Alpha globin gene copy number and incident ischemic stroke risk among Black Americans

A. Parker Ruhl[1,2], Neal Jeffries[3], Yu Yang[4], Steven D. Brooks[1], Rakhi P. Naik[5,] Lydia H. Pecker[5], Bryan T. Mott[6], Cheryl A. Winkler[7], Nicole D. Armstrong[8], Neil A. Zakai[9,10], Orlando M. Gutierrez[8,11], Suzanne E. Judd[8], Virginia J. Howard[8,11], George Howard[8,11], Marguerite R. Irvin[8], Mary Cushman[9,10], Hans C. Ackerman[1]

[1]Laboratory of Malaria and Vector Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland; [2]Pulmonary Branch, National Heart, Lung, and Blood Institute, Bethesda, Maryland; [3]Office of Biostatistics Research, National Heart, Lung, and Blood Institute, Bethesda, Maryland; [4]Division of Blood Diseases and Resources, National Heart, Lung, and Blood Institute, Rockville, Maryland; [5]Division of Hematology, Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland; [6]Wake Forest University School of Medicine, Winston-Salem, North Carolina: [7]Basic Research Laboratory, National Cancer Institute, Frederick National Laboratory for Cancer Research, Frederick, Maryland; [8]Department of Epidemiology, University of Alabama at Birmingham, Birmingham, Alabama; [9]Department of Medicine, Larner College of Medicine at the University of Vermont, Burlington, Vermont; [10]Department of Pathology & Laboratory Medicine, Larner College of Medicine at the University of Vermont, Burlington, Vermont; [11]Department of Medicine, University of Alabama at Birmingham, Birmingham, Alabama,

Correspondence: A. Parker Ruhl, MD, 10 Center Dr., B3-4207, Bethesda, MD, 20892. parker.ruhl@nih.gov

Short title: Alpha globin and ischemic stroke

Keywords: alpha globin, alpha thalassemia, ischemic stroke, hemoglobin, nitric oxide

Word count: 1980

Tables: 4

Figures: 1

Author contribution statement:
APR, SB, and HCA conceived and designed the study and drafted the first version of the manuscript. YY, BTM, and CAW performed sample analyses. NJ designed and implemented the statistical analyses. All authors contributed to data interpretation and revised and approved the manuscript.

ABSTRACT

Introduction: People with African ancestry have greater stroke risk and greater heritability of stroke risk than people of other ancestries. Given the importance of nitric oxide (NO) in stroke, and recent evidence that alpha globin restricts nitric oxide release from vascular endothelial cells, we hypothesized that alpha globin gene (*HBA)* deletion would be associated with reduced risk of incident ischemic stroke.

Methods: We evaluated 8,947 participants self-reporting African ancestry in the national, prospective Reasons for Geographic And Racial Differences in Stroke (REGARDS) cohort. Incident ischemic stroke was defined as non-hemorrhagic stroke with focal neurological deficit lasting ≥ 24 hours confirmed by the medical record or focal or non-focal neurological deficit with positive imaging confirmed with medical records. Genomic DNA was analyzed using droplet digital PCR to determine *HBA* copy number. Multivariable Cox proportional hazards regression was used to estimate the hazard ratio (HR) of *HBA* copy number on time to first ischemic stroke.

Results: Four-hundred seventy-nine (5.3%) participants had an incident ischemic stroke over a median (IQR) of 11.0 (5.7, 14.0) years' follow-up. *HBA* copy number ranged from 2 to 6: 368 (4%) -α/-α, 2,480 (28%) -α/αα, 6,014 (67%) αα/αα, 83 (1%) ααα/αα and 2 (<1%) ααα/ααα. The adjusted HR of ischemic stroke with *HBA* copy number was 1.04; 95%CI 0.89, 1.21; p = 0.66.

Conclusions: Although a reduction in *HBA* copy number is expected to increase endothelial nitric oxide signaling in the human vascular endothelium, *HBA* copy number was not associated with incident ischemic stroke in this large cohort of Black Americans.

INTRODUCTION

People with African ancestry have greater stroke risk and greater heritability of stroke risk than people of other ancestries.[1,2] While some risk is explained by genetic variants shared between European and African populations, less is known about African-specific variants.[3] Alpha globin (*HBA*), which regulates endothelial nitric oxide (NO) signaling in resistance arteries, varies in gene copy number among people of African and Asian descent.[4,5] Genetic deletion of *HBA* is associated with protection from kidney disease, possibly through increased vascular NO signaling.[6]

NO plays an important role as a signaling molecule which regulates cerebrovascular blood flow both at rest and during injury.[7] The regulation of cerebrovascular blood flow via NO occurs via two main mechanisms: 1) autoregulation to keep a consistent blood flow and 2) neurovascular coupling in which increased neuronal activity is influenced by the local regulation of cerebral blood flow. When there is reduced cerebral blood flow, for example during either an ischemic stroke or a hemorrhagic stroke, depletion of NO may exacerbate neuronal injury. However, the role of NO in the cerebral vasculature is complex and an overabundance of NO leading to significant vasodilation and lower cerebral perfusion pressure could potentially contribute to cerebral injury during stroke event. In addition, NO may impact the risk of developing vasospasm in the setting of cerebral hemorrhage.[8] Given the importance of NO in stroke,[7,9] and evidence that the alpha subunit of hemoglobin restricts nitric oxide release from vascular endothelial cells, we examined the association between *HBA* copy number and incident ischemic stroke in the Reasons for Geographic And Racial Differences in Stroke (REGARDS) cohort.[10] We hypothesized that lower *HBA* copy number would be associated with lower risk of incident ischemic stroke.

METHODS

Study Design

REGARDS is a longitudinal cohort study designed to determine the reasons for racial disparities in stroke and cognitive decline in Black and White Americans aged ≥ 45 years.[10] REGARDS enrolled 30,239 participants from the 48 continental United States from 2003 to 2007. All self-reported Black participants consenting to genetic research were included in this study (Figure 1). All participants provided oral and written informed consent. The REGARDS study was approved by the Institutional Review Boards of participating centers. This study followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting guideline. The analytic plan was prespecified and approved by the REGARDS Publications Committee. Among 9,999 Black REGARDS participants with *HBA* and *HBB* genotyping available, 8,947 participants had no self-reported history of stroke at baseline and complete follow-up data (Figure 1).

Incident ischemic stroke was defined as non-hemorrhagic stroke with a focal neurological deficit lasting ≥ 24 hours confirmed by the medical record or a focal or non-focal neurological deficit with positive imaging confirmed with medical records in participants without a prior history of stroke. Genomic DNA was analyzed using droplet digital PCR[6] to determine *HBA* copy number. The covariates age, sex, race, health insurance (yes or no), highest education level obtained (less than high school, high school, some college, college or more), annual income (≤ $20K, $20-34K, $35-74K, ≥ $75K), and smoking status (categorized by never, past, or current smoker), history of hyperlipidemia, regular use of lipid lowering medication, and regular use of aspirin were self-reported. Self-reported use of lipid lowering medication was restricted to those also reporting hyperlipidemia. Region was defined as previously described in three geographic areas: stroke belt buckle, stroke belt, and stroke nonbelt.[11] Atrial fibrillation was defined by self-report of a physician diagnosis or ECG evidence. The presence of left ventricular hypertrophy (LVH) was defined by 12-lead ECG. Hypertension was defined as systolic blood pressure >=130, diastolic blood pressure >=80, or self-reported current medication use to control blood pressure. Chronic kidney disease (CKD) was defined by the

2021 CKD-Epi creatinine-cysteine equation and estimated glomerular filtration rate less than 60mL/min/1.73m^2, including those with end-stage kidney disease, and/or urine albumin to creatinine ratio $\geq$ 30 mg/g measured on urine collected during the baseline in-home examination. Fasting glucose levels $\geq$ 126 mg/dL, random glucose $\geq$ 200 mg/dL, or self-reported use of glucose-lowering medication was used to define diabetes mellitus. Hemoglobin, mean corpuscular volume (MCV), mean corpuscular hemoglobin (MCH), mean corpuscular hemoglobin concentration (MCHC), and red-cell distribution width-coefficient of variation (RDW-CV) values were measured or calculated from blood collected during the in-home examination. The first 10 principal components of ancestry were calculated from Infinium Expanded Multi-Ethnic Genotyping Array data on 7,032 (79%) participants. *HBA* regulatory single nucleotide polymorphisms (SNPs) (rs11248850, rs11865131, rs7203560)[12,13] and a NOS3 SNP (rs1549758)[14] were determined from these array data. The *HBA* regulatory SNPs rs11248850 and rs11865131 were in high linkage disequilibrium and were concordant in all but five participants. Therefore, we evaluated only rs11248850 and not rs11865131 in the models. The Framingham Stroke Risk Score (10-Year probability of stroke risk percentage)[15] and Atherosclerosis Risk in Communities Study (ARIC) Stroke Risk Score, 10-year probability of ischemic stroke risk (percentage) were calculated among those who self-reported at baseline never having had a stroke.[16]

Statistical methods

Multivariable Cox proportional hazards regression modeling was used to estimate the hazard ratio of *HBA* copy number on time to first ischemic stroke. Covariates included age, sex, region, insurance status, education level, income, hypertension, atrial fibrillation, left ventricular hypertrophy, smoking status, and diabetes mellitus. Pre-specified tests of interaction between *HBA* copy number and age, sex, and sickle cell trait were performed on fully adjusted models. Pre-specified sensitivity analyses were performed by adding each of the following to the model:

hemoglobin, CKD, both hemoglobin and CKD, regular aspirin use, regular statin use, the first ten principal components of ancestry, and putative *HBA* regulatory single nucleotide polymorphisms rs11248850 and rs7203560. A post-hoc sensitivity analysis for the association of the rs1549758 NOS3 (nitric oxide synthase 3 or endothelial nitric oxide synthase) SNP with incident ischemic stroke was performed.

For the ischemic stroke time-to-event analysis, time to event was defined as the number of years between the initial in-home interview date and date D where D was the minimum of D1 and D2 which were defined as follows. D1 was the last follow-up date provided by REGARDS as the last time the participant was contacted for status. The outcome associated with this date was either “No Event” or “Death”. D2 was the date when the individual had an ischemic stroke. Individuals with a D2 date prior to their initial in-home interview were excluded from consideration in this paper. Individuals with a D2 date after interview but prior to D1 time were recorded with an ischemic stroke event and D = D2. Individuals with D2 after D1 had stroke onset after their last REGARDS follow-up and they were recorded with “No Event” and censored at date D = D1. This decision to censor stroke events occurring after the end of REGARDS follow-up avoids bias associated with having extended follow-up only for one type of subgroup - those having a stroke event.

Missing data for the primary outcome, secondary outcomes, and explanatory variables were typically rare (< 0.5%) with some exceptions, e.g., hemoglobin values (Table 1). Multiple imputation methods were used in the multivariable analyses. The R package “mice” Version 3.14.0 was used to create and analyze the resulting imputations.[17]

For diagnostic modeling of the model of ischemic stroke incidence using Cox proportional hazard techniques, Schoenfeld residuals were examined over follow-up time to detect violations of proportional hazards assumptions for the covariates in the analysis of ischemic outcomes. Examination of Schoenfeld residuals of the non-imputed and imputed data

sets showed no suggestion of violation of proportional hazards for the regression covariates (p = 0.98).

For incident ischemic stroke, we performed an a priori power calculation based upon the Cox proportional-hazards model. We assumed an additive model where each additional *HBA* copy increased the hazard ratio for prevalent stroke by 25%. Given n=479 incident ischemic stroke events observed among those without baseline self-reported stroke, and the distribution of alpha globin alleles, the power to reject the hypothesis of no linear trend in *HBA* copy number on the log of the hazard rate is approximately 77%. In contrast, there were only 62 incident hemorrhagic stroke events and power to reject a linear trend in *HBA* copy number on the log of the hazard rate was only 13%; therefore, analysis of hemorrhagic stroke was not performed.

RESULTS

Of the 8,947 REGARDs participants with available data (Figure 1), 479 (5.3%) participants had an incident ischemic stroke over a median (IQR) of 11.0 (5.7, 14.0) years' follow-up. Of these participants with incident ischemic stroke, 393 (82%) were diagnosed with a non-hemorrhagic stroke with a focal neurological deficit lasting ≥ 24 hours and 86 (18%) were found to have a non-focal neurological deficit with positive imaging. There were 62 participants who developed incident hemorrhagic stroke. *HBA* gene copy number ranged from 2 to 6: 368 (4%) -α/-α, 2,480 (28%) -α/αα, 6,014 (67%) αα/αα, 83 (1%) ααα/αα and 2 (<1%) ααα/ααα (Table 1). The HR of ischemic stroke with *HBA* copy number, fully adjusted for age, sex, region, insurance status, education, income, hypertension, atrial fibrillation, LVH, smoking, and diabetes, was 1.04; 95%CI 0.89, 1.21; p = 0.66; (Table 2). There were no interactions between *HBA* copy number and age, sex, or sickle cell trait (Table 3). In pre-specified sensitivity analyses, the addition of hemoglobin, chronic kidney disease, aspirin use, statin use, or the first ten principal components of ancestry did not materially change the observed associations (Table 4). In a post-hoc

sensitivity analysis we evaluated whether the putative NOS3 SNP was associated with incidence ischemic stroke in this population and no association was found (HR 1.03; 95%CI 0.84, 1.27; p = 0.79).

DISCUSSION

Although a reduction in *HBA* copy number is expected to increase endothelial nitric oxide signaling in the human vascular endothelium, *HBA* copy number was not associated with incident ischemic stroke in this large longitudinal cohort of Black Americans. This finding is consistent with a prior study demonstrating no association between *HBA* copy number and hypertension but differs from a study demonstrating an association with chronic and end-stage kidney disease in which *HBA* gene deletions are protective.[6,18] Together, these population studies suggest that the physiological roles of alpha globin may differ between the renal and cerebral vascular beds.

The strengths of this study include a robust measurement of *HBA* copy number with ddPCR with the analysis performed in a well-characterized, large, national cohort of Black Americans with clearly defined cerebrovascular outcomes followed for an extensive period. Moreover, we adjusted for key stroke risk factors including demographic, social, and biomedical variables to address the novel question of whether *HBA* copy number is associated with stroke risk. We explored the relevance of the NOS3 SNP rs1549758 and found no association between this SNP and incident ischemic stroke in a fully adjusted model. This replicates the finding of no association with stroke among participants with African ancestry in a prior GWAS study.[14] Given the absence of a significant main effect for either *HBA* CNV or the NOS3 SNP on incident ischemic stroke in this cohort, we did not pursue a model with both main effects nor did we test for interaction.

Our study has limitations, including that the population is limited to Black Americans and the associations of *HBA* copy number and ischemic stroke risk may vary in other populations

with *HBA* copy number variation and outside of the U.S. We did not have sufficient power to evaluate other stroke subtypes such as hemorrhagic stroke, though increases in vascular NO signaling could affect bleeding risk through its effects on platelets. An evaluation of the association of *HBA* copy number and hemorrhagic stroke risk could be evaluated in other populations with a higher frequency of *HBA* copy number variation. For example, in Sub-Saharan Africa, there is a slightly higher frequency of hemorrhagic stroke than reported than in the U.S.[19]

CONCLUSION

In a large longitudinal cohort of Black American adults, we found that *HBA* copy number was not associated with incident ischemic stroke.

REFERENCES

1. Prapiadou S, Demel SL, Hyacinth HI. Genetic and Genomic Epidemiology of Stroke in People of African Ancestry. *Genes*. 2021;12:1825.

2. Howard G, Cushman M, Kissela BM, Kleindorfer DO, McClure LA, Safford MM, Rhodes JD, Soliman EZ, Moy CS, Judd SE, Howard VJ. Traditional risk factors as the underlying cause of racial disparities in stroke: lessons from the half-full (empty?) glass. *Stroke*. 2011;42.

3. 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.

4. Straub AC, Lohman AW, Billaud M, Johnstone SR, Dwyer ST, Lee MY, Bortz PS, Best AK, Columbus L, Gaston B, Isakson BE. Endothelial cell expression of haemoglobin α regulates nitric oxide signalling. *Nature*. 2012;491:473–477.

5. Dozy AM, Kan YW, Embury SH, Mentzer WC, Wang WC, Lubin B, Davis JR, Koenig HM. alpha-Globin gene organisation in blacks precludes the severe form of alpha-thalassaemia. *Nature*. 1979;280:605–607.

6. Ruhl AP, Jeffries N, Yang Y, Naik RP, Patki A, Pecker LH, Mott BT, Zakai NA, Winkler CA, Kopp JB, Lange LA, Irvin MR, Gutierrez OM, Cushman M, Ackerman HC. Alpha Globin Gene Copy Number Is Associated with Prevalent Chronic Kidney Disease and Incident End-Stage Kidney Disease among Black Americans. *J Am Soc Nephrol*. 2022;33:213–224.

7. Garry PS, Ezra M, Rowland MJ, Westbrook J, Pattinson KTS. The role of the nitric oxide pathway in brain injury and its treatment — From bench to bedside. *Exp Neurol*. 2015;263:235–243.

8. Siuta M, Zuckerman SL, Mocco J. Nitric Oxide in Cerebral Vasospasm: Theories, Measurement, and Treatment. *Neurol Res Int*. 2013;2013:972417.

9. Woodhouse L, Scutt P, Krishnan K, Berge E, Gommans J, Ntaios G, Wardlaw J, Sprigg N, Bath PM. Effect of Hyperacute Administration (Within 6 Hours) of Transdermal Glyceryl Trinitrate, a Nitric Oxide Donor, on Outcome After Stroke. *Stroke*. 2015;46:3194–3201.

10. 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.

11. Howard G, Howard VJ. Twenty Years of Progress Toward Understanding the Stroke Belt. *Stroke*. 2020;51:742–750.

12. Raffield LM, Ulirsch JC, Naik RP, Lessard S, Handsaker RE, Jain D, Kang HM, Pankratz N, Auer PL, Bao EL, Smith JD, Lange LA, Lange EM, Li Y, Thornton TA, Young BA, Abecasis GR, Laurie CC, Nickerson DA, McCarroll SA, Correa A, Wilson JG, Consortium NT-O for PM (TOPMed), Hemostasis H&, Diabetes, Groups and SVTopmW, Lettre G, Sankaran VG, Reiner AP. Common α-globin variants modify hematologic and other clinical phenotypes in sickle cell trait and disease. *PLOS Genet*. 2018;14:e1007293.

13. Milton JN, Rooks H, Drasar E, McCabe EL, Baldwin CT, Melista E, Gordeuk VR, Nouraie M, Kato GR, Kato GJ, Minniti C, Taylor J, Campbell A, Luchtman-Jones L, Rana S, Castro O, Zhang Y, Thein SL, Sebastiani P, Gladwin MT, Walk-PHAAST Investigators, Steinberg MH. Genetic determinants of haemolysis in sickle cell anaemia. *Br J Haematol*. 2013;161:270–278.

14. Mishra A, Malik R, Hachiya T, Jürgenson T, Namba S, Posner DC, Kamanu FK, Koido M, Le Grand Q, Shi M, He Y, Georgakis MK, Caro I, Krebs K, Liaw Y-C, Vaura FC, Lin K, Winsvold BS, Srinivasasainagendra V, Parodi L, Bae H-J, Chauhan G, Chong MR, Tomppo L, Akinyemi R, Roshchupkin GV, Habib N, Jee YH, Thomassen JQ, Abedi V, Cárcel-Márquez J, Nygaard M, Leonard HL, Yang C, Yonova-Doing E, Knol MJ, Lewis AJ, Judy RL, Ago T, Amouyel P, Armstrong ND, Bakker MK, Bartz TM, Bennett DA, Bis JC, Bordes C, Børte S, Cain A, Ridker PM, Cho K, Chen Z, Cruchaga C, Cole JW, de Jager PL, de Cid R, Endres M, Ferreira LE, Geerlings MI, Gasca NC, Gudnason V, Hata J, He J, Heath AK, Ho Y-L, Havulinna AS, Hopewell JC, Hyacinth HI, Inouye M, Jacob MA, Jeon CE, Jern C, Kamouchi M, Keene KL, Kitazono T, Kittner SJ, Konuma T, Kumar A, Lacaze P, Launer LJ, Lee K-J, Lepik K, Li J, Li L, Manichaikul A, Markus HS, Marston NA, Meitinger T, Mitchell BD, Montellano FA, Morisaki T, Mosley TH, Nalls MA, Nordestgaard BG, O'Donnell MJ, Okada Y, Onland-Moret NC, Ovbiagele B, Peters A, Psaty BM, Rich SS, Rosand J, Sabatine MS, Sacco RL, Saleheen D, Sandset EC, Salomaa V, Sargurupremraj M, Sasaki M, Satizabal CL, Schmidt CO, Shimizu A, Smith NL, Sloane KL, Sutoh Y, Sun YV, Tanno K, Tiedt S, Tatlisumak T, Torres-Aguila NP, Tiwari HK, Trégouët D-A, Trompet S, Tuladhar AM, Tybjærg-Hansen A, van Vugt M, Vibo R, Verma SS, Wiggins KL, Wennberg P, Woo D, Wilson PWF, Xu H, Yang Q, Yoon K, Millwood IY, Gieger C, Ninomiya T, Grabe HJ, Jukema JW, Rissanen IL, Strbian D, Kim YJ, Chen P-H, Mayerhofer E, Howson JMM, Irvin MR, Adams H, Wassertheil-Smoller S, Christensen K, Ikram MA, Rundek T, Worrall BB, Lathrop GM, Riaz M, Simonsick EM, Kõrv J, França PHC, Zand R, Prasad K, Frikke-Schmidt R, de Leeuw F-E, Liman T, Haeusler KG, Ruigrok YM, Heuschmann PU, Longstreth WT, Jung KJ, Bastarache L, Paré G, Damrauer SM, Chasman DI, Rotter JI, Anderson CD, Zwart J-A, Niiranen TJ, Fornage M, Liaw Y-P, Seshadri S, Fernández-Cadenas I, Walters RG, Ruff CT, Owolabi MO, Huffman JE, Milani L, Kamatani Y, Dichgans M, Debette S. Stroke genetics informs drug discovery and risk prediction across ancestries. *Nature*. 2022;611:115–123.

15. McClure LA, Kleindorfer DO, Kissela BM, Cushman M, Soliman EZ, Howard G. Assessing the Performance of the Framingham Stroke Risk Score in the Reasons for Geographic and Racial Differences in Stroke Cohort. *Stroke*. 2014;45:1716–1720.

16. Chambless LE, Heiss G, Shahar E, Earp MJ, Toole J. Prediction of ischemic stroke risk in the Atherosclerosis Risk in Communities Study. *Am J Epidemiol*. 2004;160:259–269.

17. Van Buuren S, Goorthius-Oudshoom K. mice: Multivariate Imputation by Chained Equations in R. *J Stat Softw*. 2011;45:1–67.

18. Ruhl AP, Jeffries N, Yang Y, Gutierrez OM, Muntner P, Naik RP, Pecker LH, Mott BT, Zakai NA, Safford MM, Lange LA, Winkler CA, Irvin MR, Cushman M, Ackerman HC. Alpha globin gene copy number and hypertension risk among Black Americans. *PLOS ONE*. 2022;17:e0271031.

19. Akinyemi RO, Ovbiagele B, Adeniji OA, Sarfo FS, Abd-Allah F, Adoukonou T, Ogah OS, Naidoo P, Damasceno A, Walker RW, Ogunniyi A, Kalaria RN, Owolabi MO. Stroke in Africa: profile, progress, prospects and priorities. *Nat Rev Neurol*. 2021;17:634–656.

DATA AVAILABILITY

Data are available from the University of Alabama at Birmingham (regardsadmin@uab.edu).

The genetic data is available in dbGaP with study accession number phs002719.v1.p1.

FUNDING

This is an ancillary study supported by cooperative agreement U01 NS041588 (M.C., V.J.H., G.H.) co-funded by the National Institute of Neurological Disorders and Stroke (NINDS) and the National Institute on Aging (NIA), National Institutes of Health, Department of Health and Human Service. This research was supported in part by the Divisions of Intramural Research, National Institute of Allergy and Infectious Diseases project AI001150 (A.P.R., H.C.A), National Heart, Lung, and Blood Institute (NHLBI) project HL006196 (A.P.R., Y.Y., S.D.B., L.H.P., H.C.A.). This work was also funded in part by the National Cancer Institute (NCI) Intramural Research Program under contract HHSN26120080001E (C.A.W), and the NHLBI grants K08HL12510 (R.P.N.) and K08HL096841 (N.A.Z.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NINDS, NIA, NIAID, NCI, or NHLBI. The content of this publication does not necessarily reflect the view or policy of the Department of Health and Human Services, nor does mention of trade names, commercial products or organizations imply endorsement by the government. The interpretation and reporting of these data are the responsibility of the author(s) and in no way should be seen as an official policy or interpretation of the U.S. government.

DISCLOSURES

Dr. Gutierrez receives funding from GlaxoSmithKline; consulting fees from Reata, AstraZeneca, and Ardelyx; funding and consulting fees from Akebia Therapeutics, Amgen. He serves on a data monitoring committee for QED Therapeutics.

Figure 1. Incident ischemic stroke and *HBA* copy number cohort study flow diagram

REGARDS = Reasons for Geographic and Racial Differences in Stroke (REGARDS) longitudinal cohort study; *HBA* = alpha globin gene; *HBB* = beta globin gene

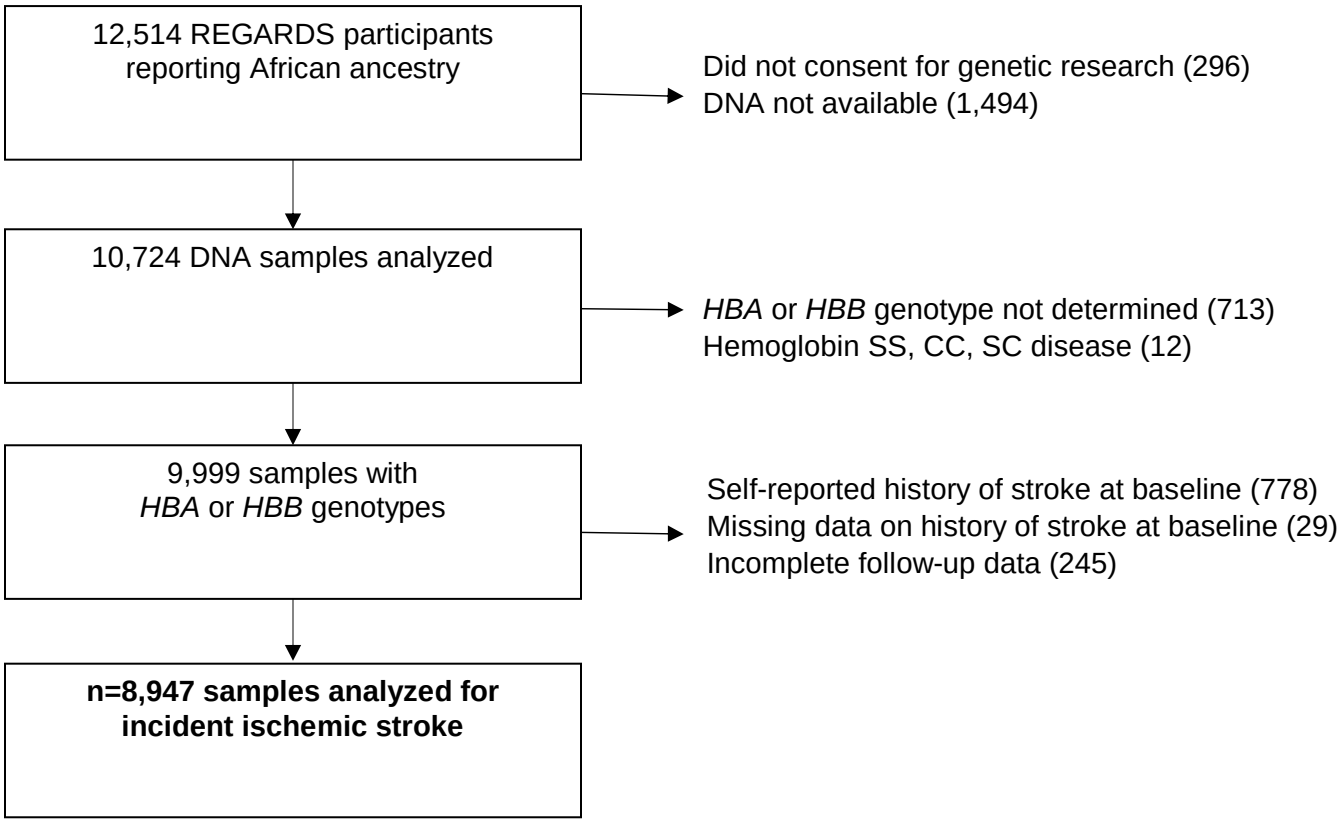

Table 1. Clinical and demographic characteristics by *HBA* copy number

		***HBA* copy number**			
	All Subjects	**2**	**3**	**4**	**≥ 5***
Subjects, N (%)	8947 (100)	368 (4)	2480 (28)	6014 (67)	85 (1)
Age, years	63.0 (57, 70)	64.0 (58, 70)	64.0 (57, 70)	63.0 (57, 70)	67.0 (61, 72)
Female sex, N (%)	5549 (62)	237 (64)	1532 (62)	3730 (62)	50 (59)
Smoking status, N (%)					
Never	4098 (46)	159 (44)	1160 (47)	2738 (46)	41 (48)
Past	3313 (37)	149 (41)	905 (37)	2227 (37)	32 (38)
Present	1493 (17)	57 (16)	402 (16)	1022 (11)	12 (14)
Region, N (%)					
Non-Belt	4430 (50)	162 (44)	1255 (51)	2959 (49)	54 (64)
Belt	2963 (33)	129 (35)	790 (32)	2026 (34)	18 (21)
Buckle	1554 (17)	77 (21)	435 (18)	1029 (17)	13 (15)
Medically insured, N (%)	8043 (90)	332 (90)	2241 (91)	5393 (90)	77 (91)
Education level, N (%)					
Less than high school	1640 (18)	63 (17)	464 (19)	1098 (18)	15 (18)
High school graduate	2462 (28)	102 (28)	6987 (28)	1624 (27)	27 (32)
Some college	2435 (27)	99 (27)	698 (28)	1615 (27)	23 (27)

College graduate or more	2404 (28)	104 (28)	607 (24)	1673 (28)	20 (24)
Income, N (%)					
≤ $20K	2306 (29)	98 (30)	660 (30)	1528 (29)	20 (26)
$20K - $34K	2341 (30)	100 (30)	674 (31)	1561 (30)	29 (38)
$35K - $74K	2392 (30)	102 (31)	674 (31)	1592 (30)	24 (31)
≥ $75K	864 (11)	30 (9)	198 (9)	632 (12)	4 (5)
Atrial fibrillation	636 (7)	23 (6)	184 (8)	422 (7)	7 (8)
Left ventricular hypertrophy	1265 (14)	51 (14)	358 (15)	845 (14)	11 (13)
Hypertension[†], N (%)	7571 (86)	315 (87)	2114 (87)	5066 (86)	76 (92)
Diabetes mellitus, N (%)	2491 (28)	95 (26)	719 (29)	1655 (28)	22 (26)
Chronic Kidney disease, N (%)	2314 (27)	76 (21)	645 (27)	1566 (27)	27 (34)
Regular aspirin use	3268 (37)	130 (35)	917 (37)	2188 (36)	33 (39)
Lipid-lowering medication use	2640 (30)	108 (30)	729 (30)	1767 (30)	36 (43)
Framingham Stroke Risk Score	7.1 (3.8, 13.5)	6.9 (3.6, 13.9)	7.1 (3.9, 13.7)	7.1 (3.7, 13.4)	8.1 (5.1, 16.5)
ARIC Stroke Risk Score	6.2 (2.8, 14.1)	6.0 (2.7, 14.6)	6.3 (2.9, 14.6)	6.2 (2.8, 13.7)	6.0 (4.1, 18.5)
***HBA* regulatory SNP rs11248850**					
G/G	4059 (59)	234 (81)	1311 (68)	2481 (54)	35 (53)

A/G	2492 (36)	53 (18)	577 (30)	1837 (40)	24 (36)
A/A	372 (5)	3 (1)	44 (2)	318 (7)	7 (11)
***HBA* regulatory SNP rs7203560**					
A/A	6005 (87)	158 (54)	1388 (72)	4394 (95)	65 (98)
A/G	882 (13)	108 (37)	535 (28)	238 (5)	1 (2)
G/G	36 (<1)	24 (8)	8 (<1)	4 (<1)	0 (0)
Sickle cell trait, N (%)	672 (8)	23 (6)	184 (7)	461 (8)	4 (5)
NOS3 SNP rs1549758					
C/C	5310 (78)	230 (81)	1478 (78)	3549 (78)	53 (82)
T/C	1392 (20)	47 (17)	389 (20)	945 (21)	11 (17)
T/T	109 (2)	7 (2)	38 (2)	63 (1)	1 (2)
Hemoglobin, g/dL	13.1 (12.2, 14.0)	12.3 (11.5, 13.2)	12.9 (12.1, 13.8)	13.3 (12.4, 14.1)	13.1 (12.2, 14.1)
MCV, fL	88.0 (84.0, 92.0)	74.0 (72.0, 77.0)	84.0 (82.0, 87.0)	90.0 (87.0, 93.0)	88.0 (86.0, 92.0)
MCH, pg	29.8 (27.9, 30.9)	23.8 (22.9, 24.8)	27.9 (26.9, 28.9)	30.3 (29.2, 31.4)	29.8 (29.1, 30.9)
MCHC, g/dL	33.4 (32.9, 33.9)	32.1 (31.6, 32.5)	33.0 (32.6, 33.5)	33.7 (33.2, 34.1)	33.7 (33.2, 33.9)
RDW-CV, %	13.9 (13.3, 14.8)	15.0 (14.4, 15.9)	14.2 (13.5, 15.1)	13.8 (13.2, 14.6)	13.6 (13.2, 14.3)

HBA = alpha globin gene; P = p value; N = number; K = thousand; RBC = red blood cell; MCV = mean cell volume; MCH = mean cell hemoglobin; MCHC = mean corpuscular hemoglobin concentration; RDW-CV = red cell distribution width coefficient of variation; ARIC = Atherosclerosis Risk in Communities Study; SNP = single nucleotide polymorphism; No. = number. NOS3 = nitric oxide synthase 3. Values are median (25th, 75th percentile) except where otherwise indicated.

*83 subjects had 5 *HBA* gene copies and 2 subjects had 6 *HBA* copies; †P values for tests of differences by *HBA* genotype generated from the chi-squared test for categorical variables and the Kruskal-Wallis non-parametric ANOVA test for continuous variables.

Missing data are as follows: medically insured (n=10, <0.01%); education (n=6, <0.01%); income (n=1,044 refused, 12%); hypertension (n=181, 2%); atrial fibrillation (n=229,3%); left ventricular hypertrophy (n=145, 2%); regular aspirin use (n=4, <0.01%); lipid-lowering medication use (n=87, 1%) kidney disease (n=359, 4%); diabetes mellitus (n=44, <0.01%); smoking status (n=43, <0.01%); Rs11248850, *HBA* regulatory SNP (n=2,023); rs7203560, *HBA* regulatory SNP (n=2,024); hemoglobin (n=2,853, 32%); MCV (n=2,858, 32%); MCH (n=2,853, 32%); MCHC (n=2,853, 32%); RDW-CV (n=2,863, 32%)' Framingham risk score (n=503, 6%); ARIC risk score (n=390, 4%). All other variables in Table 1 had no missing values.

Table 2. Association of *HBA* copy number with incident ischemic stroke – fully adjusted analyses.

	HR	95% CI	P value[†]
***HBA* copy number**	1.04	(0.89,1.21)	0.66
Age, per year	1.05	(1.03, 1.06)	<0.001
Sex			
Female (ref) [§]			
Male	0.80	(0.66, 0.96)	0.02
Region			
Non-Belt (ref)			
Belt	1.15	(0.94, 1.41)	0.18
Buckle	1.19	(0.92, 1.52)	0.19
Medically insured			
No (ref)			
Yes	0.87	(0.62, 1.21)	0.41
Education level			
< HS Grad (ref)			
HS Grad	1.13	(0.87, 1.45)	0.36
Some College	0.78	(0.59, 1.04)	0.09
≥ College Grad	0.75	(0.55, 1.02)	0.07
Income			
< $20K (ref)			
$20K - $34K	1.09	(0.85, 1.40)	0.48
$35K - $74K	1.03	(0.77, 1.37)	0.83
≥ $75K	0.72	(0.45, 1.17)	0.19
Hypertension	1.51	(1.09, 2.08)	0.01
Atrial fibrillation	1.48	(1.09, 1.99)	0.01
Left ventricular hypertrophy	1.32	(1.04, 1.67)	0.02
Smoking status			
Never (ref)			

Past	1.04	(0.84, 1.26)	0.79
Present	1.62	(1.25, 2.10)	<0.001
Diabetes mellitus			
No (ref)			
Yes	1.67	(1.39, 2.01)	<0.001

HBA = alpha globin gene; HR = hazard ratio; CI = confidence interval; ref = reference; HS = high school. Grad = graduate; K = thousand.

[†]P values were generated with a Cox proportional hazards multivariable regression employing a linear effect (i.e., additive model for risk) of *HBA* allele count on the log of the hazard ratio. All variables in the table were included in the multivariable model. Multiple imputations were performed for missing data.

Table 3. **Pre-specified tests for interaction between *HBA* genotype and age, sex, and sickle cell trait on incident ischemic stroke in fully adjusted models**

Interaction term	P value
Age**HBA*	0.61
Sex**HBA*	0.77
Sickle cell trait**HBA*	0.74

HBA= alpha globin gene; HR= hazard ratio; CI= 95% confidence interval

*P values were generated with Cox proportional hazards multivariable regression models employing a linear effect (i.e., additive model for risk) of *HBA* allele count on the log of the hazard ratio. Each model was adjusted for HBA genotype, age per year, male sex, region, medically insurance status, education level, income, hypertension, atrial fibrillation, left ventricular hypertrophy, diabetes mellitus, and smoking status with age, sex, and sickle cell trait individually added with interaction terms. Multiple imputations were performed for missing data.

Table 4. Pre-specified sensitivity analyses for the association of *HBA* genotype with incident ischemic stroke – fully adjusted models with the separate addition of covariates for each model.

	HR	95% CI	P value*
Hemoglobin			
HBA copy number	1.05	(0.89,1.24)	0.55
Hemoglobin	0.96	(0.86, 1.08)	0.52
Chronic kidney disease			
HBA copy number	1.03	(0.88,1.20)	0.75
Chronic kidney disease	1.81	(1.49, 2.20)	<0.001
Hemoglobin and Chronic kidney disease ‡			
HBA copy number	1.03	(0.88, 1.21)	0.71
Hemoglobin	0.99	(0.88, 1.10)	0.83
Chronic kidney disease	1.81	(1.49, 2.20)	<0.001
Regular aspirin use			
HBA copy number	1.04	(0.88, 1.21)	0.67
Regular aspirin use	1.26	(1.05, 1.51)	0.01
Regular statin use			
HBA copy number	1.04	(0.88, 1.21)	0.66

Regular statin use	1.05	(0.87, 1.28)	0.60
Principal components of ancestry			
HBA copy number	1.01	(0.86,1.19)	0.89
PC1	1.01	(0.91,1.12)	0.86
PC2	0.97	(0.88,1.07)	0.55
PC3	0.99	(0.90,1.09)	0.89
PC4	0.97	(0.89,1.07)	0.55
PC5	1.07	(0.98,1.18)	0.14
PC6	1.05	(0.96,1.16)	0.27
PC7	0.92	(0.84,1.01)	0.08
PC8	1.02	(0.92,1.12)	0.75
PC9	0.90	(0.82,0.99)	0.03
PC10	1.04	(0.95,1.15)	0.36
rs11248850, *HBA* regulatory SNP			
HBA copy number	1.04	(0.89,1.22)	0.63
Rs11248850	0.98	(0.83, 1.15)	0.81
rs7203560, *HBA* regulatory SNP			

HBA copy number	1.06	(0.89, 1.25)	0.52
rs7203560	1.09	(0.83, 1.44)	0.52

HBA = alpha globin gene; HR = hazard ratio; CI = confidence interval; PC = principal component; SNP = single nucleotide polymorphism

*P values were generated with Cox proportional hazards multivariable regression models employing a linear effect (i.e., additive model for risk) of *HBA* allele count on the log of the hazard ratio. Each model was adjusted for *HBA* genotype, age per year, sex, region, medically insurance status, education level, income, hypertension, atrial fibrillation, left ventricular hypertrophy, diabetes mellitus, and smoking status with listed covariates added in separate models. Multiple imputations were performed for missing data. [†]Principal components of ancestry model with n=7,032 participants and all other models with n=8,947. [‡]This sensitivity analysis included both chronic kidney disease and hemoglobin added to the fully adjusted base model.