Published in final edited form as:
Circ Cardiovasc Genet. 2012 April 1; 5(2): 210–216. doi:10.1161/CIRCGENETICS.111.962191.

Associations Between Incident Ischemic Stroke Events and Stroke and Cardiovascular Disease-Related GWAS SNPs in the Population Architecture Using Genomics and Epidemiology (PAGE) Study

Cara L. Carty, PhD[1], **Petra Bůžková, PhD**[2], **Myriam Fornage, PhD**[3,4], **Nora Franceschini, MD**[5], **Shelley Cole, PhD**[6], **Gerardo Heiss, MD, PhD**[5], **Lucia A. Hindorff, PhD, MPH**[7], **Barbara V. Howard, PhD**[8,9], **Sue Mann, MPH**[1], **Lisa W. Martin, MD**[9], **Ying Zhang, PhD**[10], **Tara C. Matise, PhD**[11], **Ross Prentice, PhD**[1], **Alexander P. Reiner, MD, MS**[12], and **Charles Kooperberg, PhD**[1]

[1]Public Health Sciences, Fred Hutchinson Cancer Rsrch Ctr

[2]Dept of Biostatistics, University of Washington, Seattle, WA

[3]Inst of Molecular Medicine, University of Texas Health Sciences Ctr at Houston

[4]Division of Epidemiology, School of Public Health, University of Texas Health Sciences Ctr, Houston, TX

[5]Dept of Epidemiology, University of North Carolina, Chapel Hill, NC

[6]Dept of Genetics, Texas Biomedical Rsrch Inst, San Antonio, TX

[7]Office of Population Genomics, National Human Genome Rsrch Inst, Bethesda, MD

[8]Medstar Health Rsrch Inst, Washington, DC

[9]George Washington University School of Medicine, Washington, DC

[10]University of Oklahoma Health Sciences Ctr, Oklahoma City, OK

[11]Dept of Genetics, Rutgers University, Piscataway, NJ

[12]Dept of Epidemiology, University of Washington, Seattle, WA

Abstract

Background—Genome-wide association studies (GWAS) have identified loci associated with ischemic stroke (IS) and cardiovascular disease (CVD) in European-descent individuals, but their replication in different populations has been largely unexplored.

Methods and Results—Nine single-nucleotide polymorphisms (SNPs) selected from GWAS and meta-analyses of stroke and 86 SNPs previously associated with myocardial infarction and CVD risk factors including blood lipids (HDL, LDL, triglycerides), type 2 diabetes and body mass index were investigated for associations with incident IS in European Americans (EA) N=26,276;

Corresponding Author: Cara L. Carty, MS, PhD Fred Hutchinson Cancer Research Center 1100 Fairview Ave N. / M3-A410 Seattle, WA 98109 Tel: 206-667-2873 Fax: 206-667-4142 ccarty@fhcrc.org.

Conflict of Interest Disclosures: none

African Americans (AA) N=8970; and American Indians (AI) N= 3570 from the Population Architecture using Genomics and Epidemiology Study. Ancestry-specific fixed effects meta-analysis with inverse variance weighting was used to combine study-specific log hazard ratios from Cox proportional hazards models. Two of 9 stroke SNPs (rs783396 and rs1804689) were associated with increased IS hazard in AA; none were significant in this large EA cohort. Of 73 CVD risk factor SNPs tested in EA, two (HDL and triglycerides SNPs) were associated with IS. In AA, SNPs associated with LDL, HDL and BMI were significantly associated with IS (3 of 86 SNPs tested). Out of 58 SNPs tested in AI, one LDL SNP was significantly associated with IS.

Conclusions—Our analyses showing lack of replication in spite of reasonable power for many stroke SNPs and differing results by ancestry highlight the need to follow-up on GWAS findings and conduct genetic association studies in diverse populations. We found modest IS associations with BMI and lipids SNPs, though these findings require confirmation.

Keywords

genetics of stroke; risk factors for stroke; genetics of cardiovascular disease; epidemiology

Introduction

Family studies suggest that stroke has a substantial genetic component.[1, 2] To date, a small number of genetic variants associated with stroke have been identified in genome-wide association studies (GWAS). These studies were conducted in largely European-descent populations, though the burden of stroke is higher in minority populations in the US.[3] Stroke incidence and mortality in African Americans and stroke incidence in American Indians are nearly twice that of European-Americans.[4]

Stroke is a heterogeneous disease consisting of several distinct subtypes, each having their own etiologies.[5] Ischemic stroke, the most common subtype accounting for ~87% of strokes,[6] shares many risk factors with cardiovascular disease due to a common etiological factor, atherosclerosis.[7] Shared risk factors for these diseases may also encompass genetic factors, such as gene variants involved in atherosclerosis. We sought to replicate associations between incident ischemic stroke and ischemic stroke SNPs identified in previous genome-wide association (GWA) or meta-analysis studies of European populations in a large sample of European Americans (EA), and investigated whether associations were consistent for African Americans (AA). Furthermore, we tested whether SNPs associated with myocardial infarction (MI) and cardiovascular disease risk factors (blood lipids, type 2 diabetes [T2D] and body mass index [BMI]) are also associated with the incident ischemic stroke hazard in EA, AA and American Indians (AI) from the Population Architecture using Genomics and Epidemiology (PAGE)[8] Study.

Methods

As part of the PAGE Study, four cohort studies with adjudicated stroke data were included in these analyses: Atherosclerosis Risk in Communities (ARIC),[9] Cardiovascular Health Study (CHS),[10] Strong Heart Study (SHS),[11] and Women's Health Initiative (WHI).[12] Participants were censored at the first incident ischemic stroke event and individuals with a history of stroke or transient ischemic attack at baseline were excluded. All studies were approved by local institutional review boards and all participants gave informed consent. The study populations are briefly described in Table 1 and below.

The Women's Health Initiative (WHI) recruited 161,838 postmenopausal women aged 50–79 yrs. old from 40 clinical centers in the US between 1993 and 1998.[12, 13] WHI consists of an observational study, two clinical trials 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 the WHI cohort, n=21,000, were selected for PAGE genotyping and for inclusion in these analyses. Women were selected based on self-reported history of disease, incident event outcomes, DNA availability and consent, and racial/ethnic diversity. In the association analyses, inverse-probability weighting was used to account for this non-random sampling. Sample weights ranged from 1 to 60.4. DNA was extracted from blood samples collected at baseline. Incident stroke events were identified by semi-annual questionnaires and adjudicated following medical record review.

The Atherosclerosis Risk in Communities (ARIC) Study is a bi-racial population-based cohort recruited from four U.S. communities: Forsyth County, North Carolina; Jackson, Mississippi; suburban areas of Minneapolis, Minnesota; and Washington County, Maryland, USA.[9] The 15,792 men and women in ARIC, including 11,478 non-Hispanic white participants, were between 45-64 years of age at baseline and were followed up for possible stroke events (http://www.cscc.unc.edu/aric/) through annual phone interviews, follow-up examinations, community hospital surveillance, and death certificates. Reported hospitalizations led to screening and, if suitable, to medical record abstraction. Potential stroke events were selected for medical records abstraction if the discharge diagnosis included a cerebrovascular disease code (International Classification of Diseases, 9th Revision, codes 430 to 438), if a cerebrovascular procedure was mentioned in the discharge summary, or if the CT or MRI report showed evidence of acute cerebrovascular disease.[14] All suspected events were classified by computer algorithm and also by an expert physician reviewer, blinded to the automated results. A second physician reviewer adjudicated disagreements between the computer and the initial reviewer.

The Cardiovascular Health Study (CHS) is a population-based longitudinal study of risk factors for cardiovascular disease in adults 65 years of age or older.[15] A total of 5,201 predominantly European Americans were recruited in 1989-1990 from random samples of Medicare eligibility lists at four field centers (Forsyth County, North Carolina; Sacramento County, California; Washington County, Maryland; Pittsburgh, Pennsylvania), followed by an additional 687 African Americans recruited in 1992-1993 (total n=5,888). CHS participants completed standardized clinical examinations and questionnaires at study enrollment and at up to nine annual follow-up visits and are being followed for clinical events. Incident strokes were identified at semi-annual exams and through community surveillance and adjudicated by neurologists.[10]

Funded by the National Heart, Lung, and Blood Institute, the Strong Heart Study (SHS) is a population-based longitudinal study of 4,549 American Indian men and women recruited from 13 communities and from centers located in 3 US geographic areas (http://strongheart.ouhsc.edu/). SHS participants aged 45-74 years underwent a baseline (1989-1991) and then two follow-up (1993-1995 and 1996-1999) examinations. Incident ischemic stroke events were identified at yearly interviews and/or examinations and adjudicated by neurologists.

SNPs were identified using the National Human Genome Research Institute's Catalog of Published GWAS[16], a comprehensive database of GWA investigations. GWAS SNPs and well-replicated SNPs from meta-analyses associated with stroke (n=9), myocardial infarction/coronary heart disease (CHD) (n=16), T2D (n=19), high density lipoprotein cholesterol (HDL) (n=17), low density lipoprotein cholesterol (LDL) (n=16), triglycerides (TG) (n=4) and BMI (n=14) for a maximum total of 86 SNPs as of Jan 1, 2009 were genotyped by each study in PAGE (see Table 2 and Supplemental Tables 1-3 for lists of included SNPs).[17181920] PAGE quality control measures called for exclusion of SNPs with call rates <95% or with <98% concordance among duplicate samples, and exclusion of

individuals with call rates<95%, though WHI used more stringent criteria. For SNPs in ancestry-specific Hardy Weinberg disequilibrium ($p<5\times10^{-5}$), intensity plot clustering was reviewed manually and SNPs were excluded if clustering was judged to be poor. SNPs that were successfully genotyped, passed quality control measures and were available in at least two PAGE studies are included in the EA and AA analysis. For the AI analysis, all CVD risk factor SNPs passing QC in SHS were tested; however, the 9 stroke SNPs were not genotyped.

Genotype data were coded assuming additive genetic models, with each SNP coded as a count of the variant alleles (0, 1, or 2), unless otherwise specified. Ancestry-stratified Cox proportional hazards models were minimally adjusted for age, sex and study site (as appropriate in each sub-study). African American models were additionally adjusted for global ancestry to account for population substructure in WHI and ARIC; in CHS, study site was used as a proxy for ancestry. In SHS, AI models were adjusted for self-reported AI ancestry obtained by questionnaire. Study-specific results (or in the case of SHS, site-specific results) were combined in a fixed effects meta-analysis with inverse variance weighting using METAL.[21] Per allele hazard ratios (HR) and 95% confidence intervals (95%CI) from the meta-analysis are reported. Heterogeneity for the SNP effects across PAGE cohorts was assessed using I^2 and the Q-test statistic.[22] Aggregate data from the meta-analysis and individual tests of association from each PAGE study will be made available via dbGaP.[23] Results are presented unadjusted for multiple testing with presentation of the multiple testing implications for interpretation in the discussion. Power calculations were performed using Quanto[24] assuming unrelated participants, population risk of ischemic stroke=0.04, additive genetic models (or dominant or recessive models as appropriate based on prior reports), unmatched case-control study design, effect size equal to the 95% CI closest to the null based on prior reports in European-descent populations, and the average ancestry-specific allele frequencies and number of incident ischemic stroke cases listed in Table 1. Power estimates for GWAS-identified stroke SNPs were also corrected for 'winner's curse'[25] bias using the above assumptions with the exception of the effect size, which was based on bias-reduced estimates obtained using the method detailed in Zhong and Prentice.[26] The 'winner's curse' refers to the ascertainment bias affecting gene association results; genetic associations which are significant in the initial study (GWAS) may be overestimated, while other associations (potentially underestimated) are less likely to be published and selected for follow-up. Because the 'winner's curse' results are inflated, they can be difficult to replicate in follow-up studies, and if used for power calculations, can result in inflated power estimates. Zhong and Prentice[26] propose a method using conditional maximum likelihood estimation and quartile-based confidence interval procedures to estimate bias-corrected betas and selection-adjusted confidence intervals. We contrast power calculations using these bias-adjusted estimates with calculations using the original GWAS estimates.

Results

During follow-up, there were a total of 3,239 incident ischemic strokes in EA, 655 in AA and 163 in AI. A total of 9 stroke SNPs were investigated for their associations with IS in EA and AA (genotypes were unavailable in AI). None of these SNPs was significantly associated with risk of incident ischemic stroke in EA (Table 2). In AA, two SNPs (rs783396 and rs1804689) were significantly associated with ischemic stroke at $p<0.05$. The original reports for 3 of the 9 stroke SNPs described associations using dominant and recessive models rather than additive models; when we modeled SNPs similarly, results from the meta-analyses in AA and EA changed only slightly and remained non-significant (data not shown). In sensitivity analyses, using a smaller subset of the WHI EA women (n=1205) having more detailed stroke subtype information, we tested for stroke SNP

associations with large artery (n=189), cardioembolic (n=531) and lacunar strokes (n=485). Rs10486776 was significantly associated with large artery ischemic stroke, p=0.005, and rs7506045 was significantly associated with cardioembolic ischemic stroke, p=0.027; no other significant associations with ischemic stroke subtypes were found.

In EA, a total of 73 CVD risk factor SNPs were investigated (Supplemental Table 1); of these, the HDL SNP rs2156552 (p=0.006) and the triglycerides SNP rs2954029 (p=0.048) were significantly associated with ischemic stroke (Table 3). In AA, we tested a total of 86 CVD risk factor SNPs (Supplemental Table 2); three SNPs were associated with ischemic stroke, the HDL SNP rs1800961 (p=0.0006), the LDL SNP rs6544713 (p=0.022) and the BMI SNP rs11084753 (p=0.024) (Table 3). In American Indians (AI), a smaller subset of 58 CVD risk factor SNPs were tested (Supplemental Table 3). Rs754523, a SNP previously associated with LDL levels, was associated with ischemic stroke, p=0.0026 (Table 3).

Discussion

In this analysis of multi-ethnic populations from four large cohort studies, we followed up on previous gene association study findings for ischemic stroke in diverse ancestral groups and explored whether GWAS SNPs for CVD risk factors are also associated with ischemic stroke. While several SNPs have been previously associated with ischemic stroke, including exonic variants in the *MTHFR*, *AIM1* and *F5* genes, we were unable to replicate these associations in our large EA population. Interestingly, two SNPs were significantly associated with incident ischemic stroke in AA, rs783396 and rs1804689. Located in an exon of the absent in melanoma-1 (*AIM1)* gene whose expression is associated with melanoma tumor suppression, rs783396-A (Ala->Glu) is a non-synonymous variant. The variant was previously associated with an increased risk of ischemic stroke in EA[19] with an OR=5.79 (95% CI:2.66-12.59), which is consistent with our finding of an increased risk of stroke in AA. However, it is possible that the original estimate was inflated due to ascertainment bias termed the winner's curse.[25] Indeed, adjustment for winner's curse bias greatly attenuated the original estimates as well as our hypothetical power to replicate them. Rs1804689 (A allele) in the 5′ UTR of the *HPS1* gene was previously associated with an increased risk of ischemic stroke in CHS EA[20], but in our AA population (ARIC and WHI) it was associated with a significantly reduced risk of ischemic stroke. Different patterns of linkage disequilibrium in this gene region could potentially explain these contradictory findings; accordingly, the average allele frequencies for the PAGE AA (0.14) and EA (0.31) are markedly different.

Since the ischemic stroke subtype may be heterogeneous in its etiology and thus have different risk factors[27], we explored ischemic stroke subtypes in a smaller subset of WHI women having more detailed data in sensitivity analyses. We identified 2 SNPs associated with large artery and cardioembolic stroke in EA—though interestingly, these SNPs were not strongly associated with either ischemic stroke subtype in the original GWAS.[19] One important difference is that our analyses were conducted in EA women only, while the GWAS included both sexes. However, both subtype analyses have small sample sizes which may be more susceptible to spurious results.

Given our lack of replication along with our high power to replicate the lower confidence limit of previously reported or bias-adjusted effect sizes for almost half of the stroke SNPs in EA, we could suggest that previous findings may reflect type 1 error. Indeed, several others have commented on the lack of reproducibility of stroke-gene associations[28-30] citing low power, study design issues, population stratification and study heterogeneity as potential causes. Several of the previous stroke studies were conducted in the case-control setting and may be prone to potential selection and survival biases, in contrast to our analyses which

were conducted in cohorts with many years of longitudinal follow-up and which also included some fatal stroke cases. Yet, these explanations do not appear to account for the differences in population-specific findings in our analysis. A potential explanation for the differences in population-specific results is gene-environment interaction. It might be that these SNPs exert their effects most strongly in the presence of environmental or host risk factors which may differ between studies or between ancestry groups in the same study. Comparison of the prevalence of stroke risk factors reveals that BMI tends to be higher and T2D is more common in PAGE AA and AI than EA; hypertension is also much more common in AA than in the other groups. However, we saw the most suggestive findings for the lipids SNPs; the use of lipid lowering medications at baseline and during early follow-up was low in all groups since recruitment for each of the studies primarily occurred before the introduction and widespread use of the recent generation of lipid-lowering medications.

Although ischemic stroke and MI/CHD share many risk factors, SNPs associated with MI/CHD traits were not significantly associated with incident ischemic stroke in our population. The 9p21 locus has been previously associated with coronary heart disease[31] and myocardial infarction[32] in multiple GWAS, although findings for stroke outcomes have been mixed. A recent study[33] attributes these mixed results to stroke subtype heterogeneity and reported that the 9p21 locus was significantly associated with large artery ischemic stroke, which is consistent with prior associations for atherosclerotic disease of the coronary arteries and abdominal aorta.[32, 34] Chromosome 9p21 variants, rs10757278, rs1333049, rs2383206, rs2383207 and rs4977574 were included among the MI/CHD 9p21 SNPs we tested, but in sensitivity analyses of ischemic stroke subtypes, none were significantly associated with large artery (p=0.47-0.61) or cardioembolic (p=0.06-0.24) stroke in WHI EA. However, all of these 9p21 variants were modestly associated with small vessel (lacunar) ischemic stroke (p=0.04-0.07) in EA. While inconsistent with the previous report, this finding is exploratory and was conducted in a small sample of postmenopausal women.

Overall, few CVD risk factor SNPs, including other 9p21 variants, were associated with risk of ischemic stroke in our study, with the exception of 5 lipid SNPs and 1 BMI SNP. In AA, an LDL SNP (rs6544713) and an HDL SNP (rs1800961) were significantly associated with ischemic stroke. The LDL SNP was also associated with increased LDL concentrations in all three ancestral groups in PAGE, while the HDL SNP, a non-synonymous variant in the hepatocyte nuclear factor 4 alpha gene, was associated with lower HDL levels in EA and AI, and showed a similar trend in AA that was not significant.[35] These findings suggesting SNP-related lipid level differences are consistent with the increased risk of stroke that we find in our analysis, i.e. SNPs associated with adverse lipid concentration trends are associated with increased risk of stroke. Similarly, in AI, the *APOB* SNP rs754523-T was associated with a reduced risk of stroke; this SNP has been also been associated with reduced LDL levels in the PAGE study.[35] EA findings include rs2156522 which was significantly associated with ischemic stroke; this SNP has been previously associated with HDL concentrations in PAGE EA, but not in PAGE AA.[35] Conversely, the triglycerides SNP rs2954029 was significantly associated with an increased risk of stroke in EA, but slightly reduced triglyceride levels, which is contrary to what we might expect. And finally, the BMI SNP rs11084753, was associated with ischemic stroke in AA, but was not significantly associated with BMI in any of the PAGE race/ethnicity groups (Fesinmeyer MD, *et al.,* manuscript submitted to *Obesity*, 2011). These lipid and BMI SNP findings were generally not consistent between ancestral groups. In part, these differences may be explained by different linkage disequilibrium patterns across the loci in the different population groups. This scenario would be expected if the originally identified SNPs act as tags for the functional variant and the tags do not extend across race/ethnicity groups due to differences in linkage disequilibrium. Indeed, the majority of SNPs identified in GWAS are noncoding, located in intergenic or intronic regions,[16] although our non-synonymous HDL

SNP (rs1800961) is a notable exception. This SNP is also interesting because it is in a gene, hepatocyte nuclear factor 4 alpha (*HNF4A*), with potential pleiotropic effects. *HNF4A* variants have been associated with lower HDL levels, as well as with Type 2 diabetes and with levels of the coagulation factor, Factor VII, suggesting that this gene is involved in multiple pathways relevant to the stroke outcome.

Potential limitations of this study include study heterogeneity, lack of detailed ischemic stroke subtype information on the entire population and reduced power to detect modest effect sizes in the smaller American Indian population and for a number of SNPs in EA and AA. Our analysis of CVD risk factor SNPs was exploratory and with the exception of the HDL SNP rs1800961 in AA, our findings were not statistically significant after Bonferroni correction for multiple testing and thus should be interpreted cautiously until confirmed. We tested for study heterogeneity in the meta-analysis and also assessed the magnitude of heterogeneity using I^2, but these tests may be less robust when small numbers of sub-studies are analyzed[36], such as in our case. The significant SNPs we report did not show significant evidence of study heterogeneity; but it is possible that heterogeneity may have reduced our power to detect significant associations for other SNPs. In general, allele frequencies were consistent between the different studies for the EA and AA populations, though more heterogeneity was seen for the AI, likely due to differing levels of admixture at the different regional sites in SHS. Additionally, we report power estimates adjusted for the 'winner's curse' bias which tend to be more conservative than the naïve power estimates. Effect estimates (and power) for several SNPs dropped substantially when accounting for this bias suggesting that replication of these GWAS SNPs will be difficult in even large populations such as the PAGE EA with more than 3200 ischemic strokes.

In spite of these limitations, our study includes large numbers of ancestrally diverse individuals with adjudicated ischemic stroke outcomes, standardized methods and long follow-up. Much of our sample is population-based and not subject to biases that might be present in hospital-based case-control studies of stroke.

Conclusions

As fine-mapping and re-sequencing studies begin to investigate GWAS findings in more detail, data regarding GWAS SNP replication in independent samples and across different ancestry/ethnicity groups will be increasingly important for use in prioritizing SNPs for follow-up[37] and for enhancing our understanding of population differences in common diseases. Our study contributes information about the reproducibility and magnitude of prior GWAS-identified SNPs for ischemic stroke. While none replicated in our EA PAGE population, two associations did generalize to AA. Furthermore, in analyses investigating CVD pleiotropy for largely GWAS-based SNPs, we identified additional CVD risk factor SNPs which may also be associated with stroke in diverse US ancestral/ethnic groups, including AA and AI who are at high risk of this common disease.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The PAGE consortium thanks the staff and participants of all PAGE studies for their important contributions. The complete list of PAGE members can be found at http://www.pagestudy.org. In addition, the authors thank the WHI investigators and staff for their dedication, and the study participants for making the program possible. A full listing of WHI investigators can be found at: http://www.whiscience.org/publications/WHI_investigators_shortlist.pdf . The authors also thank the staff and participants of the ARIC study for their important contributions.

Funding Sources: The Population Architecture Using Genomics and Epidemiology (PAGE) program is funded by the National Human Genome Research Institute, supported by U01HG004803 (Causal Variants Across the Life Course), U01HG004798 (Epidemiologic Architecture of Genes Linked to Environment), U01HG004802 (Multi-Ethnic Cohort), U01HG004790 (Women's Health Initiative), and U01HG004801 (PAGE Coordinating Center). The contents of this paper are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health. The data and materials included in this report result from a collaboration between the following studies: Funding support for the "Epidemiology of putative genetic variants: The Women's Health Initiative" study is provided through the NHGRI PAGE program (U01HG004790). The WHI program is funded by the National Heart, Lung, and Blood Institute; NIH; and U.S. 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.

Funding support for the Genetic Epidemiology of Causal Variants Across the Life Course (CALiCo) program was provided through the NHGRI PAGE program (U01HG004803). The following studies contributed to this manuscript and are funded by the following agencies: The Atherosclerosis Risk in Communities Study is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute contracts (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C), R01HL087641, R01HL59367 and R01HL086694; National Human Genome Research Institute contract U01HG004402; and National Institutes of Health contract HHSN268200625226C. The Cardiovascular Health Study (CHS) is supported by contracts N01-HC-85079 through N01-HC-85086, N01-HC-35129, N01-HC-15103, N01 HC-55222, N01-HC-75150, N01-HC-45133, grants U01HL080295 and R01 HL087652 from the National Heart, Lung, and Blood Institute, with additional contribution from the National Institute of Neurological Disorders and Stroke. CHS GWAS DNA handling and genotyping was supported in part by National Center for Research Resources grant M01-RR00425 to the Cedars-Sinai General Clinical Research Center Genotyping core and National Institute of Diabetes and Digestive and Kidney Diseases grant DK063491 to the Southern California Diabetes Endocrinology Research Center. The Strong Heart Study (SHS) is supported by NHLBI grants U01 HL65520, U01 HL41642, U01 HL41652, U01 HL41654, and U01 HL65521. The opinions expressed in this paper are those of the author(s) and do not necessarily reflect the views of the Indian Health Service.

The PAGE coordinating center (U01HG004801-01) provides assistance with study design, phenotype harmonization, SNP selection and annotation, data cleaning, data management, integration and dissemination, and general study coordination. Genotype calling, genotype quality control, and statistical analyses are also performed by the coordinating center for some PAGE studies. The National Institute of Mental Health also contributes to the support for the coordinating center.

References

1. Liao D, Myers R, Hunt S, Shahar E, Paton C, Burke G, et al. Familial history of stroke and stroke risk. The Family Heart Study. Stroke. 1997; 28:1908–1912. [PubMed: 9341694]
2. Bak S, Gaist D, Sindrup SH, Skytthe A, Christensen K. Genetic liability in stroke: a long-term follow-up study of Danish twins. Stroke. 2002; 33:769–774. [PubMed: 11872902]
3. Lloyd-Jones D, Adams R, Carnethon M, De Simone G, Ferguson TB, Flegal K, et al. Heart disease and stroke statistics--2009 update: a report from the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. Circulation. 2009; 119:e21–181. [PubMed: 19075105]
4. Lloyd-Jones D, Adams RJ, Brown TM, Carnethon M, Dai S, De Simone G, et al. Heart disease and stroke statistics--2010 update: a report from the American Heart Association. Circulation. 2010; 121:e46–e215. [PubMed: 20019324]
5. Saeed M. Editorial comment--Unraveling the pagodian architecture of stroke as a complex disorder. Stroke. 2004; 35:824–825. [PubMed: 15044810]
6. Kleindorfer D, Broderick J, Khoury J, Flaherty M, Woo D, Alwell K, et al. The unchanging incidence and case-fatality of stroke in the 1990s: a population-based study. Stroke. 2006; 37:2473–2478. [PubMed: 16946146]
7. Duvall WL, Vorchheimer DA. Multi-bed vascular disease and atherothrombosis: scope of the problem. J Thromb Thrombolysis. 2004; 17:51–61. [PubMed: 15277788]
8. Matise TC, Ambite JL, Buyske S, Carlson CS, Cole SA, Crawford DC, et al. The next PAGE in understanding complex traits: design for the analysis of Population Architecture Using Genetics and Epidemiology (PAGE) study. Am J Epidemiol. 2011; 174:849–859. [PubMed: 21836165]

9. The ARIC Investigators. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. Am J Epidemiol. 1989; 129:687–702. [PubMed: 2646917]
10. Ives DG, Fitzpatrick AL, Bild DE, Psaty BM, Kuller LH, Crowley PM, et al. Surveillance and ascertainment of cardiovascular events. The Cardiovascular Health Study. Ann Epidemiol. 1995; 5:278–285. [PubMed: 8520709]
11. Lee ET, Welty TK, Fabsitz R, Cowan LD, Le NA, Oopik AJ, et al. The Strong Heart Study. A study of cardiovascular disease in American Indians: design and methods. Am J Epidemiol. 1990; 132:1141–1155. [PubMed: 2260546]
12. The 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. [PubMed: 9492970]
13. Anderson GL, Manson J, Wallace R, Lund B, Hall D, Davis S, et al. Implementation of the Women's Health Initiative study design. Ann Epidemiol. 2003; 13:S5–17. [PubMed: 14575938]
14. 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. [PubMed: 10187871]
15. Fried LP, Borhani NO, Enright P, Furberg CD, Gardin JM, Kronmal RA, et al. The Cardiovascular Health Study: design and rationale. Ann Epidemiol. 1991; 1:263–276. [PubMed: 1669507]
16. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc Natl Acad Sci U S A. 2009; 106:9362–9367. [PubMed: 19474294]
17. Casas JP, Hingorani AD, Bautista LE, Sharma P. Meta-analysis of genetic studies in ischemic stroke: thirty-two genes involving approximately 18,000 cases and 58,000 controls. Arch Neurol. 2004; 61:1652–1661. [PubMed: 15534175]
18. Gretarsdottir S, Thorleifsson G, Manolescu A, Styrkarsdottir U, Helgadottir A, Gschwendtner A, et al. Risk variants for atrial fibrillation on chromosome 4q25 associate with ischemic stroke. Ann Neurol. 2008; 64:402–409. [PubMed: 18991354]
19. Matarin M, Brown WM, Scholz S, Simon-Sanchez J, Fung HC, Hernandez D, et al. A genome-wide genotyping study in patients with ischaemic stroke: initial analysis and data release. Lancet Neurol. 2007; 6:414–420. [PubMed: 17434096]
20. Luke MM, O'Meara ES, Rowland CM, Shiffman D, Bare LA, Arellano AR, et al. Gene variants associated with ischemic stroke: The Cardiovascular Health Study. Stroke. 2009; 40:363–368. [PubMed: 19023099]
21. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010; 26:2190–2191. [PubMed: 20616382]
22. Huedo-Medina TB, Sanchez-Meca J, Marin-Martinez F, Botella J. Assessing heterogeneity in meta-analysis: Q statistic or I^2 index? Psychol Methods. 2006; 11:193–206. [PubMed: 16784338]
23. Mailman MD, Feolo M, Jin Y, Kimura M, Tryka K, Bagoutdinov R, et al. The NCBI dbGaP database of genotypes and phenotypes. Nat Genet. 2007; 39:1181–1186. [PubMed: 17898773]
24. Gauderman, W.; Morrison, J. [Accessed November 23, 2010] QUANTO 1.1: A computer program for power and sample size calculations for genetic-epidemiology studies. 2006. http://hydra.usc.edu/gxe
25. Zollner S, Pritchard JK. Overcoming the winner's curse: estimating penetrance parameters from case-control data. Am J Hum Genet. 2007; 80:605–615. [PubMed: 17357068]
26. Zhong H, Prentice RL. Bias-reduced estimators and confidence intervals for odds ratios in genome-wide association studies. Biostatistics. 2008; 9:621–634. [PubMed: 18310059]
27. Kolominsky-Rabas PL, Weber M, Gefeller O, Neundoerfer B, Heuschmann PU. Epidemiology of ischemic stroke subtypes according to TOAST criteria: incidence, recurrence, and long-term survival in ischemic stroke subtypes: a population-based study. Stroke. 2001; 32:2735–2740. [PubMed: 11739965]
28. Meschia JF. New information on the genetics of stroke. Curr Neurol Neurosci Rep. 2011; 11:35–41. [PubMed: 21058051]
29. Markus HS. Unravelling the genetics of ischaemic stroke. PLoS Med. 2010; 7:e1000225. [PubMed: 20209141]

30. Lanktree MB, Dichgans M, Hegele RA. Advances in genomic analysis of stroke: what have we learned and where are we headed? Stroke. 2010; 41:825–832. [PubMed: 20167918]

31. McPherson R, Pertsemlidis A, Kavaslar N, Stewart A, Roberts R, Cox DR, et al. A common allele on chromosome 9 associated with coronary heart disease. Science. 2007; 316:1488–1491. [PubMed: 17478681]

32. Helgadottir A, Thorleifsson G, Manolescu A, Gretarsdottir S, Blondal T, Jonasdottir A, et al. A common variant on chromosome 9p21 affects the risk of myocardial infarction. Science. 2007; 316:1491–1493. [PubMed: 17478679]

33. Anderson CD, Biffi A, Rost NS, Cortellini L, Furie KL, Rosand J. Chromosome 9p21 in ischemic stroke: population structure and meta-analysis. Stroke. 2010; 41:1123–1131. [PubMed: 20395606]

34. Ye S, Willeit J, Kronenberg F, Xu Q, Kiechl S. Association of genetic variation on chromosome 9p21 with susceptibility and progression of atherosclerosis: a population-based, prospective study. J Am Coll Cardiol. 2008; 52:378–384. [PubMed: 18652946]

35. Dumitrescu L, Carty CL, Taylor K, Schumacher FR, Hindorff LA, Ambite JL, et al. Genetic determinants of lipid traits in diverse populations from the Population Architecture using Genomics and Epidemiology (PAGE) study. PLoS Genet. 2011; 7:e1002138. [PubMed: 21738485]

36. Thompson JR, Attia J, Minelli C. The meta-analysis of genome-wide association studies. Brief Bioinform. 2011; 12:259–269. [PubMed: 21546449]

37. Manolio TA. Genomewide association studies and assessment of the risk of disease. N Engl J Med. 2010; 363:166–176. [PubMed: 20647212]

Recent genome-wide association studies (GWAS) have identified a number of genetic variants associated with stroke and cardiovascular disease (CVD). However, data regarding GWAS variant replication in independent samples and across different ancestry/ethnicity groups are lacking but are important for prioritizing genetic variants for translational research and for furthering our understanding of population differences in complex diseases such as stroke. We sought to replicate previously identified single nucleotide polymorphisms (SNPs) associated with ischemic stroke in GWAS and meta-analyses using a large, well-characterized, multi-ethnic population. In addition, we tested whether SNPs previously associated with CVD risk factors were also associated with ischemic stroke. In spite of reasonable power, we did not replicate several of the previous stroke findings in European descent individuals, but did identify associations in African Americans for two SNPs in the *AIM1* and *HPS1* genes. In exploratory analyses investigating CVD, we identified additional CVD risk factor (lipids and body mass index) SNPs which may be associated with stroke in diverse US ancestral groups, including African Americans and American Indians who are at high risk of stroke. Our findings highlight the importance of replication and consideration of power in genetic studies and support the investigation of non-European descent populations for identifying genetic factors associated with complex disease.

Table 1

Summary of studies and population characteristics at baseline

Study	WHI		ARIC		CHS		SHS
Design	Observational Study and Controlled Clinical Trials		Observational Study		Observational Study		Observational Study
Focus	Chronic disease risk factors and prevention in postmenopausal women		Atherosclerosis, atherosclerotic diseases, and CVD risk factors		CVD risk factors in older adults		CVD risk factors in American Indians
Recruitment	1993-94		1987-89		1989-90, 1992-93		1989-91
Stroke Ascertainment	through 2009		through 2007		through 2007		through 2006
Population	**EA**	**AA**	**EA**	**AA**	**EA**	**AA**	**AI**
N with Genotype Data	11,153	4173	10,804	4031	4319	766	3466
Female, N(%)	11,153 (100)	4173 (100)	5710 (52.8)	2509 (62.2)	2482 (57.5)	489 (63.8)	2070 (59.7)
Age in years, mean ± SD	67 ± 7	62 ± 7	54 ± 6	54 ± 6	73 ± 6	73 ± 6	56 ± 8
BMI in kg/m^2, mean ± SD	28.9 ± 6.7	33.1 ± 7.7	27.0 ± 4.9	29.6 ± 6.1	26.4 ± 4.5	28.6 ± 5.6	30.8±6.4
Type 2 Diabetes *, N (%)	380 (2.9)	216 (5.2)	723 (6.7)	669 (17.0)	321 (7.5)	54 (7.3)	1517 (44.6)
Hypertension, N (%)	5226 (40.0)	1803 (43.0)	2320 (21.6)	1926 (48.0)	2370 (54.9)	553 (72.3)	1345 (38.9)
History of MI *, N (%)	62 (0.5)	21 (0.5)	425 (4.0)	127 (3.2)	396 (9.2)	55 (7.2)	82 (2.4)
Lipid-lowering Medication Use, N (%)	1276 (9.8)	313 (7.5)	353 (3.3)	53 (1.3)	223 (5.2)	50 (6.6)	17 (0.5)
Incident Ischemic Stroke Events, N	2170	220	443	336	626	99	163
Median Follow-up Time in years	10.9	11.0	18.9	18.5	11.8	10.4	16.1

* prevalent disease was an exclusion criterion for a large subset of the WHI sample

Table 2

Meta-analysis results for associations between incident ischemic stroke and SNPs previously associated with ischemic stroke*

SNP	Ref	Locus†	Gene (Location)	%Power in EA‡	%Power in AA‡	Alleles§	European Americans				African Americans			
							N	CAF	HR (95% CI)	p-value	N	CAF	HR (95% CI)	p-value
rs1801133	meta-analysis; recessive[17]	1p36.22	*MTHFR (exon)*	30	6	**T**/C	26,463	0.35	1.05 (0.97, 1.12)	0.208	7564	0.12	1.17 (0.97, 1.41)	0.099
rs6025	meta-analysis; dominant[17]	1q24.2	*F5 (exon)*	26	14	A/**G**	23,138	0.03	1.08 (0.80, 1.45)	0.613	6971	0.01	0.32 (0.04, 2.34)	0.260
rs2200733	GWAS[18]	4q25	*PITX2 (intergenic)*	>99 (>99)	>99 (93)	**T**/C	26,518	0.13	1.07 (0.96, 1.19)	0.201	7606	0.21	0.98 (0.84, 1.15)	0.849
rs783396	GWAS[19]	6q21	*AIM1 (exon)*	>99 (22)	82 (8)	**A**/C	25,341	0.07	0.95 (0.82, 1.10)	0.462	7281	0.06	1.32 (1.03, 1.69)	**0.027**
rs10486776	GWAS[19]	7p21.2	*MEOX2 (intergenic)*	>99 (>99)	>99 (23)	**A**/G	15,946	0.07	0.96 (0.80, 1.16)	0.681	4721	0.02	1.17 (0.52, 2.63)	0.706
rs1804689‖	candidate gene[20]	10q24.2	*HPS1 (5´UTR)*	79	16	**A**/C	23,110	0.31	1.04 (0.94, 1.15)	0.425	3932	0.14	0.76 (0.59, 0.99)	**0.042**
rs1799963	meta-analysis; dominant[17]	11p11.2	*F2 (3´UTR)*	31	5	**A**/G	23,228	0.01	0.90 (0.67, 1.20)	0.482	7112	0.002	1.04 (0.31, 3.41)	0.954
rs9536591	GWAS[19]	13q14.3	*OLFM4 (intergenic)*	>99 (49)	>99 (14)	C/**A**	25,326	0.46	0.96 (0.89, 1.04)	0.298	7284	0.43	0.99 (0.86, 1.14)	0.863
rs7506045	GWAS[19]	18p11.21	*IMPA2 (intron)*	>99 (>99)	>99 (>99)	**T**/C	25,458	0.07	0.95 (0.83, 1.09)	0.508	7305	0.12	1.15 (0.94, 1.4)	0.188

Ref=SNP reference and type of study, and genetic model if not additive; N=total number of participants included in analysis; CAF= coded allele frequency; HR=hazard ratio; 95%CI=95% confidence interval

*Meta-analysis results indicate the per allele change in risk of ischemic stroke, using the coded allele. Genotype data for these SNPs were not available in the American Indians.

†Locus information based on genome build GRCh37/hg19

‡Percent power estimates shown using reported GWAS estimate and correcting for winner's curse (in parentheses).

§Alleles column lists the coded allele/other allele. Risk alleles identified in the original study are bolded.

‖Meta-analysis included ARIC and WHI, but not CHS because the stroke association for this SNP was originally reported in the CHS population.

Table 3

Meta-analysis results for SNPs previously associated with CVD risk factors that were significant in ≥ 1 PAGE race/ethnic group

SNP	Locus*	Trait	Alleles†	European Americans				African Americans				American Indians			
				N	CAF	HR (95% CI)‡	p-value	N	CAF	HR (95% CI)‡	p-value	N	CAF	HR (95% CI)‡	p-value
Stroke Risk Factor SNPs Significant in ≥ 1 race/ethnic group															
rs6544713	2p21	LDL	T/C	23,201	0.31	1.03 (0.94, 1.14)	0.494	7880	0.17	1.21 (1.03, 1.42)	**0.022**	3464	0.09	0.80 (0.47, 1.36)	0.413
rs754523	2p24.1	LDL	T/C	15,938	0.69	0.93 (0.84, 1.03)	0.151	4722	0.79	0.87 (0.70, 1.08)	0.205	3452	0.69	0.70 (0.55, 0.88)	**0.003**
rs2954029	8q24.13	TG	T/A	23,104	0.46	1.11 (1.00, 1.23)	**0.049**	6970	0.34	1.00 (0.87, 1.16)	0.958	n/a			
rs2156552	18q21.1	HDL	A/T	16,921	0.84	0.85 (0.76, 0.96)	**0.006**	6783	0.96	1.24 (0.84, 1.48)	0.274	3460	0.95	0.55 (0.22, 1.37)	0.200
rs11084753	19q13.11	BMI	A/G	16,927	0.33	0.95 (0.87, 1.04)	0.247	6774	0.36	1.18 (1.02, 1.35)	**0.024**	3449	0.30	1.14 (0.90, 1.43)	0.276
rs1800961	20q13.12	HDL	T/C	n/a				2497	0.01	4.00 (1.81, 8.83)	**0.0006**	3461	0.03	1.52 (0.86, 2.68)	0.146

N=total number of participants included in analysis; CAF= coded allele frequency; HR=hazard ratio; 95%CI=95% confidence interval

*Locus information based on genome build GRCh37/hg19

†Alleles column lists the coded allele/other allele.

‡Meta-analysis results indicate the per allele change ischemic stroke hazard, using the coded allele.