Supplementary Online Content

Hyacinth HI, Carty CL, Seals SR, et al. Association of sickle cell trait with ischemic stroke among African Americans: a meta-analysis. *JAMA Neurol.* Published online April 23, 2018. doi:10.1001/jamaneurol.2018.0571

eMethods. Supplemental Methods

This supplementary material has been provided by the authors to give readers additional information about their work.

eMethods. Supplemental Methods

Cohort Descriptions

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.[1-4] 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.[5]

A total of 2,133 subjects (having the relevant phenotype and covariate data) were included in the current study, of which 1,827 were directly genotyped for rs334 via sequencing, the rest were imputed.

Multi-Ethnic Study of Atherosclerosis (MESA)

The MESA study was designed to investigate the correlates of subclinical CVD progression in a longitudinal multi-ethnic cohort free of CVD at baseline. Detailed methods of MESA were previously reported.[6] Participants were excluded if they had physician-diagnosed cardiovascular disease prior to enrollment, including angina, myocardial infarction, heart failure, stroke or TIA, resuscitated cardiac arrest or a cardiovascular intervention (e.g., CABG, angioplasty, valve replacement, or pacemaker/defibrillator placement). Pre-specified recruitment plans identified four racial/ethnic groups (White European-American, African-American, Hispanic-American, and Chinese-American) for enrollment, with targeted oversampling of minority groups to enhance statistical power. Between July 2000 and August 2002, a total of 10,966 individuals were screened, and 6,814 individuals self-identified as White, African-American, Hispanic, or Chinese who were aged 45-84 years were deemed eligible and enrolled in MESA from 6 centers in the United States (Baltimore, MD; Chicago, IL; Forsyth County, NC; Los Angeles, CA; New York, NY; and St. Paul, MN). The institutional review board at each participating institution approved the study, and all participants signed informed consent.

For this study, participant characteristics were obtained from data collected at the enrollment visit from physical measures, standardized questionnaires, and laboratory tests. These included

demographic information (age, sex, race/ethnicity), medical history, medications, and alcohol and tobacco use. Resting blood pressure was determined by taking three measurements with the participant in the seated position. Systolic and diastolic blood pressures were recorded as the average value of the last two measurements from both the first and second study examinations. Participants have returned for 4 visits, in 2002–2004 (exam 2), 2004–2005 (exam 3) 2005–2007 (exam 4), and 2010-2012 (exam 5).

A total of 1,556 subjects (who had complete genotype, phenotype and covariate data) were included in the current study, of which 152 were directly genotyped for rs334 via sequencing and the rest were imputed using data from genome-wide SNP genotyping array.

REason for Geographic And Racial Difference in Stroke (REGARDS)

REGARDS is a longitudinal population-based cohort study designed to investigate the factors associated with excess stroke mortality among blacks and residents of the stroke belt region of the United States of America.[7] The stroke belt is defined as the 8 Southern states of North Carolina, South Carolina, Georgia, Tennessee, Mississippi, Alabama, Louisiana, and Arkansas.[8] REGARDS sampled a total of 30,239 community-dwelling adult individuals aged ≥45 years, self-reported as Non-Hispanic Black or White. Sampling was started in January, 2003 and was completed by October, 2007. Potential enrollees where determined from a well characterized commercially available list from the GENESYS database/sampling system. The exclusion criteria included any self-reported medical conditions (e.g. cancer) that would prevent long-term participation, or being on a waiting list for a nursing home. The sample design was such that participants were balanced on race and sex, and across the stroke buckle and stroke belt, and the rest of the contiguous United States. The resulting sample was such that, 21% were from the

stroke buckle, 35% from the rest of the stroke belt area (i.e. minus the buckle), 44% from the other 40 contiguous states, 42% of the total sample were Blacks, and 55% were women.

Initial study contact involved mailed information about the study. This was followed by a computer-assisted telephone interview to obtain health history and other risk factors, then an in home visit where written informed consent was obtained, physical measures and an electrocardiogram completed, and blood samples collected for laboratory measures. Participants or their proxies were contact every six months to obtain medical history and identify hospitalizations, emergency department visits, overnight stays in nursing homes or rehabilitation centers, or death within the last 6 months. For those who reported events that could be stroke, medical records were obtained and events were adjudicated by an expert panel of trained adjudicators, based on prespecified adjudication criteria (see below).

A total of 10,489 participants had direct Taqman® genotyping for rs334 of which 9,759 had the relevant covariate data, and were included in this study.

Women's Health Initiative (WHI)

WHI is one of the largest prospective population-based cohort study investigating post-menopausal women's health in the U.S. A total of 161, 808 women aged 50–79 years old were recruited from 40 U.S. clinical centers between 1993 and 1998 to participate in the observational study (OS) and in clinical trials (CT): postmenopausal hormone therapy (estrogen alone or estrogen plus progestin), a calcium and vitamin D supplement trial, and a dietary modification trial.[9] A diverse population including 26,045 (17%) women from minority groups were recruited from 1993-1998 at 40 clinical centers across the U.S. Recruitment was done through mass

mailing to age-eligible women obtained from voter registration, driver's license and Health Care Financing Administration or other insurance list, with emphasis on recruitment of minorities and older women. Exclusions included participation in other randomized trials, predicted survival < 3 years, alcoholism, drug dependency, mental illness, and dementia. For the clinical trial, women were ineligible if they had a systolic BP > 200 mm Hg or diastolic BP > 105 mm Hg, a history of hypertriglyceridemia or endometrial cancer.

A total of 6015 subjects from WHI (having the relevant phenotype and covariate data) were included in the current study, of which 566 were directly genotyped for rs334 via sequencing and the rest were imputed.

Exome Sequencing

Exome sequencing (N=2,052) was performed through the National Heart, Lung, and Blood Institute (NHLBI) Exome Sequencing Project (ESP) (CHS=90, MESA=146, JHS=367, WHI-567) and in the NHLBI Minority Health Genomics and Translational Research Bio-Repository Database (MH-GRID) (N=311 from JHS) as described in Fu et al.,[10] and in the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) Type 2 Diabetes Genetic Exploration by Next-Generation Sequencing in Multi-Ethnic Samples (T2D-GENES) (N=571 from JHS).[11] In all three sequencing studies, the rs334 variant was covered at high depth (mean depth = 131(ESP), (MH-GRID), (T2D-GENES).

Genotyping and Imputation

Direct genotype data for the rs334 variant were obtained by custom genotyping from the REGARDS study. Whole blood DNA was isolated from the buffy coat layer using the Gentra Puregene Blood Kit (Qiagen, Inc., Valencia, CA; www.qiagen.com). Carriers of HbS were

identified from biallelic variation in the single nucleotide polymorphism, rs334. In addition, carriers of HbC were identified from rs3393016. Genotyping was performed using functionally tested TaqMan® SNP Genotyping Assays in accordance with manufacturer protocols (Life Technologies, Grand Island, NY; www.lifetechnologies.com). The following custom primer and probe sequences were used to capture biallelic variation: rs334 (A/T) Forward-TCAAACAGACACCATGGTGCAT, Reverse-CCCCACAGGGCAGTAACG, VIC-CTGACTCCTGAGGAGAA-MGB, 6FAM-CTGACTCCTGTGGAGAA-MGB; and rs33930165 (A/G) Forward-AAACAGACACCATGGTGCATCT, Reverse-CCCCACAGGGCAGTAACG, VIC-CAGACTTCTCCTTAGGAGTC-MGB, 6FAM-ACTTCTCCTCAGGAGTC-MGB (designed on complement strand). PCR product in a 5.5 µL reaction volume was amplified utilizing 0.9 µM of each forward and reverse primer, 0.2 µM of each FAM and VIC sequence-specific probe, 3 ng DNA, and 1X TaqMan Universal PCR Master Mix containing AmpliTaq Gold DNA Polymerase and no AmpErase UNG. After an initial step of 10 min at 95ºC, the products were amplified using 50 cycles of 15 s at 92ºC and 1 min at 60ºC. Allele detection and genotype calling were performed using the ABI 7900HT and the Sequence Detection System software (Life Technologies, formerly Applied Biosystems). Quality control measures utilized a blind duplicate program in which 5% of samples were re-genotyped at random.

Imputation of the rs334 variant was carried out as described in Auer et al.[12] Briefly, 2,163 participants (1,692 AAs, 471 EAs) from ESP with Affymetrix 6.0 genome-wide genotyping data were selected to form an imputation reference panel. The imputation target panel consisted of 15,826 AAs from WHI, ARIC, MESA, CARDIA, and JHS, of which 6,664 had sufficient phenotype information to include in this analysis. Standard quality control measures were performed.[12] Both target and reference panels were pre-phased using BEAGLE.[13] The reference

panel was then imputed into the target using minimac.[14] The imputation quality score Rsq (which

is equivalent to the squared correlation between proximal imputed and genotyped SNP) was

0.86. Individuals with the rs334 genotype derived via imputation were coded as having 0, 1, or 2

risk alleles using the most probable genotype. The imputation of genotype in these cohorts, have

been done as part of multiple projects[15-18] and the similarity in description with this supplemental

methods is due to the fact that the same set of data is being utilized for the current study.

Difference in the number of participants between projects is due to phenotype and/or covariate

missingness, which is purely random.

Using N=1,132 samples with both exome sequence and imputed genotype data from JHS, we

validated our imputed genotypes (Table). The kappa correlation for sequenced and imputed

values was 0.88 (95%CI 0.84 - 0.92).

Table. Cross tabulation of imputed genotypes versus genotypes via imputation, for the rs334 variant in 1,132 samples from JHS.

rs334 via imputation	rs334 via sequencing			
	HbAA	HbAS	HBSS	Total
HbAA	1,018	6	0	1,024
HbAS	15	91	0	106
HbSS	0	1	1	2
Total	1,033	98	1	1,132

Incident Stroke Definitions by cohort

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 cases where the WHO criteria was not fully satisfied, but there is enough 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.[19,20]

MESA

In addition to the standard MESA examinations, participants or next of kin were contacted by telephone every 9-12 months for information on interim changes in medical status. Copies of death certificates, and medical records of hospitalizations and outpatient cardiovascular diagnoses were obtained. Incident strokes were determined via adjudication by an expert panel of trained stroke neurologists (vascular neurologists). Stroke was defined as "a rapid onset of

neurologic deficit, headache, or meningismus, with neurologic deficits not secondary to brain trauma (closed head injury), tumor, infection (e.g., encephalitis or meningitis), or other nonvascular cause. Although clinical evidence or suspicion of embolic stroke secondary to SBE were counted as stroke, as well as clinically relevant lesion on brain imaging (irrespective of duration, imaging modality or stroke subtype). Definition of stroke was similar to that of the WHO.[21]

REGARDS

After initial review by a stroke nurse to exclude obvious none cases, medical records were reviewed by at least 2 physician members of a committee of stroke experts to validate and classify potential strokes. Disagreements were resolved by full committee review. Over time, if an adjudicator's rate of disagreement with other adjudicators was >20%, the involved adjudicator underwent retraining. Stroke events were defined following World Health Organization (WHO) definition as ''rapidly developing clinical signs of focal, at times global, disturbance of cerebral function, lasting more than 24 hours or leading to death with no apparent cause other than that of vascular origin.[21]'' Events not meeting this definition but characterized by symptoms lasting <24 hours, with neuroimaging consistent with acute ischemia or hemorrhage were classified as ''clinical strokes.'' ''Probable stroke'' was defined for cases in which adjudicators agreed that the event was likely a stroke or death related to stroke but information was incomplete for WHO or clinical classification. Strokes were further classified as ischemic or hemorrhagic. Sensitivity analyses were performed to assess if results differed when analyses were restricted to only WHO strokes.

WHI

WHI participants were contacted every 6-12 months for self-reported stroke related hospitalizations or death. Participants in the clinical trial component of the study were contacted semi-annually, while those in the observational study were contacted annually. When a stroke event is reported following the semiannual or annual phone-call/survey, the medical record is retrieved, abstracted and the event adjudicated by a panel of trained vascular neurologists. For WHI, only hospitalized stroke events were recorded and adjudicated. In addition, only the first stroke after enrollment were adjudicated and recorded. WHI also used a federated (local) adjudication system for adjudicating stroke events for a while, before switching to central adjudication. Definition of stroke was based on the WHO criteria,[21] but where there is brain imaging (usually the case) with clinical evidence, duration of symptom did not need to meet the 24 hours threshold stipulated in the WHO definition.

eReferences

1.   Taylor HA, Jr. The Jackson Heart Study: an overview. *Ethn Dis*. 2005;15(4 Suppl 6):S6-1-3.
2.   Carpenter MA, Crow R, Steffes M, et al. Laboratory, reading center, and coordinating center data management methods in the Jackson Heart Study. *Am J Med Sci*. 2004;328(3):131-144.
3.   Fuqua SR, Wyatt SB, Andrew ME, et al. Recruiting African-American research participation in the Jackson Heart Study: methods, response rates, and sample description. *Ethn Dis*. 2005;15(4 Suppl 6):S6-18-29.
4.   Payne TJ, Wyatt SB, Mosley TH, et al. Sociocultural methods in the Jackson Heart Study: conceptual and descriptive overview. *Ethn Dis*. 2005;15(4 Suppl 6):S6-38-48.
5.   Wyatt SB, Akylbekova EL, Wofford MR, et al. Prevalence, awareness, treatment, and control of hypertension in the Jackson Heart Study. *Hypertension*. 2008;51(3):650-656.
6.   Bild DE, Bluemke DA, Burke GL, et al. Multi-ethnic study of atherosclerosis: objectives and design. *Am J Epidemiol*. 2002;156(9):871-881.

7.   Howard VJ, Cushman M, Pulley L, et al. The Reasons for Geographic and Racial Differences in Stroke Study: Objectives and Design. *Neuroepidemiology*. 2005;25(3):135-143.

8.   Howard G, Anderson R, Johnson NJ, Sorlie P, Russell G, Howard VJ. Evaluation of Social Status as a Contributing Factor to the Stroke Belt Region of the United States. *Stroke*. 1997;28(5):936-940.

9.   Borel MJ, Buchowski MS, Turner EA, Goldstein RE, Flakoll PJ. Protein turnover and energy expenditure increase during exogenous nutrient availability in sickle cell disease. *Am J Clin Nutr*. 1998;68(3):607-614.

10.   Fu W, O'Connor TD, Jun G, et al. Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. *Nature*. 2013;493(7431):216-220.

11.   Consortium STD, Williams AL, Jacobs SB, et al. Sequence variants in SLC16A11 are a common risk factor for type 2 diabetes in Mexico. *Nature*. 2014;506(7486):97-101.

12.   Auer PL, Johnsen JM, Johnson AD, et al. Imputation of exome sequence variants into population-based samples and blood-cell-trait-associated loci in African Americans: NHLBI GO Exome Sequencing Project. *The American Journal of Human Genetics*. 2012;91(5):794-808.

13.   Browning SR, Browning BL. Rapid and Accurate Haplotype Phasing and Missing-Data Inference for Whole-Genome Association Studies By Use of Localized Haplotype Clustering. *The American Journal of Human Genetics*;81(5):1084-1097.

14.   Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet*. 2012;44(8):955-959.

15.   Naik RP, Derebail VK, Grams ME, et al. Association of sickle cell trait with chronic kidney disease and albuminuria in African Americans. *JAMA*. 2014;312(20):2115-2125.

16.   Lacy ME, Wellenius GA, Sumner AE, et al. Association of Sickle Cell Trait With Hemoglobin A1c in African Americans. *Jama*. 2017;317(5):507-515.

17.   Bello NA, Hyacinth HI, Roetker NS, et al. Sickle cell trait is not associated with an increased risk of heart failure or abnormalities of cardiac structure and function. *Blood*. 2017;129(6):799-801.

18.   Naik RP, Wilson JG, Ekunwe L, et al. Elevated D-dimer levels in African Americans with sickle cell trait. *Blood*. 2016;127(18):2261-2263.

19.   Keku E, Rosamond W, Taylor HA, Jr., et al. Cardiovascular disease event classification in the Jackson Heart Study: methods and procedures. *Ethn Dis*. 2005;15(4 Suppl 6):S6-62-70.

20.   Rosamond WD, Folsom AR, Chambless LE, 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(4):736-743.

21.   Stroke--1989. Recommendations on stroke prevention, diagnosis, and therapy. Report of the WHO Task Force on Stroke and other Cerebrovascular Disorders. *Stroke*. 1989;20(10):1407-1431.