes of Health (NIH) Roadmap for Medical Research. CHS (Cardiovascular Health Study): supported by NHLBI 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 was 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. VISP (Vitamin Intervention for Stroke Prevention): funded by the National Institute of Neurological Disorders and Stroke (R01-NS34447). Genome-wide association study (GWAS) data for a subset of VISP participants were supported by the National Human Genome Research Institute (U01-HG005160), as part of the GARNET (Genomics and Randomized Trials Network; Principal Investigator [PI]: Drs Sale and Worrall). INTERSTROKE has received unrestricted grants from the Canadian Institutes of Health Research, Heart and Stroke Foundation of Canada, Canadian Stroke Network, Pfizer Cardiovascular Award, Merck, AstraZeneca, and Boehringer Ingelheim. JHS (Jackson Heart Study): supported and conducted in collaboration with Jackson State University (HHSN268201800013I), Tougaloo College (HHSN268201800014I), the Mississippi State Department of Health (HHSN268201800015I) and the University of Mississippi Medical Center (HHSN268201800010I, HHSN268201800011I, and HHSN268201800012I) contracts from the NHLBI and the National Institute for Minority Health and Health Disparities (NIMHD). SIGNET-REGARDS (Sea Islands Genetics Network–Reasons for Geographic and Racial Differences in Stroke: SIGNET was supported by R01 DK084350 (Dr Sale) and consists of data from the REGARDS cohort, (U01 NS041588; G Howard). GEOS Study (The Genetics of Early Onset Stroke): GEO Study was supported by the NIH Genes, Environment, and Health Initiative grant U01 HG004436, as part of the Gene Environment Association Studies consortium, with additional support provided by the Mid-Atlantic Nutrition and Obesity Research Center (P30 DK072488) and the Office of Research and Development, Medical Research Service, and the Baltimore Geriatrics Research, Education, and Clinical Center of the Department of Veterans Affairs. Genotyping services were provided by the Johns Hopkins University Center for Inherited Disease Research (CIDR), which is fully funded through a federal contract from the NIH to the Johns Hopkins University (contract No. HHSN268200782096C). Assistance with data cleaning was provided by the Gene Environment Association Studies Consortium Coordinating Center (U01 HG 004446; PI Bruce S. Weir). Study recruitment and assembly of data sets were supported by a Cooperative Agreement with the Division of Adult and Community Health, Centers for Disease Control and by grants from the National Institute of Neurological Disorders and Stroke (NINDS) and the NIH Office of Research on Women's Health (R01 NS45012, U01 NS069208-01).

NINDS-SiGN Groups 4: The SiGN study was funded by a cooperative agreement grant from the NINDS U01 NS069208. Genotyping services were provided by the Johns Hopkins University CIDR, which is fully funded through a federal contract from the NIH to the Johns Hopkins University (contract No.HHSN268200782096C). The Biostatistics Department Genetics Coordinating Center at the University of Washington (Seattle) provided more extensive quality control of the genotype data through a subcontract with CIDR. Additional support to the Administrative Core of SiGN was provided by the Dean's Office, University of Maryland School of Medicine.

SLESS (South London Ethnicity and Stroke Study). Recruitment to SLESS was supported by a program grant from the Stroke Association. This study represents independent research part-funded by the National Institute for Health Research Biomedical Research Centre at South London and Maudsley National Health Service (NHS) Foundation Trust and King's College London, and the National Institute for Health Research Biomedical Research Centre at Guy's and St Thomas' NHS Foundation Trust and King's College London. Hugh Markus is supported by a National Institute for Health Research (NIHR) Senior Investigator Award, and his work is supported by NIHR Comprehensive Biomedical Research Unit funding awarded to Cambridge University Hospitals Trust. ISGS (Ischemic Stroke Genetics Study) and SWISS (Siblings with Ischemic Stroke Study): ISGS and SWISS were supported by the National Institute of Neurological Disorders and Stroke grants (R01 NS42733; Dr Meschia) and (R01NS39987; Dr Me-

schia), respectively. Both studies received additional support, in part, from the Intramural Research Program of the National Institute of Aging (Z01 AG000954-06; PI Singleton), and used samples and clinical data from the NIH-NINDS Human Genetics Resource Center DNA and Cell Line Repository (http://ccr.cori- ell.org/ninds), human subjects protocol numbers 2003–081 and 2004–147. WHI (Women's Health Initiative): supported by the NHLBI, NIH, and the US 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. The funders had no role in study design, data collection and analysis, decision to publish, or manuscript preparation. Dr Hyacinth is supported by NIH/NHLBI grants U01HL117721, R01HL138423, and R56HL136210. HANDLS (Healthy Aging in Neighborhoods of Diversity Across the Life Span): supported by the Intramural Research Program of the NIH, National Institute of Aging (project no. Z01-AG000513 and human subjects protocol no. 09AGN248).

Disclosures

Dr Nall's participation is supported by a consulting contract between Data Tecnica International and the National Institute on Aging, National Institutes of Health (NIH). Dr Nall consults for the Michael J. Fox Foundation, Vivid Genomics, Lysosomal Therapies, Inc, Illumina, Inc, and SK Therapeutics. Dr Furie is a Deputy Editor for *Stroke* journal. Dr Psaty serves on the Steering Committee of the Yale Open Data Access Project funded by Johnson & Johnson. Dr Cruchaga receives research support from Biogen, EISAI, Alector, and Parabon. Dr Cruchaga is a member of the advisory board of Vivid genetics, Halia Therapeutics, and ADx Healthcare. Dr Worrall serves as Deputy Editor for the journal *Neurology*. H.S. Markus has received personal fees from BIBA. Dr Gottesman is a former Associate Editor for the journal *Neurology*. Dr Lee consults for Regenera. The other authors report no conflicts.

REFERENCES

1. Benjamin EJ, Muntner P, Alonso A, Bittencourt MS, Callaway CW, Carson AP, Chamberlain AM, Chang AR, Cheng S, Das SR, et al; American Heart Association Council on Epidemiology and Prevention Statistics Committee and Stroke Statistics Subcommittee. Heart disease and stroke statistics-2019 update: a report from the American heart association. *Circulation*. 2019;139:e56–e528. doi: 10.1161/CIR.0000000000000659
2. Hajat C, Dundas R, Stewart JA, Lawrence E, Rudd AG, Howard R, Wolfe CD. Cerebrovascular risk factors and stroke subtypes: differences between ethnic groups. *Stroke*. 2001;32:37–42. doi: 10.1161/01.str.32.1.37
3. Traylor M, Bevan S, Baron JC, Hassan A, Lewis CM, Markus HS. Genetic architecture of lacunar stroke. *Stroke*. 2015;46:2407–2412. doi: 10.1161/STROKEAHA.115.009485
4. Bevan S, Traylor M, Adib-Samii P, Malik R, Paul NL, Jackson C, Farrall M, Rothwell PM, Sudlow C, Dichgans M, et al. Genetic heritability of ischemic stroke and the contribution of previously reported candidate gene and genomewide associations. *Stroke*. 2012;43:3161–3167. doi: 10.1161/STROKEAHA.112.665760
5. Howard G, Moy CS, Howard VJ, McClure LA, Kleindorfer DO, Kissela BM, Judd SE, Unverzagt FW, Soliman EZ, Safford MM, et al; REGARDS Investigators*. Where to focus efforts to reduce the black-white disparity in stroke mortality: incidence versus case fatality? *Stroke*. 2016;47:1893–1898. doi: 10.1161/STROKEAHA.115.012631
6. White H, Boden-Albala B, Wang C, Elkind MS, Rundek T, Wright CB, Sacco RL. 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
7. Carty CL, Keene KL, Cheng YC, Meschia JF, Chen WM, Nalls M, Bis JC, Kittner SJ, Rich SS, Tajuddin S, et al; COMPASS and METASTROKE Consortia. Meta-analysis of genome-wide association studies identifies genetic risk factors for stroke in African Americans. *Stroke*. 2015;46:2063–2068. doi: 10.1161/STROKEAHA.115.009044
8. Gretarsdottir S, Thorleifsson G, Manolescu A, Styrkarsdottir U, Helgadottir A, Gschwendtner A, Kostulas K, Kuhlenbäumer G, Bevan S, Jonsdottir T, 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
9. Dichgans M, Malik R, König IR, Rosand J, Clarke R, Gretarsdottir S, Thorleifsson G, Mitchell BD, Assimes TL, Levi C, 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. Bellenguez C, Bevan S, Gschwendtner A, Spencer CC, Burgess AI, Pirinen M, Jackson CA, Traylor M, Strange A, Su Z, et al. Genome-wide association study identifies a variant in hdac9 associated with large vessel ischemic stroke. *Nature genetics*. 2012;44:328–333. doi: 10.1038/ng.1081
11. Holliday EG, Maguire JM, Evans TJ, Koblar SA, Jannes J, Sturm JW, Hankey GJ, Baker R, Golledge J, Parsons MW, 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
12. Ikram MA, Seshadri S, Bis JC, Fornage M, DeStefano AL, Aulchenko YS, Debette S, Lumley T, Folsom AR, van den Herik EG, et al. Genomewide association studies of stroke. *N Engl J Med*. 2009;360:1718–1728. doi: 10.1056/NEJMoa0900094
13. Kilarski LL, Achterberg S, Devan WJ, Traylor M, Malik R, Lindgren A, Pare G, Sharma P, Slowik A, Thijs V, 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
14. Traylor M, Farrall M, Holliday EG, Sudlow C, Hopewell JC, Cheng YC, Fornage M, Ikram MA, Malik R, Bevan S, 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
15. Malik R, Traylor M, Pulit SL, Bevan S, Hopewell JC, Holliday EG, Zhao W, Abrantes P, Amouyel P, Attia JR, et al; ISGC Analysis Group; METASTROKE collaboration; Wellcome Trust Case Control Consortium 2 (WTCCC2); NINDS Stroke Genetics Network (SiGN). Low-frequency and common genetic variation in ischemic stroke: the METASTROKE collaboration. *Neurology*. 2016;86:1217–1226. doi: 10.1212/WNL.0000000000002528
16. Malik R, Chauhan G, Traylor M, Sargurupremraj M, Okada Y, Mishra A, Rutten-Jacobs L, Giese AK, van der Laan SW, Gretarsdottir S, et al; AFGen Consortium; Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium; International Genomics of Blood Pressure (iGEN-BP) Consortium; INVENT Consortium; STARNET; BioBank Japan Cooperative Hospital Group; COMPASS Consortium; EPIC-CVD Consortium; EPIC-InterAct Consortium; International Stroke Genetics Consortium (ISGC); METASTROKE Consortium; Neurology Working Group of the CHARGE Consortium; NINDS Stroke Genetics Network (SiGN); UK Young Lacunar DNA Study; MEGASTROKE Consortium. Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nat Genet*. 2018;50:524–537. doi: 10.1038/s41588-018-0058-3
17. The ARIC Investigators. The atherosclerosis risk in communities (aric) study: design and objectives. *Am J Epidemiol*. 1989;129:687–702.
18. Fried LP, Borhani NO, Enright P, Furberg CD, Gardin JM, Kronmal RA, Kuller LH, Manolio TA, Mittelmark MB, Newman A. The cardiovascular health study: design and rationale. *Ann Epidemiol*. 1991;1:263–276. doi: 10.1016/1047-2797(91)90005-w
19. Sempos CT, Bild DE, Manolio TA. Overview of the Jackson heart study: a study of cardiovascular diseases in African American men and women. *Am J Med Sci*. 1999;317:142–146. doi: 10.1097/00000441-199903000-00002
20. Wilson JG, Rotimi CN, Ekunwe L, Royal CD, Crump ME, Wyatt SB, Steffes MW, Adeyemo A, Zhou J, Taylor HA Jr, et al. Study design for genetic analysis in the Jackson heart study. *Ethn Dis*. 2005;15(4 Suppl 6):S6–30.
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. doi: 10.1016/s0197-2456(97)00078-0
22. O'Donnell M, Xavier D, Diener C, Sacco R, Lisheng L, Zhang H, Pias P, Truelsen T, Chin SL, Rangarajan S, et al; INTERSTROKE investigators. Rationale and design of INTERSTROKE: a global case-control study of risk factors for stroke. *Neuroepidemiology*. 2010;35:36–44. doi: 10.1159/000306058
23. Howard VJ, Cushman M, Pulley L, Gomez CR, Go RC, Prineas RJ, Graham A, Moy CS, Howard G. The reasons for geographic and racial differences in stroke study: objectives and design. *Neuroepidemiology*. 2005;25:135–143. doi: 10.1159/000086678
24. Meschia JF, Brott TG, Brown RD Jr, Crook RJ, Frankel M, Hardy J, Merino JG, Rich SS, Silliman S, Worrall BB; Ischemic Stroke Genetics Study. The

ischemic stroke genetics study (ISGS) protocol. *BMC Neurol.* 2003;3:4. doi: 10.1186/1471-2377-3-4
25. Spence JD, Howard VJ, Chambless LE, Malinow MR, Pettigrew LC, Stampfer M, Toole JF. Vitamin intervention for stroke prevention (VISP) trial: rationale and design. *Neuroepidemiology.* 2001;20:16–25. doi: 10.1159/000054753
26. Williams SR, Hsu FC, Keene KL, Chen WM, Nelson S, Southerland AM, Madden EB, Coull B, Gogarten SM, Furie KL, et al; METASTROKE, The Genomics and Randomized Trials Network (GARNET) Collaborative Research Group; METASTROKE The Genomics and Randomized Trials Network GARNET Collaborative Research Group. Shared genetic susceptibility of vascular-related biomarkers with ischemic and recurrent stroke. *Neurology.* 2016;86:351–359. doi: 10.1212/WNL.0000000000002319
27. Traylor M, Rutten-Jacobs L, Curtis C, Patel H, Breen G, Newhouse S, Lewis CM, Markus HS. Genetics of stroke in a UK African ancestry case-control study: South London ethnicity and stroke study. *Neurol Genet.* 2017;3:e142. doi: 10.1212/NXG.0000000000000142
28. Cheng YC, O'Connell JR, Cole JW, Stine OC, Dueker N, McArdle PF, Sparks MJ, Shen J, Laurie CC, Nelson S, et al. Genome-wide association analysis of ischemic stroke in young adults. *G3 (Bethesda).* 2011;1:505–514. doi: 10.1534/g3.111.001164
29. Meschia JF, Arnett DK, Ay H, Brown RD Jr, Benavente OR, Cole JW, de Bakker PI, Dichgans M, Doheny KF, Fornage M, et al; NINDS SiGN Study. Stroke Genetics Network (SiGN) study: design and rationale for a genome-wide association study of ischemic stroke subtypes. *Stroke.* 2013;44:2694–2702. doi: 10.1161/STROKEAHA.113.001857
30. 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.
31. 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. doi: 10.1186/1471-2350-3-1
32. 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
33. Zhou YJ, Yin RX, Hong SC, Yang Q, Cao XL, Chen WX. Association of the hnf1a polymorphisms and serum lipid traits, the risk of coronary artery disease and ischemic stroke. *J Gene Med.* 2017;19:e2941. doi: 10.1002/jgm.2941.
34. López-Mejías R, Genre F, Remuzgo-Martínez S, González-Juanatey C, Robustillo-Villarino M, Llorca J, Corrales A, Vicente E, Miranda-Filloy JA, Magro C, et al. Influence of elevated-CRP level-related polymorphisms in non-rheumatic Caucasians on the risk of subclinical atherosclerosis and cardiovascular disease in rheumatoid arthritis. *Sci Rep.* 2016;6:31979. doi: 10.1038/srep31979
35. Shi H, Leng S, Liang H, Zheng Y, Chen L. Association study of C-reactive protein associated gene HNF1A with ischemic stroke in Chinese population. *BMC Med Genet.* 2016;17:51. doi: 10.1186/s12881-016-0313-3
36. Hildick-Smith GJ, Cooney JD, Garone C, Kremer LS, Haack TB, Thon JN, Miyata N, Lieber DS, Calvo SE, Akman HO, et al. Macrocytic anemia and mitochondriopathy resulting from a defect in sideroflexin 4. *Am J Hum Genet.* 2013;93:906–914. doi: 10.1016/j.ajhg.2013.09.011
37. Zheng H, Ji C, Zou X, Wu M, Jin Z, Yin G, Li J, Feng C, Cheng H, Gu S, et al. Molecular cloning and characterization of a novel human putative transmembrane protein homologous to mouse sideroflexin associated with sideroblastic anemia. *DNA Seq.* 2003;14:369–373. doi: 10.1080/10425170310001605491
38. Seliger SL, Zhang AD, Weir MR, Walker L, Hsu VD, Parsa A, Diamantidis CJ, Fink JC. Erythropoiesis-stimulating agents increase the risk of acute stroke in patients with chronic kidney disease. *Kidney Int.* 2011;80:288–294. doi: 10.1038/ki.2011.49
39. Horne BD, Hauser ER, Wang L, Muhlestein JB, Anderson JL, Carlquist JF, Shah SH, Kraus WE. Validation study of genetic associations with coronary artery disease on chromosome 3q13-21 and potential effect modification by smoking. *Ann Hum Genet.* 2009;73(pt 6):551–558. doi: 10.1111/j.1469-1809.2009.00540.x
40. Wang L, Hauser ER, Shah SH, Pericak-Vance MA, Haynes C, Crosslin D, Harris M, Nelson S, Hale AB, Granger CB, et al. Peakwide mapping on chromosome 3q13 identifies the kalirin gene as a novel candidate gene for coronary artery disease. *Am J Hum Genet.* 2007;80:650–663. doi: 10.1086/512981
41. Boroumand M, Ziaee S, Zarghami N, Anvari MS, Cheraghi S, Abbasi SH, Jalali A, Pourgholi L. The kalirin gene rs9289231 polymorphism as a novel predisposing marker for coronary artery disease. *Lab Med.* 2014;45:302–308. doi: 10.1309/LMLS813ZDPHRFLUU
42. Dang M, Wang Z, Zhang R, Li X, Peng Y, Han X, Sun L, Tian J. KALRN rare and common variants and susceptibility to ischemic stroke in Chinese han population. *Neuromolecular Med.* 2015;17:241–250. doi: 10.1007/s12017-015-8352-z
43. Raj B, Irimia M, Braunschweig U, Sterne-Weiler T, O'Hanlon D, Lin ZY, Chen GI, Easton LE, Ule J, Gingras AC, et al. A global regulatory mechanism for activating an exon network required for neurogenesis. *Mol Cell.* 2014;56:90–103. doi: 10.1016/j.molcel.2014.08.011
44. Chung J, Wang X, Maruyama T, Ma Y, Zhang X, Mez J, Sherva R, Takeyama H, Lunetta KL, Farrer LA, et al; Alzheimer's Disease Neuroimaging Initiative. Genome-wide association study of Alzheimer's disease endophenotypes at prediagnosis stages. *Alzheimers Dement.* 2018;14:623–633. doi: 10.1016/j.jalz.2017.11.006
45. Rusconi F, Paganini L, Braida D, Ponzoni L, Toffolo E, Maroli A, Landsberger N, Bedogni F, Turco E, Pattini L, et al. LSD1 Neurospecific alternative splicing controls neuronal excitability in mouse models of epilepsy. *Cereb Cortex.* 2015;25:2729–2740. doi: 10.1093/cercor/bhu070
46. Popejoy AB, Fullerton SM. Genomics is failing on diversity. *Nature.* 2016;538:161–164. doi: 10.1038/538161a
47. Howard G, Kissela BM, Kleindorfer DO, McClure LA, Soliman EZ, Judd SE, Rhodes JD, Cushman M, Moy CS, Sands KA, et al. Differences in the role of black race and stroke risk factors for first vs. recurrent stroke. *Neurology.* 2016;86:637–642. doi: 10.1212/WNL.0000000000002376

SUPPLEMENTAL MATERIAL

A GWAS Meta-Analysis of Stroke in 22,000 individuals of African descent identifies novel associations with stroke

Supplemental Methods

Meta-Analysis Methods.
All studies imputed SNPs using 1000 Genomes Phase I Version 3 Haplotypes (1KGp1v3), except SLESS and WHI, which used 1000G Phase III data (1KGp3) reference populations (http://www.internationalgenome.org/1000-genomes-browsers/). SNPs were excluded from subsequent analyses if they had invalid or missing alleles, P-Values, or Beta values; minor allele frequencies (MAF) < 1%; imputation quality (r^2) <0.3; and were located on sex chromosomes. We analyzed SNPs available in two or more studies, for a total of ~16.9 million SNPs. We used logistic regression (additive genetic model) analyses with a count of variant alleles (0, 1, or 2) for each genotyped SNP or allelic dose for imputed SNPs. 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 due to population stratification; no large deviations were noted. We combined study-specific results in fixed effects meta-analyses with inverse variance weighting and sample size weighting using METAL,[1], P-value meta-analyses were also conducted. We set the genome-wide significance threshold at $P<5x10^{-8}$ for the GWAS discovery but investigated all SNPs with $P<10^{-6}$. The genomic inflation factors were estimated and determined to be 1.02 for both the sample size weighting and inverse variance weighting meta-analyses for ischemic stroke, indicating minimal evidence of inflation of the test statistics due to stratification.

Power. Using GAS Power Calculator (www.csg.sph.umich.edu/abecasis/cats/gas_power_calculator) we estimated ≥99% power to detect associations at the $5x10^{-8}$ p-value threshold for low frequency common variants ($0.05 \geq MAF \leq 0.07$) in our population of 3,734 AA stroke cases and 18,317 controls for a genotype relative risk of 1.5. Using the same parameters, we have considerably less power to detect associations for variants with 1% (power of 6%), 2% (power of 50.9%), and 3% (power of 87.1%) allele frequencies, while achieving 98.6% power at a 4% MAF.

Study Descriptions ARIC, CHS, VISP, INTERSTROKE, JHS, SIGNET-REGARDS, GEOS, SiGN (group 4), SLESS, SWISS, ISGS, WHI and HANDLS.

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 to1000G Phase III data (1KGp3) reference populations. MACH v1.0.16[3, 4] 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.[5] Briefly, the stroke diagnosis was assigned according to criteria adapted from the National Survey of Stroke.[6] 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.[7] 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.[8, 9] 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.

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.[10, 11] 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).

INTERSTROKE

INTERSTROKE is an international, multi-centered, case-control study of stroke investigating the global burden of risk factors across 32 countries and 18 different ethnic groups around the world. A detailed report of the study design has been published.[12] Briefly, cases were patients with acute first stroke (within 5 days of symptoms onset and 72 hours of hospital admission) in

whom neuroimaging (CT or MRI) was performed. The TOAST classification system was used to define ischemic stroke subtypes. Cases were excluded if 1) they were unable to communicate due to severe stroke without a valid surrogate respondent (e.g. first-degree relative or spouse), 2) they were hospitalized for acute coronary syndrome/myocardial infarction, or 3) stroke was attributed to non-vascular causes (e.g. tumor). Controls were selected from the community and had no history of stroke. The study was approved by the ethics committees in all participating centers. All participants, or their proxy, provided written informed consent before taking part in the study.

Genotyping. A subset of INTERSTROKE participants consenting to genetic analysis with sufficient DNA quantities were genotyped on the Illumina Infinium Cardiometabo BeadChip. All samples were genotyped at a central site (the Genetic Molecular Epidemiology Laboratory in Hamilton, Ontario, Canada). Samples were excluded if they had 1) a high proportion of missing variants (missingness > 0.05), 2) inconsistencies between reported and genetically determined sex or ethnicity or 3) exhibited cryptic relatedness. Genotyped variants were excluded if they were rare (MAF < 0.01), exhibited high missingness across samples (missingness > 0.01), or deviated from hardy-weinberg equilibrium ($P<5x10^{-6}$). Pre-phasing and imputation were performed with SHAPEIT2[13] and IMPUTE[14] respectively, using the 1000Genomes Phase 1 Version 3 (November 23, 2010 subversion) reference panel. Imputed variants were removed if they were rare (MAF < 0.01) or of poor quality (INFO SCORE < 0.30).

Jackson Heart Study (JHS)

The JHS is a single-site, prospective, population-based study designed to explore the environmental, behavioral, and genetic factors that influence the development of CVD among African Americans. A total of 5,301 women and men between the ages of 21 and 94 were recruited between 2000 and 2004 from a tri-county area of Mississippi: Hinds, Madison, and Rankin Counties. Participants were recruited from four sources, including (1) randomly sampled households from a commercial listing; (2) ARIC participants; (3) a structured volunteer sample that was designed to mirror the eligible population; and (4) a nested family cohort. Overviews of the JHS including the sampling and recruitment, sociocultural, and laboratory methods have been described and published previously.[15-18] The institutional review boards of the following participating institutions approved the study: the University of Mississippi Medical Center, Jackson State University, and Tougaloo College. All of the participants provided written informed consent. Unrelated participants were between 35 and 84 years old, and members of the family cohort were ≥ 21 years old when consent for genetic testing was obtained and blood was drawn for DNA extraction.

The baseline examination consisted of a home interview, self-administered questionnaires, and a clinic visit. Medications taken in the prior 2 weeks were brought to clinic and transcribed verbatim with subsequent coding by a pharmacist. After an overnight fast, anthropometric and seated blood pressure measurements were obtained and venipuncture/urine collection was performed in accordance with the National Committee for Clinical Laboratory Standards. Blood pressure was measured by trained technicians using a Hawksley random zero manometer and determined by the arithmetic average of two readings taken 1 minute apart after a five-minute rest.[19]

Stroke Assessment in the JHS: In addition to the standard JHS examinations, participants were contacted by telephone annually beginning in 2005 to obtain interim information about cardiovascular events. (ICD-9 code 428 for hospitalizations). During the annual follow up phone call, participants or designated representative provide self-reported information of hospitalization or death. Identification and abstraction of CVD illness and death data are performed by a certified medical record abstractor. Incident stroke is defined as stroke that occurred while the participants was enrolled the study, i.e. stroke event occurred after the baseline visit. Strokes are classified as either definite or probable stroke. The definition of stroke was based on the World Health Organization (WHO) criteria for definition of stroke or clinical criteria in which case the WHO criteria might not have been satisfied, but there is clinical evidence sufficient for a diagnosis of stroke to be made. More details on identification and classification of stroke events in the JHS have already been published.[20, 21] Although not directly relevant in this study, ischemic stroke subtyping in the JHS was done the TOAST classification criteria.

Genotyping, imputation and quality control. Affymetrix 6.0 GWAS genotyping was performed in n=3,029 JHS participants with consent for genetic analyses through NHLBI's Candidate Gene Association Resource (CARe) consortium. [PMID 20400780] Genotyping and quality control have been previously described. [PMID 21347282]. Imputation to 1000 Genomes Project (1000G) Phase 3 version 5 reference panel was completed using Minimac3 on the Michigan Imputation Server [PMID 27571263]. Prior to imputation, SNPs were filtered for minor allele frequency ≥1%, call rate ≥ 90%, HWE p-value > 10-6, as well as exclusion of sites with invalid or mismatched alleles for the reference panel. Samples with sex or pedigree mismatches and principal components outliers were excluded. Analysis was conducted using the EMMAX test as implemented in EPACTS 3.2.6.

The Sea Islands Genetics Network (SIGNET) & REasons for Geographic And Racial Differences in Stroke (REGARDS)

The Sea Islands Genetics Network (SIGNET) study consists of the REasons for Geographic And Racial Differences in Stroke (REGARDS), the Sea Islands Genetic African American Registry (Project SuGAR), a COBRE for Oral Health study (COBRE), and the Systemic Lupus Erythematosus in Gullah Health study (SLEIGH). All subjects are African Americans (AA), and all provided written informed consent.

All SIGNET samples (n= 4,298) were genotyped using the Affymetrix Genome-Wide Human SNP Array 6.0. Imputation was performed using MACH (version 1.0.16) to impute all autosomal SNPs using 1000 Genomes Phase I Version 3 Haplotypes (1KGp1v3). This COMPASS effort includes only data from the REGARDS study.

REGARDS is an observational cohort of 30,239 AA and white men and women enrolled in their homes after a telephone interview in 2003-7.[22] Participants were a national sample oversampled from the southeastern stroke belt (56%) and were 58% female and 42% black by design. Participants were followed every 6 months by telephone to ascertain health outcomes, with validation of stroke, coronary heart disease, death and other ancillary study endpoints. For SIGNET, we selected all AA REGARDS type 2 diabetes (T2D) cases recruited from SC, GA, NC, and AL, and an equivalent number of race, sex, and age-strata matched diabetes-free

controls. We also included all participants not already included that were current residents of the 15-county "Low Country" region of SC and GA (SC counties Beaufort, Berkeley, Charleston, Colleton, Dorchester, Georgetown, Hampton, Horry, Jasper; GA counties Bryan, Camden, Chatham, Glynn, Liberty, McIntosh). The subset of REGARDS participants genotyped under SIGNET are referred to as SIGNET-REGARDS. GWAS genotyping was completed among 2398 SIGNET-REGARDS AA participants, including 1149 with diabetes and 1249 without diabetes.

The Genetics of Early Onset Stroke (GEOS) Study

The Genetics of Early Onset Stroke (GEOS) Study is a population-based, case-control study designed to identify genes associated with early-onset ischemic stroke (IS) and to characterize interactions of identified stroke genes and/or SNPs with environmental risk factors. Participants were recruited from the greater Baltimore-Washington area in four different periods: Stroke Prevention in Young Women-1 (SPYW-1) conducted from 1992 to 1996, Stroke Prevention in Young Women-2 (SPYW-2) conducted from 2001 to 2003, Stroke Prevention in Young Men (SPYM) conducted from 2003 to 2007, and Stroke Prevention in Young Adults (SPYA) conducted in 2008. From these samples, we identified a total of 921 cases and 941 controls that consented to having their DNA used for genetic studies of stroke. This study was conducted with the consent of all study participants and was approved by the University of Maryland at Baltimore Institutional Review Board. "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. IS with the following characteristics were excluded from participation: stroke occurring as an immediate consequence of trauma; stroke within 48 hr after a hospital procedure, stroke within 60 days after the onset of a nontraumatic subarachnoid hemorrhage, and cerebral venous thrombosis. All cases had neuroimaging that was consistent with cerebral infarction, although neuroimaging was not used for case ascertainment. The abstracted hospital records of cases were reviewed and adjudicated for IS subtype by a pair of neurologists according to previously published,[23, 24] with disagreements resolved by a third neurologist. The IS subtype classification system retains information on all probable and possible causes, and it is reducible to the more widely used TOAST system[25] that assigns each case to a single category. All cases had age of first stroke between 15 and 49 years and were recruited within three years of stroke. "Control participants" without a history of stroke were identified by random-digit dialing. Controls were balanced to cases by age and region of residence in each study and were additionally balanced for race in SPYW-2 and SPYM. Traditional stroke risk factors and other study variables, including age, race/ethnicity, history of hypertension, diabetes, myocardial infarction (MI), and current smoking status (defined as use within one month prior to event for cases and at a comparable reference time for controls), were also collected during a standardized interview. Samples were genotyped at the Johns Hopkins Center for Inherited Disease Research (CIDR), and genotyping was performed using the Illumina HumanOmni1-Quad_v1-0_B BeadChip (Illumina, San Diego, CA).

National Institute of Neurological Disorders and Stroke (NINDS) Stroke Genetics Network (NINDS-SiGN) Groups 4

The Stroke Genetics Network (SiGN) study was funded by a cooperative agreement grant from the National Institute of Neurological Disorders and Stroke (NINDS) U01 NS069208. Genotyping services were provided by the Johns Hopkins University Center for Inherited Disease Research (CIDR), which is fully funded through a federal contract from the National Institutes of Health (NIH) to the Johns Hopkins University (contract No.HHSN268200782096C). The Biostatistics Department Genetics Coordinating Center at the University of Washington (Seattle) provided more extensive quality control of the genotype data through a subcontract with CIDR. Additional support to the Administrative Core of SiGN was provided by the Dean's Office, University of Maryland School of Medicine. SiGN- Group 4 consists of AA subjects from the GASROS, GCNKSS, ISGS, MCISS, MIAMISR, NOMAS, REGARDS, SPS3, SWISS, WHI, and WUSTL studies.

MGH-GASROS: The Massachusetts General Hospital Stroke Genetics Group was supported by the NIH Genes Affecting Stroke Risks and Outcomes Study (GASROS) grant K23 NS042720, the American Heart Association/Bugher Foundation Centers for Stroke Prevention Research 0775010N, and NINDS K23NS042695, K23 NS064052, the Deane Institute for Integrative Research in Atrial Fibrillation and Stroke, and by the Keane Stroke Genetics Fund. Genotyping services were provided by the Broad Institute Center for Genotyping and Analysis, supported by grant U54 RR020278 from the National Center for Research Resources.

GCNKSS: The Greater Cincinnati/Northern Kentucky Stroke Study (GCNKSS) was supported by the NIH (NS030678).

MCISS: The Middlesex County Ischemic Stroke Study (MCISS) was supported by intramural funding from the New Jersey Neuroscience Institute/JFK Medical Center, Edison, NJ, and The Neurogenetics Foundation, Cranbury, NJ. We acknowledge Dr Souvik Sen for his advice and encouragement in the initiation and design of this study.

MIAMISR and NOMAS: The Northern Manhattan Study (NOMAS) was supported by grants from the NINDS (R37 NS029993, R01 NS27517). The Cerebrovascular Biorepository at University of Miami/Jackson Memorial Hospital (The Miami Stroke Registry, Institutional Review Board No. 20070386) was supported by the Department of Neurology at University of Miami Miller School of Medicine and Evelyn McKnight Brain Institute. Biorepository and DNA extraction services were provided by the Hussmann Institute for Human Genomics at the Miller School of Medicine.

SPS3: The Secondary Prevention of Small Subcortical Strokes trial was funded by the US National Institute of Health and Neurological Disorders and Stroke grant No. U01NS38529-04A1 (principal investigator, Oscar R. Benavente; coprincipal investigator, Robert G. Hart). The SPS3 Genetic Substudy (SPS3-GENES) was funded by R01 NS073346 (coprincipal investigators, Julie A. Johnson, Oscar R. Benavente, and Alan R. Shuldiner) and U01 GM074492-05S109 (principal investigator, Julie A. Johnson).

WUSTL: Washington University St. Louis Stroke Study (WUSTL): The collection, extraction of DNA from blood, and storage of specimens were supported by 2 NINDS NIH grants (P50 NS055977 and R01 NS8541901). Basic demographic and clinical characterization of stroke phenotype was prospectively collected in the Cognitive Rehabilitation and Recovery Group (CRRG) registry. The Recovery Genomics after Ischemic Stroke (ReGenesIS) study was supported by a grant from the Barnes-Jewish Hospital Foundation.

The South London Ethnicity and Stroke Study (SLESS)

The South London Ethnicity and Stroke Study (SLESS) is a prospective study begun in 1999 that has recruited consecutive black patients with stroke from a contiguous catchment area covered by 3 hospitals in South London (Guy's and St Thomas' Hospitals, King's College Hospital, and St George's Hospital).[26, 27] Ethnicity was defined according to the UK Census 2001 definition and classified as Black African or Black Caribbean. Recruitment of black controls was done by random selection from General Practice lists in the catchment areas of St George's, Guys and St Thomas, and King's College Hospital between 1999 and 2012. Potential controls were selected from age and gender strata matched to stroke cases. Furthermore, controls were identified within St George's University of London and St George's Hospital staff and contacted via email. Additionally, posters inviting healthy Black African and Black Caribbean individuals were displayed in local leisure centres, General Practice surgeries, churches and communities centres within the same catchment area as the that of the cases. The study was reviewed and approved by the Wandsworth Local Research Ethics Committee, and informed consent was obtained from all participants. One consultant neurologist performed stroke subtyping using data collected on a standard proforma with additional review of all original brain imaging in all patients, as well as review of original notes when necessary. The pathophysiological Trial of Org 10172 in Acute Stroke Treatment (TOAST) subtyping classification was used for subtyping of ischemic stroke.
Stroke ascertainment in SLESS: One consultant neurologist performed stroke subtyping using data collected on a standard proforma with additional review of all original brain imaging in all patients, as well as review of original notes when necessary. The pathophysiological Trial of Org 10172 in Acute Stroke Treatment (TOAST) subtyping classification was used for subtyping of ischemic stroke. We would like to acknowledge Dr. Giosue Gulli (Neurology, Frimley Park Hospital, Surrey, UK), Dr. Loes C. A. Rutten-Jacobs (Department of Clinical Neurosciences, Stroke Research group, University of Cambridge, Cambridge, UK), Dr. Lalit Kalra (Department of Basic and Clinical Neurosciences, Institute of Psychiatry, Psychology and Neurosciences, King's College London, London, UK), Anthony G. Rudd (Division of Health and Social Care Research, King's College London, London, UK), and Charles D. A. Wolfe (Division of Health and Social Care Research, King's College London, London, UK) for their significant contributions to the SLESS collaboration.

Siblings with Ischemic Stroke Study (SWISS)

Population. SWISS is a prospective multicenter affected sibling pair study of first-ever or recurrent ischemic stroke.[28] 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[29] 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.[30]

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).

Ischemic Stroke Genetic Study (ISGS)

Population. ISGS is a multicenter inception cohort study.[31] 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,[25] the Oxfordshire Community Stroke Project[32], and the Baltimore-Washington Young Stroke Study[24]. 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).

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.[33] 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.[34] 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.[35] 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. [25] 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. 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.[36] 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[37]. 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.[38, 39] 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.

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). We would like to acknowledge Dr. Salman Tajuddin for his contributions to this effort.

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.[40] 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.[41] Subtype (cardio-embolic, large-vessel, and small vessel) specific meta- analyses were also performed in METASTROKE.

SiGN Replication Cohort National Institute of Neurological Disorders and Stroke (NINDS) Stroke Genetics Network (NINDS-SiGN) Replication Cohort

The Stroke Genetics Network (SiGN) study was funded by a cooperative agreement grant from the National Institute of Neurological Disorders and Stroke (NINDS) U01 NS069208. Genotyping services were provided by the Johns Hopkins University Center for Inherited Disease Research (CIDR), which is fully funded through a federal contract from the National Institutes of Health (NIH) to the Johns Hopkins University (contract No.HHSN268200782096C). The Biostatistics Department Genetics Coordinating Center at the University of Washington (Seattle) provided more extensive quality control of the genotype data through a subcontract with

CIDR. Additional support to the Administrative Core of SiGN was provided by the Dean's Office, University of Maryland School of Medicine. The SiGN replication cohort consists of European and Hispanic individuals from the following studies: BASICMAR, GCNKSS, GRAZ, KRAKOW, Leuven Stroke Genetics Study (LSGS), Lund Stroke Register (LSR), Malmo Diet and Cancer (MDC) Study, Middlesex County Ischemic Stroke Study (MCISS), Miami Stroke Registry and Biorepository (MIAMISR, Nurses' Healthy Study (NHS), Northern Manhattan Study (NOMAS), Reasons for Geographic and Racial Differences in Stroke (REGARDS), Sahlgrenska Academy Study on Ischemic Stroke (SAHLSIS), Secondary Prevention of Small Subcortical Strokes (SPS3), and Washington University St. Louis (WUSTL).

NINDS-SIGN consortium

BASEede datos de ICtus del hospital del MAR (BASICMAR)
BASICMAR is an ongoing prospective study of all acute strokes assessed since 2005 at the IMIM-Hospital Universitari del Mar (Barcelona, Spain). It includes both first-ever and recurrent strokes. There were no exclusion criteria regarding age or race-ethnicity of the individuals. All patients had an electrocardiogram (ECG), a blood analysis, and neuroimaging at the acute stage. Additional diagnostic procedures were s performed when clinically indicated. A follow-up of three months after stroke was completed for all survivors.
Ischemic stroke etiologic subtypes were classified according to TOAST criteria.[25] For this study, only individuals of European origin with ischemic stroke were selected from BASICMAR, with eligible events defined as a clinical syndrome of any duration associated with a radiographically proven acute infarct, without radiographic evidence of a demyelinating or neoplastic disease or other structural disease including primary intracerebral hemorrhage.

GCNKSS
The GCNKSS is a population-based epidemiologic study of stroke in blacks and whites that is designed to measure temporal trends and racial differences in incidence of stroke. The catchment area includes two southwestern Ohio, U.S.A., counties (Hamilton, which includes the city of Cincinnati, and Clermont to the east) and three Northern Kentucky, U.S.A., counties (Boone, Kenton, and Campbell) to the south of Cincinnati across the Ohio River. As part of the GCNKSS, for calendar years 1999 and 2005, prospective cohorts of first-ever and recurrent ischemic stroke cases were assembled using "hot pursuit" methodology at all local hospitals in the region (18 in 1999, and 17 in 2005), except for one hospital that is solely devoted to treating pediatric cases. Participants remained eligible if they were in a treatment trial, but participation in a treatment trial was not required for enrollment. Subjects with all degrees of severity of stroke were eligible, and no particular racial group was intentionally oversampled (about 80% were white participants and 20% black). Study research nurses prospectively screened inpatient admission and emergency department logs to identify acute ischemic stroke patients. When a case was identified and the treating physician had given permission to approach the patient, a study nurse asked the subject or a proxy (the most closely related competent individual, preferably a person living with the subject prior to the stroke) to consent to participate in the cohort. After consent was granted, a study nurse performed an extensive interview, and a blood sample was obtained for genetic analysis. In addition, a study nurse abstracted information about the individual, the subject's medical history, the stroke event, and imaging studies from the hospital chart. A study physician reviewed every abstract, along with the imaging studies, to

verify that an acute stroke had occurred, and to classify the event according to TOAST[25] and CCS criteria.[42]

GRAZ
Between 1994 and 2003, subjects with first-ever and recurrent ischemic strokes admitted to the stroke unit of the Department of Neurology, Medical University of Graz (Graz, Austria) were included. All race-ethnic groups were eligible and there was no intentional oversampling. All age groups were allowed, though only subjects above the age of 18 were admitted to our department. Ischemic stroke was defined as an episode of focal neurological deficits with acute onset and lasting > 24 hours. There were no selection criteria based on stroke severity. Those individuals in treatment trials were excluded. 685 subjects were eligible to participate in this study (278 women, 407 men). All cases were Caucasian. Mean age was 68.9 ± 13.8 years with an age range from 19 – 101 years. In addition to a standardized protocol including a laboratory examination and carotid ultrasound or magnetic resonance angiography and ECG, 304 subjects underwent neuroimaging by CT and 381 by MRI. More extensive cardiac examination, including transesophageal echocardiography or transthoracic echocardiography and Holter, was performed in subjects with suspected cardiac embolism. Stroke subtypes were assessed according to modified TOAST criteria1 and were conducted by trained stroke neurologists.

KRAKOW
All consecutive subjects with ischemic stroke (fulfilling WHO criteria[43](3)) who were admitted to the Stroke Unit at the Jagiellonian University (Krakow, Poland) and who provided informed consent were included in the study. The Stroke Unit serves as a stroke emergency center for one district of Krakow, Poland (200,000 inhabitants) and as a referral center for South East Poland (up to 15% of all admissions). For this on-going, prospective single-center, hospital-based study participants with first ever or recurrent strokes were recruited from January 22, 2002 to September 9, 2010. The local research ethics committee approved the study. Participants in treatment trials were excluded. All subjects were of European origin. Stroke severity was not a criterion for inclusion or exclusion. All cases had performed clinically relevant diagnostic workup, including brain imaging with CT (100%) and/or MRI (up to 20%) as well as ancillary diagnostic investigations including duplex ultrasonography of the carotid and vertebral arteries (approximately 90%), and transthoracic echocardiography (approximately 70%). Magnetic resonance angiography (MRA), computed tomographic angiography (CTA), and ambulatory ECG monitoring, transesophageal echocardiography and blood tests for hypercoagulability were performed. Stroke cases were classified into etiologic subtypes according to TOAST.[25] All cases were phenotyped independently by two experienced stroke neurologists with review of original imaging. Cases were subsequently classified additionally using the CCS system[42].

Leuven Stroke Genetics Study (LSGS)
Cases of European descent with cerebral ischemia, defined as a clinical stroke with imaging confirmation or a TIA with a new ischemic lesion on diffusion-weighted imaging, who were admitted to the Stroke Unit of the University Hospitals (Leuven, Belgium) were enrolled in the LSGS between 2005 and 2009. All participants from the LSGS study underwent brain imaging (MRI in 91% of patients, CT in the remainder) and a standardized protocol including lab examination, carotid ultrasound or CTA and cardiac examination (echocardiography and ambulatory ECG monitoring) in all patients. Based on clinical presentation and results from the

diagnostic work-up, cases were classified into ischemic stroke etiologic subtypes according to modified TOAST criteria[25] by a single reviewer. The reviewer had access to all information and imaging. Large-vessel disease was defined as either occlusive or significant stenosis (corresponding to > 50% luminal diameter reduction according to North American Symptomatic Carotid Endarterectomy Trial (NASCET) criteria[44] of a clinically relevant pre-cerebral or cerebral artery, presumably due to atherosclerosis. Carotid ultrasound was used as a screening tool, and in principle, additional imaging with CTA or MRA was performed when a high-grade stenosis was identified. In case CTA was used as the primary imaging modality, stenosis was confirmed by carotid ultrasound. In case of posterior circulation infarcts on imaging, CTA or MRA was used as the primary imaging modality to determine the degree of stenosis. Probable causes of cardiac embolism were excluded. The presence of a patent foramen ovale was not considered a cardiac source in this context. Intracranial atherosclerosis was considered present only if repeat imaging after at least one week revealed a similar degree of stenosis or persistent occlusion. If not, the findings were interpreted as an embolism from a proximal source. Small-vessel disease was defined as a symptomatic infarct of < 20 mm on DWI in areas supplied by single, small penetrating branches from middle cerebral artery, posterior cerebral artery or basilar artery in the absence of both a cardioembolic source and significant stenosis/occlusion due to atherosclerosis of an appropriate major brain artery. Cardioembolic stroke was defined as ischemic stroke in the presence of atrial fibrillation, sick sinus syndrome, myocardial infarction in the past four weeks, cardiac thrombus, infective endocarditis, atrial myxoma, prosthetic mitral or aortic valve, valvular vegetations, left ventricular akinetic segment, dilated cardiomyopathy, or patent foramen ovale or atrial septal aneurysm. Significant stenosis/occlusion due to atherosclerosis of an appropriate pre-cerebral or cerebral artery should be excluded. Other determined cause of stroke included those with arterial dissection, vasculitis, hematologic disorders, monogenic syndromes and complications of cardiovascular procedures. Dissection was diagnosed by typical findings on contrast-enhanced MRA and T1-fat suppressed MRI. Cryptogenic stroke was defined when no cause was identified despite an extensive evaluation. Strokes associated with significant aortic arch atheroma with plaques of ≥ 4 mm were also considered cryptogenic strokes. In addition to this primary classification, cases were reclassified using CCS.[42]

Lund Stroke Register (LSR)

The LSR is an ongoing study including consecutive subjects with first-ever stroke since March 1, 2001 from the local uptake area of Skåne University Hospital, Lund (Sweden). Stroke was defined using the WHO criteria.[43] Subjects aged 18 years or older with stroke caused by cerebral infarct, intracerebral hemorrhage or subarachnoid hemorrhage are included. Cases are included regardless of stroke severity, race-ethnic group belonging, or participation in any treatment trial. Those with iatrogenic or traumatic stroke are excluded.

In the discovery phase of the SiGN study, subjects from LSR with first-ever ischemic stroke between March 1, 2001 and February 28, 2010 were included if they or their next of kin provided informed consent. Age over 90 years was set to 90 years to maintain anonymity. Every participant underwent CT, MRI, or autopsy of the brain; and ECG. Echocardiography, ultrasound, CTA or MRA of cerebral arteries was performed when judged clinically relevant. The subtype of ischemic stroke was determined using CCS.[42]

For the secondary phase of SiGN, LSR individuals not included in the SiGN discovery phase participated after genotyping in the South Swedish genome-wide association study as follows:

first ever ischemic stroke cases recruited in 2006 and 2010 to 2012, and age- and sex-matched LSR control subjects without stroke recruited in 2001 to 2002 and 2006 to 2007 from the same geographical area with use of the official Swedish population register.

Malmo Diet and Cancer (MDC) Study

The MDC study is a population-based prospective cohort study. A total of 30,447 individuals, 45 to 73 years old, 60% women, attended a baseline examination between February 1991 and September 1996. Between 1992 and 1994, a total of 6,103 randomly selected subjects attended an extended baseline examination with the purpose of studying the epidemiology of cardiovascular diseases (the MDC cardiovascular cohort, MDC-CC). At the baseline examination, 23% of the participants were smokers, 16% used anti-hypertensive medication, 14% were obese (body mass index > 30 kg/m^2), 88% were born in Sweden and > 99% were born in Europe. Genotyping was performed using the Illumina Infinium Omni5 platform with exome content. Incidence of stroke was monitored prospectively from the baseline examination in 1992 to 1996 until December 31, 2008 using national registries. Stroke was defined according to the WHO criteria.[43] By definition, patients with transient ischemic attacks are excluded. The stroke subtypes are coded according to International Classification of Diseases revision 9. Cerebral infarction (International Classification of Diseases code 434) is diagnosed when CT, MRI, or autopsy verifies the infarction in location corresponding to the focal neurology or excludes hemorrhage and nonvascular disease.

Middlesex County Ischemic Stroke Study (MCISS)

The MCISS was initiated as a prospective hospital based stroke registry at the New Jersey Neuroscience Institute (Edison, NJ, U.S.A.). All cases over age 18 years were included, and no specific ethnic/racial group was targeted or excluded. From 2000 to 2009, 1,139 subjects with ischemic strokes were enrolled in this registry. There was no selection criterion based upon stroke severity, and both first-ever and recurrent strokes were included. Cases that were participants in treatment trials were not excluded. The major race/ethnic groups are Whites (67.2%), African Americans (14.3%), Asian Indians (8.2%), Hispanic (5.5%) and others (4.8%, Chinese and other Asians). All subjects with clinical suspicion of a stroke were admitted through the emergency room to a dedicated stroke unit supervised by a vascular neurologist. After a history and neurological examination, a standardized series of investigations were performed: complete blood count and differential, comprehensive metabolic panel, electrolytes, blood urea nitrogen, creatinine, lipid panel (total cholesterol, low-density lipoprotein, high-density lipoprotein, triglyceride levels, homocysteine levels, a cerebral MRI/MRA (if the MRI could not be performed, a head CT scan was done), carotid duplex ultrasound, ECG and an echocardiogram. The diagnosis of cerebral infarct was confirmed by the imaging studies. The epidemiological and clinical data on these participants was collected prospectively. Two independent investigators (one of which was a board-certified neurologist with expertise in vascular neurology) reviewed the data, and all strokes were classified into etiological subtypes using TOAST criteria[25]. In addition, the Oxfordshire stroke classification[32] was applied, and the vascular distribution of stroke was tabulated. All procedures, including the generation of the databases and recruitment of the stroke subjects, were conducted following Institutional Review Board policies and procedures at the New Jersey Neuroscience Institute/JFK Hospital.

Miami Stroke Registry and Biorepository (MIAMISR)
The MIAMISR at the University of Miami/Jackson Memorial Hospital (Miami, FL, U.S.A.) is an ongoing prospective hospital registry of consecutive patients subjects with prevalent stroke (ischemic and hemorrhagic) and TIA with available neuroimaging (CT or MRI) who provide informed consent. There are no specific exclusion criteria with the respect to age, stroke severity, disability or participation in treatment trials. It was established in November of 2008 in order to investigate stroke type, ischemic stroke subtypes, stroke genetics and stroke outcomes in diverse ethnic population of Miami. The stroke population is predominately Hispanic (63%), with Cuba (32%), Nicaragua (4.8%), Colombia (4.8%), and Puerto Rico (4.1%) contributing the most subjects. Jackson Memorial Hospital is a 1,550-bed county hospital affiliated with the University of Miami with approximately 900 stroke and TIA admissions per year. Demographic and clinical data along with blood samples for genetic and other research have been collected prospectively during the hospitalizations. Follow-up information was obtained at 90 days by telephone interview or in person. Trained research staff obtained written informed consent from the stroke patients or the health care proxy when available for participation in MIAMISR.[45]

Nurses' Healthy Study (NHS)
The NHS cohort consists of 121,700 female registered nurses aged 30 – 55 years who were residing in 11 U.S. states and who were enrolled in 1976 through responding to a mailed questionnaire on their medical history and lifestyle practices. They have been followed with biennial mailed questionnaires collecting information on disease risk factors and health status. From 1989 – 1990, blood samples were collected from 32,826 participants. Among these participants, we prospectively identified incident strokes and confirmed ischemic stroke cases by medical record review. Clinical symptoms consistent with stroke and exclusion of alternate etiologies were required for classification of stroke. Virtually all cases had imaging, but confirmation on CT or MRI was not required. No participants were excluded based on race/ethnicity. Neither stroke severity nor enrollment in a treatment trial was part of the eligibility criteria.

Northern Manhattan Study (NOMAS)
NOMAS is an ongoing population-based study designed to determine stroke incidence, risk factors and outcome in an urban multiethnic population.[46] NOMAS started in 1993 as a case-control study of index ischemic stroke cases admitted to the Columbia University Presbyterian Medical Center (New York, NY, U.S.A.) and affiliated hospitals and matching community controls (Northern Manhattan Stroke Study, NOMASS) and continued as a prospective stroke incidence study by following up controls in 1997 (NOMAS). Demographic and clinical data were collected prospectively during the hospitalizations and annually by phone or in person. Genetic samples were derived from two sources: (a) the population-based case-control study conducted from 1993-98 (NOMASS) and (b) the ongoing prospective cohort study (NOMAS). First-ever ischemic stroke cases were identified for the case-control study by screening of patient admissions, discharge codes, and referrals for neuroimaging at 15 acute care hospitals in the defined study area and multiple approaches to monitor for non-hospitalized cases. Incident ischemic stroke cases were identified from the prospective cohort study through follow-up visits and scheduled telephone contacts. Ischemic stroke cases from both sources were followed at 6 months by telephone and then annually afterwards in order to assess functional status and other outcomes. The administrative coordinating center of NOMAS moved from New York to Miami

in 2007. The Institutional Review Boards of both institutions, Columbia University and the University of Miami (Miami, FL, U.S.A.), approved the study.

Reasons for Geographic and Racial Differences in Stroke (REGARDS)
The REGARDS study is a U.S. national, population-based, longitudinal cohort of 30,239 African American and white adults aged ≥ 45 years, recruited January 2003 to October 2007 with ongoing follow-up. Suspected stroke is queried every six months and triggered by participant self-report of stroke, stroke symptom(s), hospitalization, or proxy report of death. Stroke severity and participation in a treatment trial did not limit inclusion in this study. Medical records for these reported events are retrieved and reviewed by at least two members of a committee of stroke experts with disagreements resolved by a third adjudicator. A symptom-based approach, independent of neuroimaging outcome, is used to confirm events using the WHO definition of stroke.[43] An infarct did not need to be seen on brain imaging to be included in this study. Ischemic stroke subtype classification is conducted using the TOAST system.[25, 47]

Sahlgrenska Academy Study on Ischemic Stroke (SAHLSIS)
SAHLSIS is a case-control study of ischemic stroke based in Gothenburg, Sweden.[48] Adult subjects who presented with first-ever or recurrent acute ischemic stroke before 70 years of age were recruited consecutively at stroke units in western Sweden from 1998 to 2012. All participants were of European origin. Patients were not excluded based on stroke severity or whether they were enrolled in a treatment trial. All participants underwent ECG and neuroimaging at the acute stage (all by CT and 58% also by MRI). Additional diagnostic work-up was performed when clinically indicated. Inclusion criteria was ischemic stroke which was defined as an episode of focal neurological deficits with acute onset and lasting > 24 hours or until death, with no apparent non-vascular cause, and no signs of primary hemorrhage on brain imaging. Subjects were excluded if they had a diagnosis of cancer at advanced stage, infectious hepatitis or human immunodeficiency virus. Ischemic stroke was assigned according to modified TOAST criteria.[49] Cases in this study were also classified using the CCS system.[42]

Secondary Prevention of Small Subcortical Strokes (SPS3)
The SPS3 trial (NCT00059306) is a randomized, multicenter, Phase 3 trial of antiplatelet therapy and antihypertensive therapy. Participants are randomized to aspirin alone or the combination of aspirin and clopidogrel. Participants are also randomized to two groups of blood pressure control: either to a target systolic blood pressure of 130 – 149 mm Hg or < 130 mm Hg. Principal eligibility criteria include man or woman at least 30 years of age with clinical evidence of small subcortical stroke and brain MRI evidence of small subcortical infarct. Subjects were required to not have evidence of ipsilateral symptomatic cervical carotid stenosis or high-risk cardioembolic sources for embolism. Further details of eligibility criteria have been published.[50] Primary outcomes include ischemic and hemorrhagic stroke. DNA samples were collected from 38% (1,139/3,020) of participants in the trial. These samples were obtained from 46% (37/81) participating centers across the U.S., Canada, Spain, Mexico, Chile, Ecuador and Peru. No additional eligibility criteria were necessary beyond informed consent for participating in the DNA sub-study. A total of 0.9% (10/1,139) of DNA donors gave sample at time of randomization, with the remainder donating at a later time point in follow-up.

Washington University St. Louis (WUSTL) Study

The WUSTL patient collection included ischemic stroke cases admitted to Barnes-Jewish Hospital/Washington University Medical Center (St. Louis, MO, U.S.A.) for genetic studies starting from August 1, 2008. Participants were identified for the genetic studies by screening admissions at our tertiary care hospital (both in the Emergency Department and on the Inpatient Stroke Service) without regard to age, race or ethnicity, including both first-ever and recurrent strokes. Subjects were retained in the study if their discharge diagnosis was ischemic stroke (without requirement for the stroke to be visualized on CT or MRI). Demographic and clinical data were collected prospectively during the hospitalization and at 90 days, by phone or in person. Genetic samples were derived from subjects enrolled in 3 different studies: (a) Acute tPA pharmacogenomics study (Ischemic stroke cases who received tPA and were admitted to BJH/Washington University; serial NIHSS scores,[51] and data on hemorrhagic transformation was collected) (b) Recovery Genomics after Ischemic Stroke Study (ReGenesIS, Ischemic stroke cases with NIHSS > 3 points without underlying chronic neurological disease, and expected survival up to 3 months after stroke), and (c) the Cognitive Recovery and Rehabilitation Group (CRRG) Registry (all ischemic stroke cases admitted to BJH/Washington University who consent to entering their clinical data into a stroke registry, and the collection of blood for genetic analysis). Cases that were part of a treatment trial were excluded from the tissue plasminogen activator pharmacogenomics and ReGenesIS study, but not the CRRG registry.

Supplemental Tables I-V- see attached excel document

Supplemental Table I: Demographics of COMPASS discovery cohorts

Supplemental Table II: COMPASS Ischemic Stroke Suggestive and Genome-wide Significant Sample Size Weighted Associations

Supplemental Table III: Demographics of Look-up of Replication cohorts

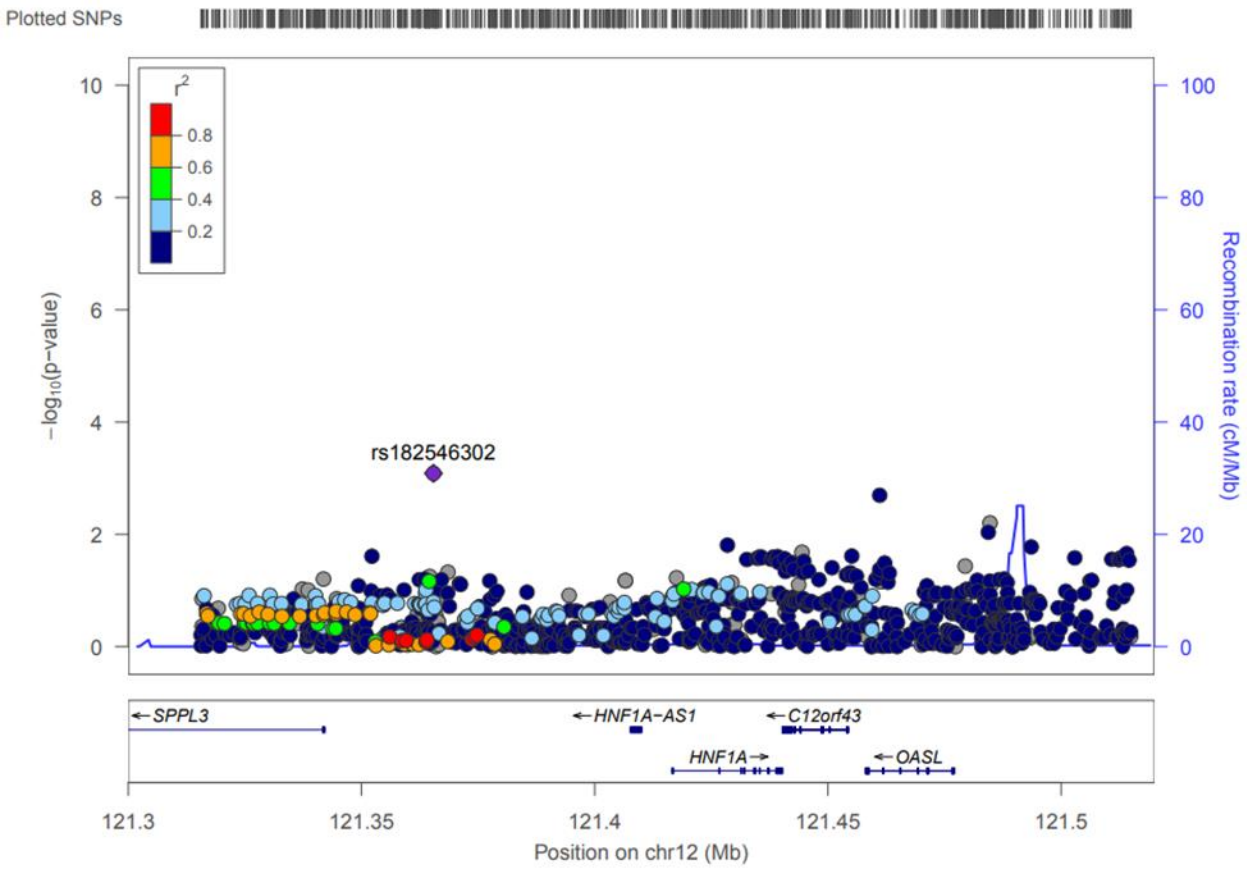
Supplemental Table IV: COMPASS HAPMAP and 1KG Association Comparisons

Supplemental Table V: COMPASS look-up of previous European-ancestry stroke associated loci

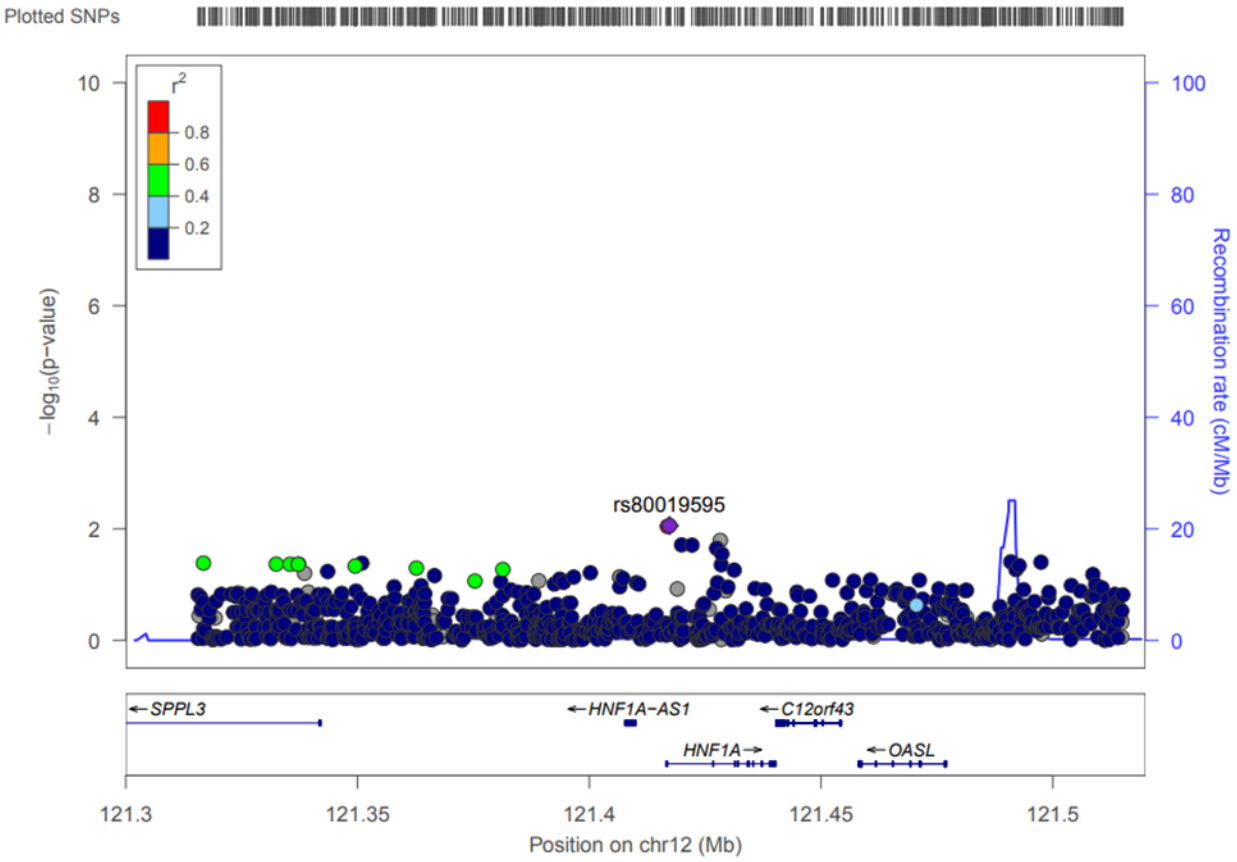
Supplemental Figures:

Supplemental Figure I. LocusZoom plots depicting "validation" associations for the HNF1A locus. A.) SiGN European validation population with LD based on hg19/1000 Genomes Nov 2014 EUR; B.) SiGN Hispanic validation population with LD based on hg19/1000 Genomes Nov 2014 AMR; and C.) METASTROKE European validation population.

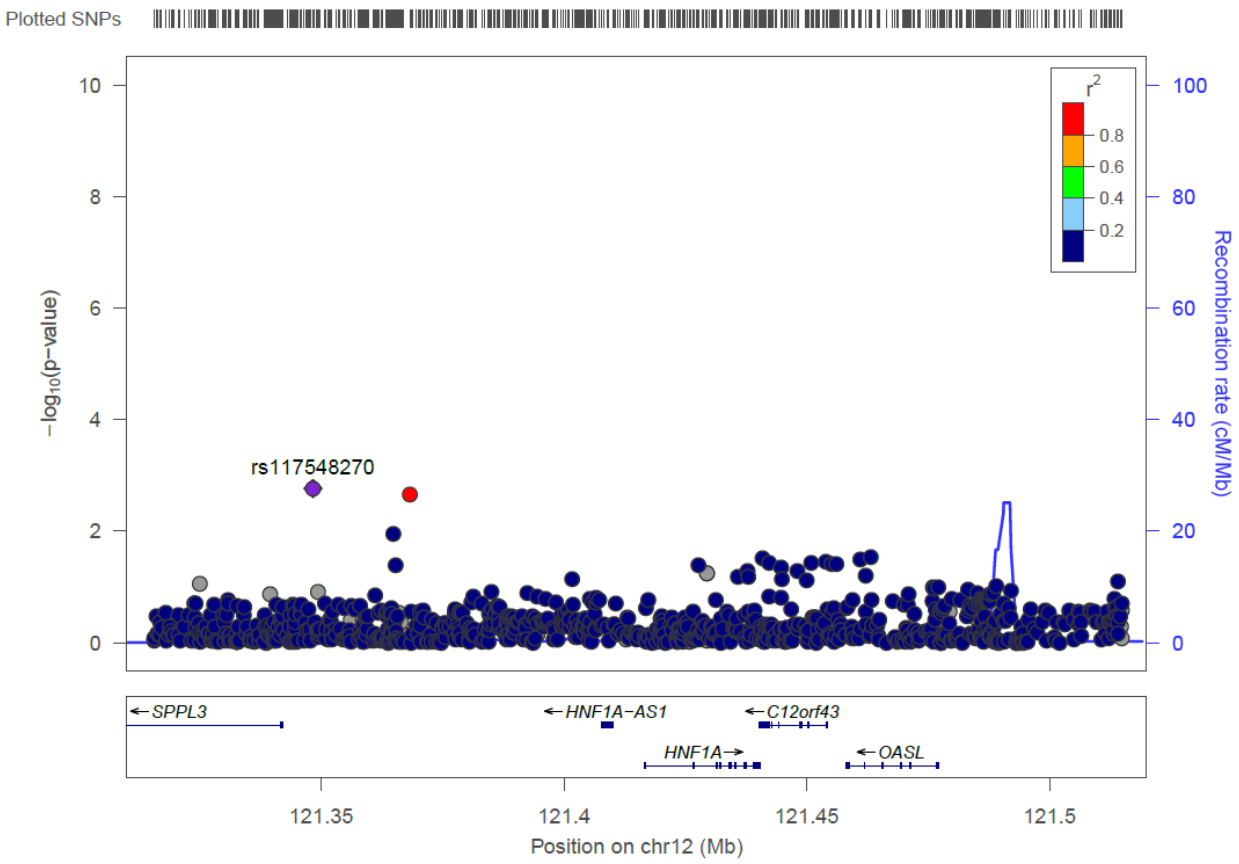
A. SiGN Europeans



B. SiGN Hispanics



C. METASTROKE Europeans



References

1. Willer CJ, Li Y, Abecasis GR. Metal: Fast and efficient meta-analysis of genomewide association scans. *Bioinformatics (Oxford, England)*. 2010;26:2190-2191
2. The ARIC Investigators. The atherosclerosis risk in communities (aric) study: Design and objectives. *Am J Epidemiol*. 1989;129:687-702
3. Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. Mach: Using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genetic epidemiology*. 2010;34:816-834
4. Li Y, Willer C, Sanna S, Abecasis G. Genotype imputation. *Annual review of genomics and human genetics*. 2009;10:387-406
5. Rosamond WD, Folsom AR, Chambless LE, Wang C-H, 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
6. National Survey of Stroke. National institute of neurological and communicative disorders and stroke. *Stroke*. 1981;12:I1-91
7. Fried LP, Borhani NO, Enright P, Furberg CD, Gardin JM, Kronmal RA, et al. The cardiovascular health study: Design and rationale. *Annals of epidemiology*. 1991;1:263-276
8. Longstreth W, 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
9. Price TR, Psaty B, O'Leary D, Burke G, Gardin J. Assessment of cerebrovascular disease in the cardiovascular health study. *Annals of epidemiology*. 1993;3:504-507
10. 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
11. Toole JF. Vitamin intervention for stroke prevention. *Journal of the neurological sciences*. 2002;203:121-124
12. O'Donnell M, Xavier D, Diener C, Sacco R, Lisheng L, Zhang H, et al. Rationale and design of interstroke: A global case-control study of risk factors for stroke. *Neuroepidemiology*. 2010;35:36-44
13. Delaneau O, Zagury J-F, Marchini J. Improved whole-chromosome phasing for disease and population genetic studies. *Nature methods*. 2013;10:5
14. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS genetics*. 2009;5
15. Taylor HA, Jr. The jackson heart study: An overview. *Ethnicity & disease*. 2005;15:S6-1-3
16. Carpenter MA, Crow R, Steffes M, Rock W, Heilbraun J, Evans G, et al. Laboratory, reading center, and coordinating center data management methods in the jackson heart study. *The American journal of the medical sciences*. 2004;328:131-144
17. Fuqua SR, Wyatt SB, Andrew ME, Sarpong DF, Henderson FR, Cunningham MF, et al. Recruiting african-american research participation in the jackson heart study: Methods, response rates, and sample description. *Ethnicity & disease*. 2005;15:S6-18-29

18. Payne TJ, Wyatt SB, Mosley TH, Dubbert PM, Guiterrez-Mohammed ML, Calvin RL, et al. Sociocultural methods in the jackson heart study: Conceptual and descriptive overview. *Ethnicity & disease*. 2005;15:S6-38-48
19. Wyatt SB, Akylbekova EL, Wofford MR, Coady SA, Walker ER, Andrew ME, et al. Prevalence, awareness, treatment, and control of hypertension in the jackson heart study. *Hypertension*. 2008;51:650-656
20. Keku E, Rosamond W, Taylor HA, Jr., Garrison R, Wyatt SB, Richard M, et al. Cardiovascular disease event classification in the jackson heart study: Methods and procedures. *Ethn Dis*. 2005;15:S6-62-70
21. 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
22. Howard VJ, Cushman M, Pulley L, Gomez CR, Go RC, Prineas RJ, et al. The reasons for geographic and racial differences in stroke study: Objectives and design. *Neuroepidemiology*. 2005;25:135-143
23. 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
24. 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
25. 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
26. Gulli G, Rutten-Jacobs LC, Kalra L, Rudd AG, Wolfe CD, Markus HS. Differences in the distribution of stroke subtypes in a uk black stroke population–final results from the south london ethnicity and stroke study. *BMC medicine*. 2016;14:77
27. Traylor M, Rutten-Jacobs L, Curtis C, Patel H, Breen G, Newhouse S, et al. Genetics of stroke in a uk african ancestry case-control study: South london ethnicity and stroke study. *Neurology Genetics*. 2017;3:e142
28. Meschia JF, Brown RD, Brott TG, Chukwudelunzu FE, Hardy J, Rich SS. The siblings with ischemic stroke study (swiss) protocol. *BMC medical genetics*. 2002;3:1
29. Worrall BB, Chen DT, Meschia JF. Ethical and methodological issues in pedigree stroke research. *Stroke*. 2001;32:1242-1249
30. Meschia JF, Brott TG, Chukwudelunzu FE, Hardy J, Brown Jr RD, Meissner I, et al. Verifying the stroke-free phenotype by structured telephone interview. *Stroke*. 2000;31:1076-1080
31. Meschia JF, Brott TG, Brown RD, Crook RJ, Frankel M, Hardy J, et al. The ischemic stroke genetics study (isgs) protocol. *BMC neurology*. 2003;3:4
32. 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
33. 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
34. 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

35. 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
36. 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
37. 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
38. Devlin B, Roeder K. Genomic control for association studies. *Biometrics*. 1999;55:997-1004
39. Bacanu SA, Devlin B, Roeder K. Association studies for quantitative traits in structured populations. *Genet. Epidemiol.* 2002;22:78-93
40. 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
41. Willer CJ, Li Y, Abecasis GR. Metal: Fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*. 2010;26:2190-2191
42. Ay H, Furie KL, Singhal A, Smith WS, Sorensen AG, Koroshetz WJ. An evidence-based causative classification system for acute ischemic stroke. *Annals of neurology*. 2005;58:688-697
43. Hatano S. Experience from a multicentre stroke register: A preliminary report. *Bulletin of the World Health Organization*. 1976;54:541-553
44. Eliasziw M, Smith RF, Singh N, Holdsworth DW, Fox AJ, Barnett HJ. Further comments on the measurement of carotid stenosis from angiograms. North american symptomatic carotid endarterectomy trial (nascet) group. *Stroke*. 1994;25:2445-2449
45. Romano JG, Arauz A, Koch S, Dong C, Marquez JM, Artigas C, et al. Disparities in stroke type and vascular risk factors between 2 hispanic populations in miami and mexico city. *Journal of stroke and cerebrovascular diseases : the official journal of National Stroke Association*. 2013;22:828-833
46. 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
47. Howard VJ, Kleindorfer DO, Judd SE, McClure LA, Safford MM, Rhodes JD, et al. Disparities in stroke incidence contributing to disparities in stroke mortality. *Annals of neurology*. 2011;69:619-627
48. Jood K, Ladenvall C, Rosengren A, Blomstrand C, Jern C. Family history in ischemic stroke before 70 years of age: The sahlgrenska academy study on ischemic stroke. *Stroke*. 2005;36:1383-1387
49. Olsson S, Holmegaard L, Jood K, Sjogren M, Engstrom G, Lovkvist H, et al. Genetic variation within the interleukin-1 gene cluster and ischemic stroke. *Stroke*. 2012;43:2278-2282
50. Benavente OR, White CL, Pearce L, Pergola P, Roldan A, Benavente MF, et al. The secondary prevention of small subcortical strokes (sps3) study. *International journal of stroke : official journal of the International Stroke Society*. 2011;6:164-175

51. Kasner SE. Clinical interpretation and use of stroke scales. *The Lancet. Neurology*. 2006;5:603-612