SUPPLEMENTAL MATERIAL

Expanded Materials & Methods

1. Participating studies in the discovery stage

Atherosclerosis Risk in Communities Study (ARIC)

The ARIC study is a population-based cohort study consisting of 15,792 men and women that were drawn from four U.S. communities (Suburban Minneapolis, Minnesota; Washington County, Maryland; Forsyth County, North Carolina, and Jackson, Mississippi) [24]. It was designed to investigate the causes of atherosclerosis and its clinical outcomes, and variation in cardiovascular risk factors, medical care, and disease by race, sex, location, and date. Participants were between age 45 and 64 years at their baseline examination in 1987-1989 when blood was drawn for DNA extraction and participants consented to genetic testing. Hospitalized strokes that occurred by December 31, 2011 are included. During annual telephone contacts, trained interviewers asked each ARIC participant to list all hospitalizations during the past year. Hospital records for any hospitalizations identified were then obtained. In addition, all local hospitals annually provided lists of stroke discharges (International Classification of Diseases, Ninth Revision, Clinical Modification codes 430 to 438), which were scrutinized for ARIC participant discharges. Details on quality assurance for ascertainment and classification of stroke are described elsewhere [25]. Briefly, the stroke diagnosis was assigned according to criteria adapted from the National Survey of Stroke [26]. Strokes secondary to trauma, neoplasm, hematologic abnormality, infection, or vasculitis were excluded, and a focal deficit lasting <24 hours was not considered to be a stroke. Out-of-hospital stroke was not ascertained and validated; thus, these potential stroke events were not included. Strokes were classified into hemorrhagic stroke (subarachnoid and intracerebral hemorrhage) and ischemic stroke. A stroke was classified as ischemic when a brain CT or MRI revealed acute infarction and showed no evidence of hemorrhage. Only individuals free of stroke or TIA at baseline were included in the analysis.

BioMe™ Biobank (BioMe)

The Charles Bronfman Institute for Personalized Medicine at Mount Sinai Medical Center (MSMC), BioMe Biobank, founded in September 2007, is an ongoing, broadly-consented electronic health record-linked clinical care biobank that enrolls participants non-selectively from the Mount Sinai Medical Center patient population. The MSMC serves diverse local communities of upper Manhattan, including Central Harlem (86% African American), East Harlem (88% Hispanic/Latino), and Upper East Side (88% Caucasian/White) with broad health disparities. For the current analysis, adult (>18 years of age) coronary artery disease (CAD) cases and controls were included. A Case-Definition-Algorithm (CDA), incorporating International Classification of Diseases (ICD) codes and Current Procedural Terminology (CPT) codes, was used to identify individuals with CAD along with suitable controls.

Cardiovascular Health Study (CHS)

The Cardiovascular Health Study (CHS) is a population-based cohort study of risk factors for coronary heart disease and stroke in adults 65 years and older conducted across four field centers [27]. The original predominantly European ancestry cohort of 5,201 persons was recruited in 1989-1990 from random samples of people on Medicare eligibility lists from four US communities. Subsequently, an additional predominantly African-American cohort of 687 persons was enrolled for a total sample of 5,888. Institutional review committees at each field center approved the CHS, and participants gave informed consent. Blood samples were drawn from all participants at their baseline examination, and DNA was subsequently extracted from available samples. These analyses were limited to participants with available DNA who also consented to genetic studies. Participants were examined annually from enrollment to 1999 and continued to be under surveillance for stroke following 1999. Since baseline, participants have also been contacted twice a year to identify potential cardiovascular events, including stroke. In addition, all hospitalizations were screened for potential stroke events. For suspected fatal and non-fatal events occurring with or without hospitalization, information was collected from the participant or next of kin, from medical records, and if needed, from the participant's physician. When available, scans or reports of

CT, MRI or both were reviewed centrally. Finally, at a consensus conference using all available information, vascular neurologists adjudicated the occurrence of fatal and non-fatal stroke, stroke types, and subtypes [28].

Framingham Heart Study (FHS)

FHS is a three-generation, single-site, community-based, ongoing cohort study that was initiated in 1948 to investigate prospectively the risk factors for CVD including stroke. It now comprises 3 generations of participants (N=10,333): the Original cohort followed since 1948 [29]; their Offspring and spouses of the Offspring, followed since 1971 [30]; and children from the largest Offspring families enrolled in 2000 (Gen 3) [31]. The Original cohort enrolled 5,209 men and women who comprised two-thirds of the adult population then residing in Framingham, MA. Survivors continue to receive biennial examinations. The Offspring cohort comprises 5,124 persons (including 3,514 biological offspring) who have been examined approximately once every 4 years. The population of Framingham was virtually entirely white (Europeans of English, Scots, Irish and Italian descent) in 1948 when the Original cohort was recruited. At the initial examination participants were asked for the country of birth and whether or not they had any Italian ancestry. At a later examination (the 8th) the Offspring cohort participants were asked to identify their race from the following choices: Caucasian or white, African-American or black, Asian, Native Hawaiian or other Pacific Islander, American Indian or Alaska native or 'prefer not to answer'. They were either asked to identify their ethnicity as either 'Hispanic or Latino' or not. Almost all the FHS Original and Offspring participants are white/Caucasian and none were excluded from the discovery cohort. At each clinic exam, participants receive questionnaires, physical examinations, and laboratory testing; between examinations, they remain under surveillance (regardless of whether or not they live in the vicinity) via physician referrals, record linkage, and annual telephone health history updates. Incident strokes have been identified since 1948 through this ongoing system of FHS clinic and local hospital surveillance and methods used have been detailed previously [32-34]; they include a review of medical records and collaboration with local general practitioners, emergency rooms, and imaging facilities. If a participant saw a physician or was admitted to the hospital, visited an emergency room or obtained any brain imaging between biennial examinations for symptoms suggestive of TIA or stroke, a stroke neurologist from the Heart Study attempted to visit the person within 48 hours and recorded a complete history and neurological examination; this was repeated at 1, 3 and 6 months. All medical records from practitioners, hospitals, imaging centers, rehabilitation centers, and nursing homes were procured for review. A panel of 3 investigators (at least 2 neurologists) adjudicated the diagnosis of stroke and determined stroke subtype in each case based on the Framingham evaluations and external records. The recruitment of Original and Offspring cohort participants at FHS had occurred long before the DNA collection with the result that the majority of stroke events in the FHS (although ascertained prospectively) were prevalent at the time of DNA collection and were excluded from these analyses.

Jackson Heart Study (JHS)

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 5,306 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 [35-39]. 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 participants provided written informed consent. Unrelated participants were between 35 and 84 years old, and members of the family cohort were $\geq$ 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 at baseline 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 [39]. 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 representatives provide self-reported information of hospitalization or death. Identification and abstraction of CVD illness and death data are performed by a certified medical record abstractor. Incident stroke is defined as stroke that occurred while the participants was enrolled the study, i.e. stroke event occurred after the baseline visit. Strokes are classified as either definite or probable stroke. The definition of stroke was based on the World Health Organization (WHO) criteria for definition of stroke or clinical criteria in which case the WHO criteria might not have been satisfied, but there is clinical evidence sufficient for a diagnosis of stroke to be made. More details on identification and classification of stroke events in the JHS have already been published [40].

Multi-Ethnic Study of Atherosclerosis (MESA)

The MESA study is a study of the characteristics of subclinical cardiovascular disease (disease detected non-invasively before it has produced clinical signs and symptoms) and the risk factors that predict progression to clinically overt cardiovascular disease or progression of the subclinical disease [41]. MESA researchers study a diverse, population-based sample of 6,814 asymptomatic men and women aged 45-84. Thirty-eight percent of the recruited participants are white, 28 percent African-American, 22 percent Hispanic, and 12 percent Asian, predominantly of Chinese descent. Participants were recruited from six field centers across the United States: Wake Forest University, Columbia University, Johns Hopkins University, University of Minnesota, Northwestern University and the University of California - Los Angeles. The first examination took place over two years, from July 2000 - July 2002. It was followed by four examination periods that were 17-20 months in length. Participants have been contacted every 9 to 12 months throughout the study to assess clinical morbidity and mortality [18]. Prevalent stroke was an exclusion criterion for MESA at baseline. New occurrences of stroke were recorded over 9-years of follow-up. In brief, a telephone interviewer contacted each participant every 9–12 months. Information about all new cardiovascular conditions, hospital admissions, cardiovascular outpatient diagnoses, treatments, and deaths were obtained. To verify self-reported diagnoses, information was collected from death certificates and medical records for all hospitalizations and outpatient cardiovascular diagnoses, using ICD-9 and ICD-10 codes. In the case of out-of-hospital deaths, next-of-kin interviews or questionnaires were administered to physicians, relatives or friends. Two physicians from the MESA study events committee independently reviewed all medical records for endpoint classification and assignment of incidence dates. The reviewers were blinded to the study data. If the reviewing physicians disagreed on the event classification, they adjudicated differences. Neurologists reviewed and classified stroke as present if there was a focal neurologic deficit lasting 24 hours or until death, or if <24h, there was a clinically relevant lesion on brain imaging and no nonvascular cause. Patients with focal neurological deficits secondary to brain trauma, tumor, infections, or other non-vascular cause were excluded [42]. Ischemic strokes were distinguished from hemorrhagic strokes using findings on imaging, surgery, autopsy, or some combination of these. Ischemic stroke subtypes were assigned based on an extension of the Trial of Org 10172 in Acute Stroke Treatment (TOAST) scheme to try to reduce the number classified as undetermined.

Women's Health Initiative (WHI)

WHI is a long-term, prospective, multi-center cohort study that investigates post-menopausal women's health [43]. WHI was funded by the National Institutes of Health and the National Heart, Lung, and Blood Institute to study strategies to prevent heart disease, breast cancer, colon cancer, and osteoporotic fractures in women 50-79 years of age. WHI involves 161,808 women recruited between 1993 and 1998 at 40 centers across the US. The study consists of two parts: the WHI Clinical Trial which was a randomized clinical

trial of hormone therapy, dietary modification, and calcium/Vitamin D supplementation, and the WHI Observational Study, which focused on many of the inequities in women's health research and provided practical information about the incidence, risk factors, and interventions related to heart disease, cancer, and osteoporotic fractures. Stroke diagnosis requiring and/or occurring during hospitalization was based on the rapid onset of a neurological deficit attributable to an obstruction or rupture of an arterial vessel system. Hospitalized incident stroke events were identified by semiannual questionnaires and adjudicated following medical record review, which occurred both locally and centrally. Ischemic strokes were further classified by the central neurologist adjudicators according to the Trial of Org 10172 Acute Stroke Trial (TOAST) criteria to examine stroke subtypes. The TOAST classification focuses on the presumed underlying stroke mechanism and requires detailed investigations (such as brain computed tomography, magnetic resonance imaging, angiography, carotid ultrasound, and echocardiography). Venous thromboembolism (VTE) cases were excluded from the control group.

UK Biobank (UKBB)

UK Biobank is a large long-term biobank study in the United Kingdom aimed to investigate the respective contributions of genetic predisposition and environmental exposure to the development of diseases. Only European ancestry and unrelated participants were included in the current analysis. Cases were defined through algorithmic combinations of coded information from UKBB's baseline assessment data collection (which included data from participants on their self-reported medical conditions, operations and medications), along with linked data from hospital admissions (diagnoses and procedures) and death registries. The classification is based on algorithms developed by the UKBB outcome adjudication group, aiming to classify disease outcomes with high positive predictive value (i.e. a high probability that people classified as being positive for a health-related event have indeed experienced that event). A total of 4,474 IS cases (and 24,000 controls), 959 ICH cases (and 4,800 controls), and 1,194 SAH cases (and 5,970 controls) of European ancestry from the UKBB were selected for analysis. Related individuals were excluded from the UKBB analysis. Controls were selected to be stroke free and maintain the same male:female ratio as the cases and an overall approximate 5:1 control-to-case ratio.

Genotyping of UKBB participants was performed using either the Affymetrix UK BiLEVE Axiom array or the Affymetrix UK Biobank Axiom® array, with QC procedures performed at both the variants and the sample level (http://www.ukbiobank.ac.uk/wp-content/uploads/2014/04/UKBiobank_genotyping_QC_documentation-web-1.pdf). Imputation was performed based on reference panels from the Haplotype Reference Consortium, UK10K, and the 1000 Genome Phase 3 using MACH (http://csg.sph.umich.edu/abecasis/MACH/index.html). Genetic variants with MAF>0.1% and imputation quality score R2>0.3 were included in the association analysis.

2. Participating studies in the replication stage

Stroke Genetics Network (SiGN) Consortium

The SiGN Consortium was formed in 2010 to identify common genetic variants associated with ischemic stroke and its subtypes using state-of-the-art stroke subtyping and the genome-wide association study approach in 16,851 cases and 32,473 controls from multiple sites [17]. To increase coverage of the genome, imputation was performed using the TOPMed WGS data as the reference panel and association analysis in each participating study was repeated using the newly imputed data. A total of 454 stroke cases from WHI were previously included in SiGN, and meta-analysis combining all studies in SiGN was performed without WHI.

TOPMed Blood Pressure Working Group

Look-up of the five novel loci in the TOPMed Blood Pressure Working Group was performed using freeze6 data. Quality control and harmonization of systolic blood pressure (SBP), diastolic blood pressure (DBP), and hypertension phenotypes across studies followed a standard protocol developed by the TOPMed BP Working Group. Prior to harmonization, each study removed unlikely BP values, such as SBP<60 mmHg, SBP>300 mmHg, or DBP>SBP. Phenotype harmonization involved using the mean of two BP measures

from a study visit to estimate SBP and DBP for each participant. For longitudinal studies, we used BP measures obtained at the first visit, which optimized sample size and minimized the percent of participants taking antihypertension medications. If a participant was taking antihypertension medication, SBP and DBP values were imputed by adding 15 and 10 mmHg, respectively, to observed values. Hypertension was defined as mean SBP>140 mmHg, mean DBP>90 mmHg, or current use of anti-hypertension medication. The harmonized BP and hypertension values were pooled across studies to create a single, multi-study phenotype dataset. Continuous systolic blood pressure (SBP) and diastolic blood pressure (DBP) phenotypes were then transformed within each ancestry strata by first regressing BP on age, sex, body mass index (BMI), study, and the first 11 ancestry principal components (PCs). BP residuals then underwent rank-based inverse normal transformation and were subsequently rescaled by multiplying the transformed value by the standard deviation of the original harmonized BP phenotype, an approach that allowed subsequent effect size estimates to reflect clinically meaningful values in mmHg.

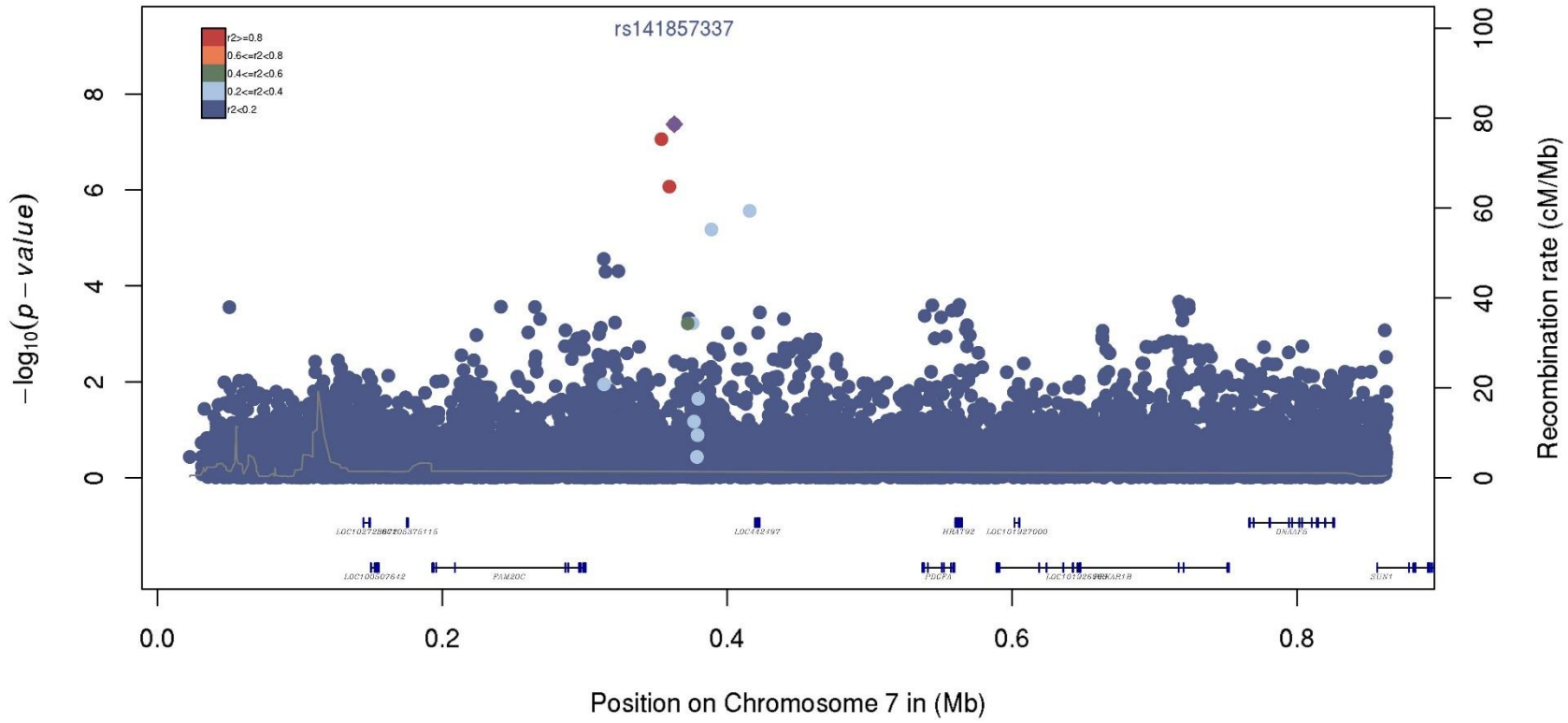
Association of each locus with the transformed SBP and DBP phenotypes as continuous outcomes and with hypertension as a dichotomous outcome was tested in 50,755 participants from 18 studies. Single nucleotide variants (SNVs) with MAC ≥10 were individually tested for association with the continuous BP phenotypes using a linear mixed model that accommodated familial relationships using a sparse kinship matrix, adjusted for age, sex, BMI, study, and ancestry PCs, and, for the multi-ancestry analyses only, fit separate (heterogeneous) residual variance components for each ancestry group. Single variant analyses of the binary hypertension phenotype employed a logistic mixed model that again accommodated familial relationships using a sparse kinship matrix, and adjusted for age, sex, BMI, study, and ancestry PCs. Association analyses were performed in all participants as well as in each ancestral group (European, African American, and Hispanic-specific analysis).

3. Functional annotation

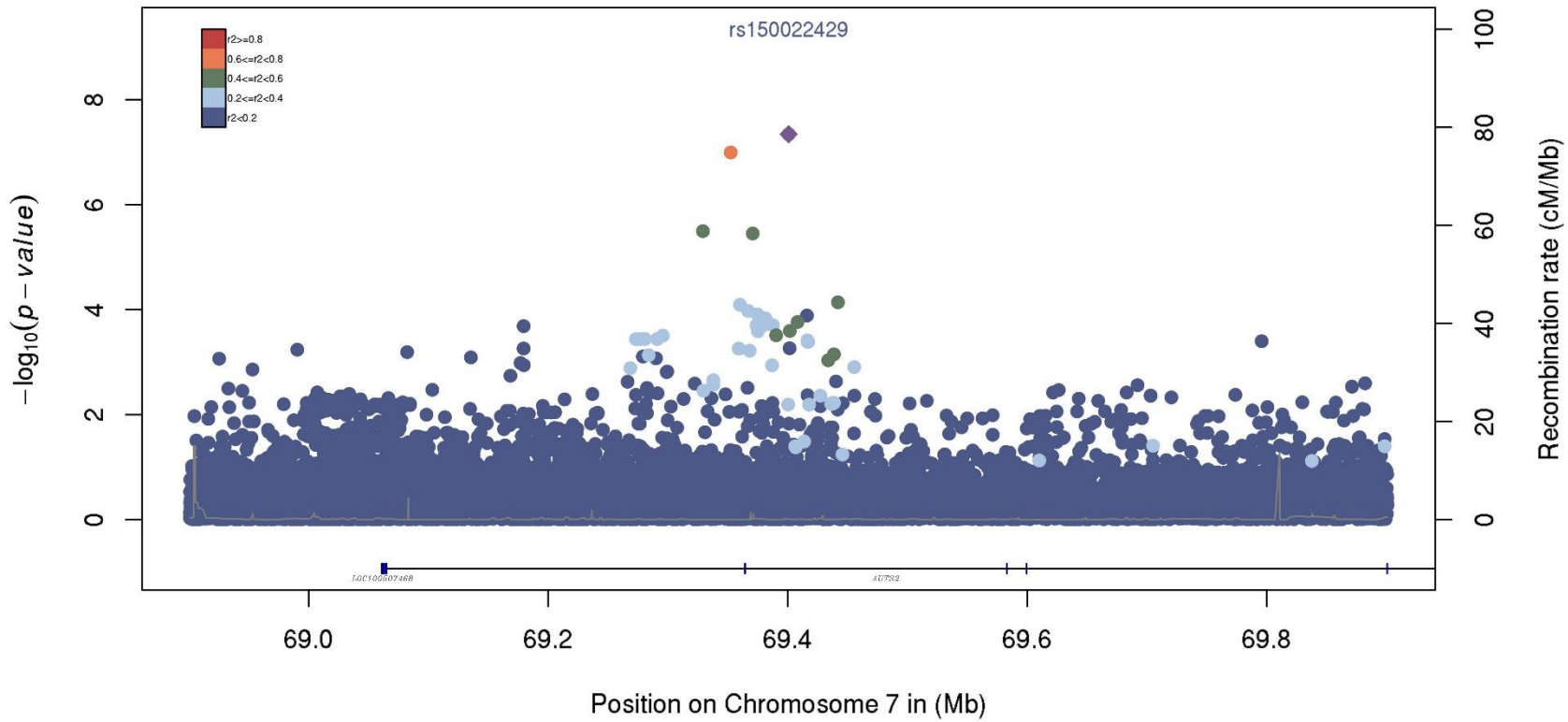
Bioinformatic follow-up was performed for each novel locus using a comprehensive functional annotation database constructed with the whole genome sequence annotator (WGSA) [50] and a custom UCSC analysis data hub visualizing enhancer and repressor activities, DNase I hypersensitive sites (DHS) and transcribed regions in selected tissues (aorta, left and right ventricle, right atrium, and all brain-related sub-tissues). The functional annotations included gene-centric function (GTEx [51]) and genome-wide functional prediction scores (DANN [52] and Eigen-PC [53]). Variants with DANN score≥0.9 were coded as deleterious, and variants with Eigen PC score>0 were coded as functional. Custom UCSC bed tracks included the lead variant of each novel or suggestive locus and variants in LD ($r^2 \geq 0.4$) with the lead variant within ±500kb. The LD matrix was generated using TOPMed WGS data and the samples included in our association analysis. The LD proxies of the five loci of interest were generated using either African-specific or Hispanic-specific data (Hispanic-specific data used for *13q33* and African-specific data used for the other four loci). Chromatin immunoprecipitation-sequencing signals associated with enhancers (H3K27ac and H3K4m1), repressors (H3K27me3), and transcribed regions (H3K36me3) were examined in each selected tissue.

Supplemental Figure I. LocusZoom plots of the five novel loci identified in TOPMed. Genetic coordinates are displayed along the x-axis (hg38) and genome-wide association significance level is plotted against the y-axis as -log10(*P-value*). LD is indicated by color scale in relationship to the most significant variant (colored as purple diamond) in each association (red: r2≥0.8, orange: 0.6≤r2<0.8, green: 0.4≤ r2<0.6, blue: 0.2≤r2<0.4, navy: r2<0.2). For *7q22*, *AUTS2*, *RAP1GAP2*, and *TEX13C* loci, LD information was extracted from African ancestry samples in TOPMed. For the *13q33* locus, LD information was extracted from Hispanic ancestry samples in TOPMed. (A) *7q22*; (B) *AUTS2*; (C) *13q33*; (D) *RAP1GAP2*; (E) *TEX13C*.

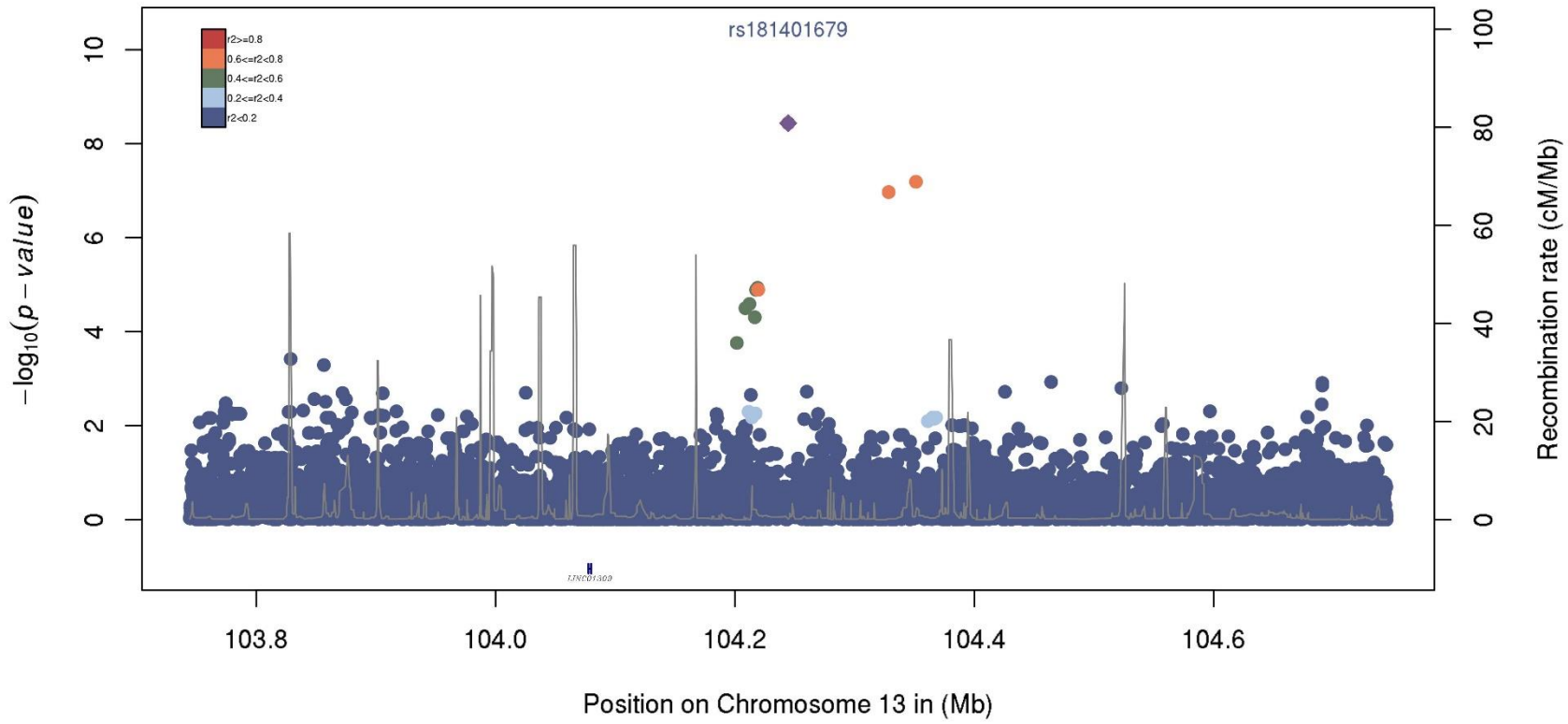
(A)



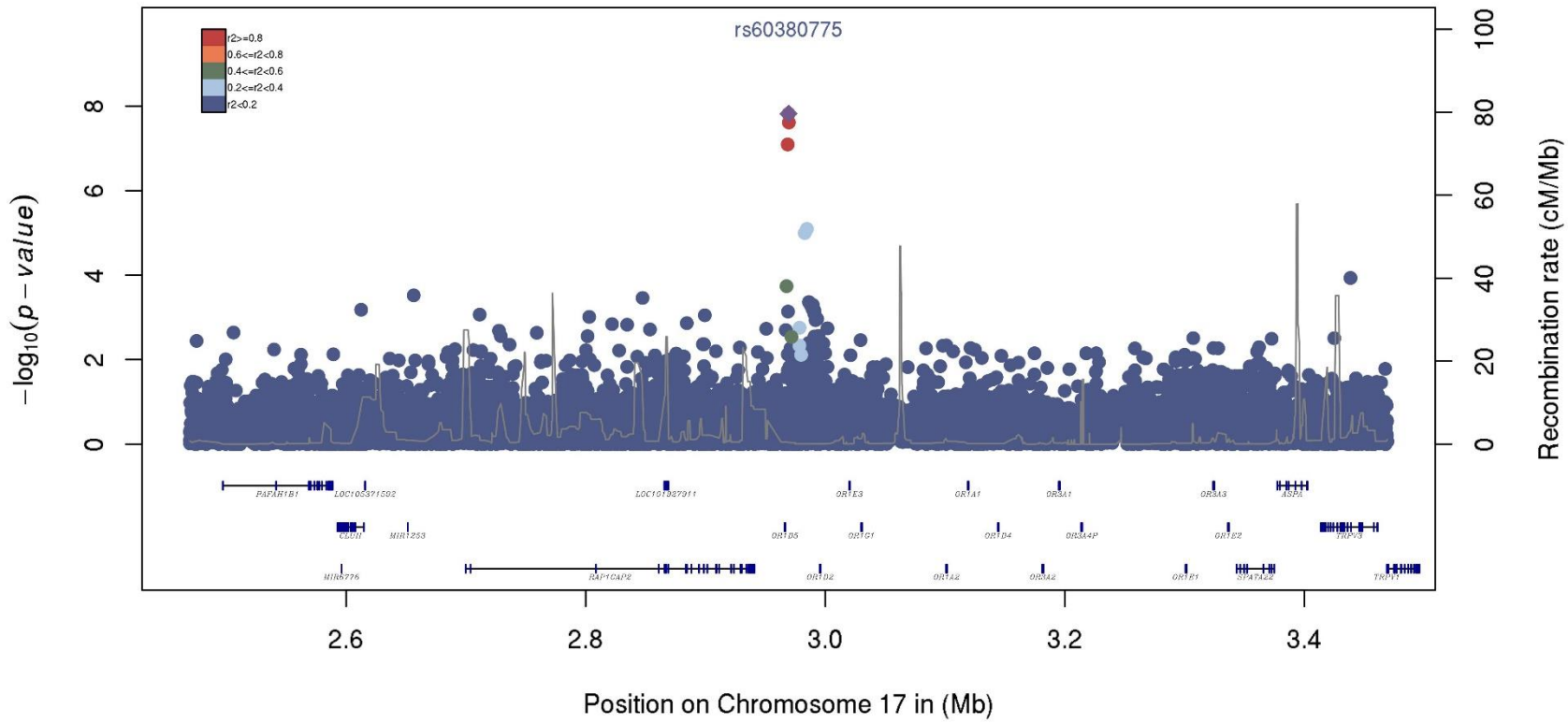
(B)


rs150022429
r2>=0.8
0.6<=r2<0.8
0.4<=r2<0.6
0.2<=r2<0.4
r2<0.2
$-\log_{10}(p-value)$
0
2
4
6
8
Recombination rate (cM/Mb)
0
20
40
60
80
100
69.0
69.2
69.4
69.6
69.8
Position on Chromosome 7 in (Mb)

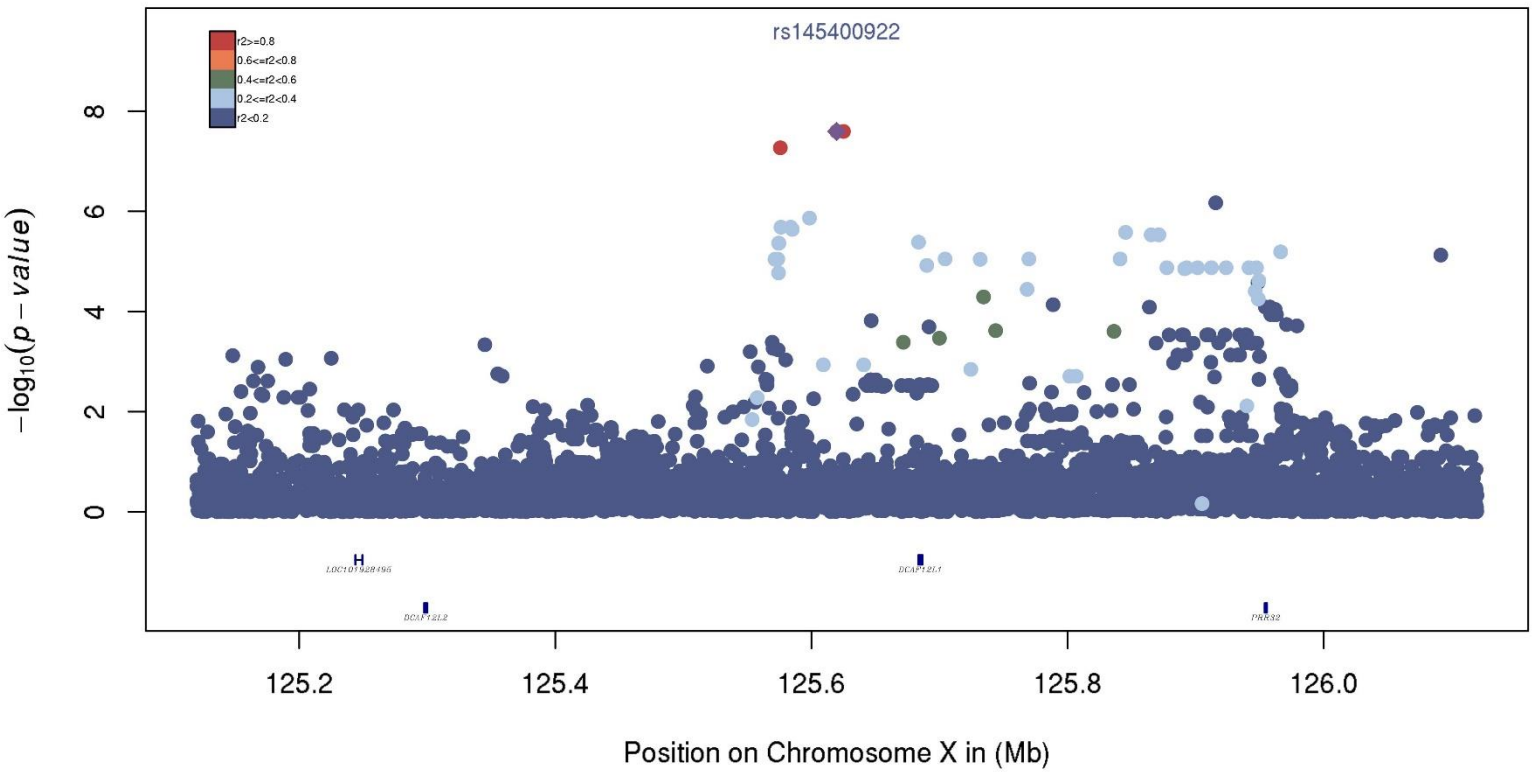
(C)


rs181401679
r2>=0.8
0.6<=r2<0.8
0.4<=r2<0.6
0.2<=r2<0.4
r2<0.2
−log10(p − value)
Recombination rate (cM/Mb)
0
2
4
6
8
10
0
20
40
60
80
100
103.8
104.0
104.2
104.4
104.6
Position on Chromosome 13 in (Mb)

(D)


rs60380775
r2>=0.8
0.6<=r2<0.8
0.4<=r2<0.6
0.2<=r2<0.4
r2<0.2
−log10(p − value)
0
2
4
6
8
Recombination rate (cM/Mb)
0
20
40
60
80
100
2.6
2.8
3.0
3.2
3.4
Position on Chromosome 17 in (Mb)

(E)


rs145400922
r2>=0.8
0.6<=r2<0.8
0.4<=r2<0.6
0.2<=r2<0.4
r2<0.2
−log10(p − value)
0
2
4
6
8
125.2
125.4
125.6
125.8
126.0
Position on Chromosome X in (Mb)

Supplemental Figure II. Functional annotation of the five novel loci using the USCS genome browser. The most significant variant at each locus was colored purple, and the LD proxies showing $r^2 \geq 0.8$, $0.8 > r^2 \geq 0.6$, and $0.6 > r^2 \geq 0.4$ were colored red, orange, and green, respectively. For *7q22*, *AUTS2*, *RAP1GAP2*, and *TEX13C* loci, LD information was extracted from African ancestry samples in TOPMed. For *13q33* locus, LD information was extracted from Hispanic ancestry samples in TOPMed. (A) *7q22*; (B) *AUTS2*; (C) *13q33*; (D) *RAP1GAP2*; (E) *TEX13C*.

A

Scale
chr7:
20 kb
hg38
340,000
345,000
350,000
355,000
360,000
365,000
370,000
375,000
380,000
385,000
390,000
Variants in LD with chr7:362834:G:A
chr7:353779:C:T
chr7:362834:G:A
chr7:372239:G:A
chr7:359366:A:ACCC
GENCODE v32 Comprehensive Transcript Set (only Basic displayed by default)
GENCODE v32
Aorta_Enhancer_H3K27ac
Aorta_Enhancer_H3K4me1
Brain_angular_gyrus_Enhancer_H3K27ac
Brain_angular_gyrus_Enhancer_H3K4me1
Brain_anterior_caudate_Enhancer_H3K27ac
Brain_anterior_caudate_Enhancer_H3K4me1
Brain_cingulate_gyrus_Enhancer_H3K27ac
Brain_cingulate_gyrus_Enhancer_H3K4me1
Brain_dorsolateral_prefrontal_cortex_Enhancer_H3K27ac
Brain_dorsolateral_prefrontal_cortex_Enhancer_H3K4me1
Brain_germinal_matrix_Enhancer_H3K27ac
Brain_germinal_matrix_Enhancer_H3K4me1
Brain_hippocampus_middle_Enhancer_H3K27ac
Brain_hippocampus_middle_Enhancer_H3K4me1
Brain_inferior_temporal_lobe_Enhancer_H3K27ac
Brain_inferior_temporal_lobe_Enhancer_H3K4me1
Brain_substantia_nigra_Enhancer_H3K27ac
Brain_substantia_nigra_Enhancer_H3K4me1
Left_Ventricle_Enhancer_H3K27ac
Left_Ventricle_Enhancer_H3K4me1
Right_Atrium_Enhancer_H3K27ac
Right_Atrium_Enhancer_H3K4me1
Right_Ventricle_Enhancer_H3K27ac
Right_Ventricle_Enhancer_H3K4me1
Aorta_Repressor_H3K27me3
Brain_angular_gyrus_Repressor_H3K27me3
Brain_anterior_caudate_Repressor_H3K27me3
Brain_cingulate_gyrus_Repressor_H3K27me3
Brain_dorsolateral_prefrontal_cortex_Repressor_H3K27me3
Brain_germinal_matrix_Repressor_H3K27me3
Brain_hippocampus_middle_Repressor_H3K27me3
Brain_inferior_temporal_lobe_Repressor_H3K27me3
Brain_substantia_nigra_Repressor_H3K27me3
Left_Ventricle_Repressor_H3K27me3
Right_Atrium_Repressor_H3K27me3
Right_Ventricle_Repressor_H3K27me3
Aorta_Transcriptn_H3K36me3
Brain_angular_gyrus_Transcriptn_H3K36me3
Brain_anterior_caudate_Transcriptn_H3K36me3
Brain_cingulate_gyrus_Transcriptn_H3K36me3
Brain_dorsolateral_prefrontal_cortex_Transcriptn_H3K36me3
Brain_germinal_matrix_Transcriptn_H3K36me3
Brain_hippocampus_middle_Transcriptn_H3K36me3
Brain_inferior_temporal_lobe_Transcriptn_H3K36me3
Brain_substantia_nigra_Transcriptn_H3K36me3
Left_Ventricle_Transcriptn_H3K36me3
Right_Atrium_Transcriptn_H3K36me3
Right_Ventricle_Transcriptn_H3K36me3
Aorta_DHS_DNase
Brain_angular_gyrus_DHS_DNase
Brain_anterior_caudate_DHS_DNase
Brain_cingulate_gyrus_DHS_DNase
Brain_dorsolateral_prefrontal_cortex_DHS_DNase
Brain_germinal_matrix_DHS_DNase
Brain_hippocampus_middle_DHS_DNase
Brain_inferior_temporal_lobe_DHS_DNase
Brain_substantia_nigra_DHS_DNase
Left_Ventricle_DHS_DNase
Right_Atrium_DHS_DNase
Right_Ventricle_DHS_DNase

B

Scale
chr7:
100 kb
hg38
69,250,000
69,300,000
69,350,000
69,400,000
69,450,000
69,500,000
69,550,000
Variants in LD with chr7:69400857:T:C
chr7:69329265:T:G
chr7:69390418:G:A
chr7:69438654:T:C
chr7:69352492:T:A
chr7:69400857:T:C
chr7:69370846:G:A
chr7:69433886:A:G
chr7:69401852:A:T
chr7:69408292:TC:T
chr7:69442037:T:C
GENCODE v32 Comprehensive Transcript Set (only Basic displayed by default)
AC092100.1
AC092100.1
AC092100.1
AC092100.1
RNU6-229P
Y RNA
Aorta_Enhancer_H3K27ac
Aorta_Enhancer_H3K4me1
Brain_angular_gyrus_Enhancer_H3K27ac
Brain_angular_gyrus_Enhancer_H3K4me1
Brain_anterior_caudate_Enhancer_H3K27ac
Brain_anterior_caudate_Enhancer_H3K4me1
Brain_cingulate_gyrus_Enhancer_H3K27ac
Brain_cingulate_gyrus_Enhancer_H3K4me1
Brain_dorsolateral_prefrontal_cortex_Enhancer_H3K27ac
Brain_dorsolateral_prefrontal_cortex_Enhancer_H3K4me1
Brain_germinal_matrix_Enhancer_H3K27ac
Brain_germinal_matrix_Enhancer_H3K4me1
Brain_hippocampus_middle_Enhancer_H3K27ac
Brain_hippocampus_middle_Enhancer_H3K4me1
Brain_inferior_temporal_lobe_Enhancer_H3K27ac
Brain_inferior_temporal_lobe_Enhancer_H3K4me1
Brain_substantia_nigra_Enhancer_H3K27ac
Brain_substantia_nigra_Enhancer_H3K4me1
Left_Ventricle_Enhancer_H3K27ac
Left_Ventricle_Enhancer_H3K4me1
Right_Atrium_Enhancer_H3K27ac
Right_Atrium_Enhancer_H3K4me1
Right_Ventricle_Enhancer_H3K27ac
Right_Ventricle_Enhancer_H3K4me1
Aorta_Repressor_H3K27me3
Brain_angular_gyrus_Repressor_H3K27me3
Brain_anterior_caudate_Repressor_H3K27me3
Brain_cingulate_gyrus_Repressor_H3K27me3
Brain_dorsolateral_prefrontal_cortex_Repressor_H3K27me3
Brain_germinal_matrix_Repressor_H3K27me3
Brain_hippocampus_middle_Repressor_H3K27me3
Brain_inferior_temporal_lobe_Repressor_H3K27me3
Brain_substantia_nigra_Repressor_H3K27me3
Left_Ventricle_Repressor_H3K27me3
Right_Atrium_Repressor_H3K27me3
Right_Ventricle_Repressor_H3K27me3
Aorta_Transcriptn_H3K36me3
Brain_angular_gyrus_Transcriptn_H3K36me3
Brain_anterior_caudate_Transcriptn_H3K36me3
Brain_cingulate_gyrus_Transcriptn_H3K36me3
Brain_dorsolateral_prefrontal_cortex_Transcriptn_H3K36me3
Brain_germinal_matrix_Transcriptn_H3K36me3
Brain_hippocampus_middle_Transcriptn_H3K36me3
Brain_inferior_temporal_lobe_Transcriptn_H3K36me3
Brain_substantia_nigra_Transcriptn_H3K36me3
Left_Ventricle_Transcriptn_H3K36me3
Right_Atrium_Transcriptn_H3K36me3
Right_Ventricle_Transcriptn_H3K36me3
Aorta_DHS_DNase
Brain_angular_gyrus_DHS_DNase
Brain_anterior_caudate_DHS_DNase
Brain_cingulate_gyrus_DHS_DNase
Brain_dorsolateral_prefrontal_cortex_DHS_DNase
Brain_germinal_matrix_DHS_DNase
Brain_hippocampus_middle_DHS_DNase
Brain_inferior_temporal_lobe_DHS_DNase
Brain_substantia_nigra_DHS_DNase
Left_Ventricle_DHS_DNase
Right_Atrium_DHS_DNase
Right_Ventricle_DHS_DNase

C

Scale
chr13:
100 kb
hg38
104,100,000
104,150,000
104,200,000
104,250,000
104,300,000
104,350,000
104,400,000
104,450,000
104,500,000
Variants in LD with chr13:104244508:C:G
chr13:104201610:A:G
chr13:104328475:G:A
chr13:104208771:A:C
chr13:104351294:C:T
chr13:104212134:G:A
chr13:104216625:G:C
chr13:104217682:T:G
chr13:104218931:G:A
chr13:104219496:A:G
chr13:104244508:C:G
GENCODE v32 Comprehensive Transcript Set (only Basic displayed by default)
AL136524.1
Aorta_Enhancer_H3K27ac
Aorta_Enhancer_H3K4me1
Brain_angular_gyrus_Enhancer_H3K27ac
Brain_angular_gyrus_Enhancer_H3K4me1
Brain_anterior_caudate_Enhancer_H3K27ac
Brain_anterior_caudate_Enhancer_H3K4me1
Brain_cingulate_gyrus_Enhancer_H3K27ac
Brain_cingulate_gyrus_Enhancer_H3K4me1
Brain_dorsolateral_prefrontal_cortex_Enhancer_H3K27ac
Brain_dorsolateral_prefrontal_cortex_Enhancer_H3K4me1
Brain_germinal_matrix_Enhancer_H3K27ac
Brain_germinal_matrix_Enhancer_H3K4me1
Brain_hippocampus_middle_Enhancer_H3K27ac
Brain_hippocampus_middle_Enhancer_H3K4me1
Brain_inferior_temporal_lobe_Enhancer_H3K27ac
Brain_inferior_temporal_lobe_Enhancer_H3K4me1
Brain_substantia_nigra_Enhancer_H3K27ac
Brain_substantia_nigra_Enhancer_H3K4me1
Left_Ventricle_Enhancer_H3K27ac
Left_Ventricle_Enhancer_H3K4me1
Right_Atrium_Enhancer_H3K27ac
Right_Atrium_Enhancer_H3K4me1
Right_Ventricle_Enhancer_H3K27ac
Right_Ventricle_Enhancer_H3K4me1
Aorta_Repressor_H3K27me3
Brain_angular_gyrus_Repressor_H3K27me3
Brain_anterior_caudate_Repressor_H3K27me3
Brain_cingulate_gyrus_Repressor_H3K27me3
Brain_dorsolateral_prefrontal_cortex_Repressor_H3K27me3
Brain_germinal_matrix_Repressor_H3K27me3
Brain_hippocampus_middle_Repressor_H3K27me3
Brain_inferior_temporal_lobe_Repressor_H3K27me3
Brain_substantia_nigra_Repressor_H3K27me3
Left_Ventricle_Repressor_H3K27me3
Right_Atrium_Repressor_H3K27me3
Right_Ventricle_Repressor_H3K27me3
Aorta_Transcriptn_H3K36me3
Brain_angular_gyrus_Transcriptn_H3K36me3
Brain_anterior_caudate_Transcriptn_H3K36me3
Brain_cingulate_gyrus_Transcriptn_H3K36me3
Brain_dorsolateral_prefrontal_cortex_Transcriptn_H3K36me3
Brain_germinal_matrix_Transcriptn_H3K36me3
Brain_hippocampus_middle_Transcriptn_H3K36me3
Brain_inferior_temporal_lobe_Transcriptn_H3K36me3
Brain_substantia_nigra_Transcriptn_H3K36me3
Left_Ventricle_Transcriptn_H3K36me3
Right_Atrium_Transcriptn_H3K36me3
Right_Ventricle_Transcriptn_H3K36me3
Aorta_DHS_DNase
Brain_angular_gyrus_DHS_DNase
Brain_anterior_caudate_DHS_DNase
Brain_cingulate_gyrus_DHS_DNase
Brain_dorsolateral_prefrontal_cortex_DHS_DNase
Brain_germinal_matrix_DHS_DNase
Brain_hippocampus_middle_DHS_DNase
Brain_inferior_temporal_lobe_DHS_DNase
Brain_substantia_nigra_DHS_DNase
Left_Ventricle_DHS_DNase
Right_Atrium_DHS_DNase
Right_Ventricle_DHS_DNase

D

Scale
chr17:
5 kb
hg38
2,965,000
2,970,000
2,975,000
Variants in LD with chr17:2969421:C:T
chr17:2967723:A:C
chr17:2971820:C:T
chr17:2968677:A:G
chr17:2969421:C:T
chr17:2969648:C:T
GENCODE v32 Comprehensive Transcript Set (only Basic displayed by default)
RAP1GAP2
RAP1GAP2
RAP1GAP2
RAP1GAP2
RAP1GAP2
AC015921.1
Aorta_Enhancer_H3K27ac
Aorta_Enhancer_H3K4me1
Brain_angular_gyrus_Enhancer_H3K27ac
Brain_angular_gyrus_Enhancer_H3K4me1
Brain_anterior_caudate_Enhancer_H3K27ac
Brain_anterior_caudate_Enhancer_H3K4me1
Brain_cingulate_gyrus_Enhancer_H3K27ac
Brain_cingulate_gyrus_Enhancer_H3K4me1
Brain_dorsolateral_prefrontal_cortex_Enhancer_H3K27ac
Brain_dorsolateral_prefrontal_cortex_Enhancer_H3K4me1
Brain_germinal_matrix_Enhancer_H3K27ac
Brain_germinal_matrix_Enhancer_H3K4me1
Brain_hippocampus_middle_Enhancer_H3K27ac
Brain_hippocampus_middle_Enhancer_H3K4me1
Brain_inferior_temporal_lobe_Enhancer_H3K27ac
Brain_inferior_temporal_lobe_Enhancer_H3K4me1
Brain_substantia_nigra_Enhancer_H3K27ac
Brain_substantia_nigra_Enhancer_H3K4me1
Left_Ventricle_Enhancer_H3K27ac
Left_Ventricle_Enhancer_H3K4me1
Right_Atrium_Enhancer_H3K27ac
Right_Atrium_Enhancer_H3K4me1
Right_Ventricle_Enhancer_H3K27ac
Right_Ventricle_Enhancer_H3K4me1
Aorta_Repressor_H3K27me3
Brain_angular_gyrus_Repressor_H3K27me3
Brain_anterior_caudate_Repressor_H3K27me3
Brain_cingulate_gyrus_Repressor_H3K27me3
Brain_dorsolateral_prefrontal_cortex_Repressor_H3K27me3
Brain_germinal_matrix_Repressor_H3K27me3
Brain_hippocampus_middle_Repressor_H3K27me3
Brain_inferior_temporal_lobe_Repressor_H3K27me3
Brain_substantia_nigra_Repressor_H3K27me3
Left_Ventricle_Repressor_H3K27me3
Right_Atrium_Repressor_H3K27me3
Right_Ventricle_Repressor_H3K27me3
Aorta_Transcriptn_H3K36me3
Brain_angular_gyrus_Transcriptn_H3K36me3
Brain_anterior_caudate_Transcriptn_H3K36me3
Brain_cingulate_gyrus_Transcriptn_H3K36me3
Brain_dorsolateral_prefrontal_cortex_Transcriptn_H3K36me3
Brain_germinal_matrix_Transcriptn_H3K36me3
Brain_hippocampus_middle_Transcriptn_H3K36me3
Brain_inferior_temporal_lobe_Transcriptn_H3K36me3
Brain_substantia_nigra_Transcriptn_H3K36me3
Left_Ventricle_Transcriptn_H3K36me3
Right_Atrium_Transcriptn_H3K36me3
Right_Ventricle_Transcriptn_H3K36me3
Aorta_DHS_DNase
Brain_angular_gyrus_DHS_DNase
Brain_anterior_caudate_DHS_DNase
Brain_cingulate_gyrus_DHS_DNase
Brain_dorsolateral_prefrontal_cortex_DHS_DNase
Brain_germinal_matrix_DHS_DNase
Brain_hippocampus_middle_DHS_DNase
Brain_inferior_temporal_lobe_DHS_DNase
Brain_substantia_nigra_DHS_DNase
Left_Ventricle_DHS_DNase
Right_Atrium_DHS_DNase
Right_Ventricle_DHS_DNase

E

Scale
chrX: 125,400,000 125,500,000 125,600,000 125,700,000 125,800,000 125,900,000 126,000,000
200 kb hg38
Variants in LD with chrX:125619640:A:G
chrX:125575676:G:A chrX:125699914:G:T chrX:125836182:G:T
chrX:125619640:A:G chrX:125743812:C:T
chrX:125624940:C:G
chrX:125671717:T:C
chrX:125734344:T:C
GENCODE v32 Comprehensive Transcript Set (only Basic displayed by default)
TEX13C AL645820.1

Aorta_Enhancer_H3K27ac
Aorta_Enhancer_H3K4me1
Brain_angular_gyrus_Enhancer_H3K27ac
Brain_angular_gyrus_Enhancer_H3K4me1
Brain_anterior_caudate_Enhancer_H3K27ac
Brain_anterior_caudate_Enhancer_H3K4me1
Brain_cingulate_gyrus_Enhancer_H3K27ac
Brain_cingulate_gyrus_Enhancer_H3K4me1
Brain_dorsolateral_prefrontal_cortex_Enhancer_H3K27ac
Brain_dorsolateral_prefrontal_cortex_Enhancer_H3K4me1
Brain_germinal_matrix_Enhancer_H3K27ac
Brain_germinal_matrix_Enhancer_H3K4me1
Brain_hippocampus_middle_Enhancer_H3K27ac
Brain_hippocampus_middle_Enhancer_H3K4me1
Brain_inferior_temporal_lobe_Enhancer_H3K27ac
Brain_inferior_temporal_lobe_Enhancer_H3K4me1
Brain_substantia_nigra_Enhancer_H3K27ac
Brain_substantia_nigra_Enhancer_H3K4me1
Left_Ventricle_Enhancer_H3K27ac
Left_Ventricle_Enhancer_H3K4me1
Right_Atrium_Enhancer_H3K27ac
Right_Atrium_Enhancer_H3K4me1
Right_Ventricle_Enhancer_H3K27ac
Right_Ventricle_Enhancer_H3K4me1
Aorta_Repressor_H3K27me3
Brain_angular_gyrus_Repressor_H3K27me3
Brain_anterior_caudate_Repressor_H3K27me3
Brain_cingulate_gyrus_Repressor_H3K27me3
Brain_dorsolateral_prefrontal_cortex_Repressor_H3K27me3
Brain_germinal_matrix_Repressor_H3K27me3
Brain_hippocampus_middle_Repressor_H3K27me3
Brain_inferior_temporal_lobe_Repressor_H3K27me3
Brain_substantia_nigra_Repressor_H3K27me3
Left_Ventricle_Repressor_H3K27me3
Right_Atrium_Repressor_H3K27me3
Right_Ventricle_Repressor_H3K27me3
Aorta_Transcriptn_H3K36me3
Brain_angular_gyrus_Transcriptn_H3K36me3
Brain_anterior_caudate_Transcriptn_H3K36me3
Brain_cingulate_gyrus_Transcriptn_H3K36me3
Brain_dorsolateral_prefrontal_cortex_Transcriptn_H3K36me3
Brain_germinal_matrix_Transcriptn_H3K36me3
Brain_hippocampus_middle_Transcriptn_H3K36me3
Brain_inferior_temporal_lobe_Transcriptn_H3K36me3
Brain_substantia_nigra_Transcriptn_H3K36me3
Left_Ventricle_Transcriptn_H3K36me3
Right_Atrium_Transcriptn_H3K36me3
Right_Ventricle_Transcriptn_H3K36me3
Aorta_DHS_DNase
Brain_angular_gyrus_DHS_DNase
Brain_anterior_caudate_DHS_DNase
Brain_cingulate_gyrus_DHS_DNase
Brain_dorsolateral_prefrontal_cortex_DHS_DNase
Brain_germinal_matrix_DHS_DNase
Brain_hippocampus_middle_DHS_DNase
Brain_inferior_temporal_lobe_DHS_DNase
Brain_substantia_nigra_DHS_DNase
Left_Ventricle_DHS_DNase
Right_Atrium_DHS_DNase
Right_Ventricle_DHS_DNase

Reporting checklist for genetic association study.

Based on the STREGA guidelines.

Instructions to authors

Complete this checklist by entering the page numbers from your manuscript where readers will find each of the items listed below.

Your article may not currently address all the items on the checklist. Please modify your text to include the missing information. If you are certain that an item does not apply, please write "n/a" and provide a short explanation.

Upload your completed checklist as an extra file when you submit to a journal.

In your methods section, say that you used the STREGAreporting guidelines, and cite them as:

Little J, Higgins JP, Ioannidis JP, Moher D, Gagnon F, von Elm E, Khoury MJ, Cohen B, Davey-Smith G, Grimshaw J, Scheet P, Gwinn M, Williamson RE, Zou GY, Hutchings K, Johnson CY, Tait V, Wiens M, Golding J, van Duijn C, McLaughlin J, Paterson A, Wells G, Fortier I, Freedman M, Zecevic M, King R, Infante-Rivard C, Stewart A, Birkett N; STrengthening the REporting of Genetic Association Studies. STrengthening the REporting of Genetic Association Studies (STREGA): An Extension of the STROBE Statement.

		Reporting Item	Page Number
Title and abstract			
Title	#1a	Indicate the study's design with a commonly used term in the title or the abstract	1
Abstract	#1b	Provide in the abstract an informative and balanced summary of what was done and what was found	4-5
Background/rationale			
	#2	Explain the scientific background and rationale for the investigation being reported	8
Objectives			
	#3	State specific objectives, including any prespecified hypotheses. State if the study is the first report of a	8

		genetic association, a replication effort, or both.	
Study design			
	#4	Present key elements of study design early in the paper	9
Setting			
	#5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	9
Eligibility criteria			
	#6a	Cohort study – Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up. Case-control study – Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls. Cross-sectional study – Give the eligibility criteria, and the sources and methods of selection of participants. Give information on the criteria and methods for selection of subsets of participants from a larger study, when relevant.	9
	#6b	Cohort study – For matched studies, give matching criteria and number of exposed and unexposed. Case-control study – For matched studies, give matching criteria and the number of controls per case.	NA
Variables			
	#7a	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	9-12
	#7b	Clearly define genetic exposures (genetic variants) using a widely-used nomenclature system. Identify variables likely to be associated with population stratification (confounding by ethnic origin).	9-12
Data sources/measurement			

	#8a	For each variable of interest give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group. Give information separately for for exposed and unexposed groups if applicable.	9-11
	#8b	Describe laboratory methods, including source and storage of DNA, genotyping methods and platforms (including the allele calling algorithm used, and its version), error rates and call rates. State the laboratory / centre where genotyping was done. Describe comparability of laboratory methods if there is more than one group. Specify whether genotypes were assigned using all of the data from the study simultaneously or in smaller batches.	10
Bias			
	#9a	Describe any efforts to address potential sources of bias	9
	#9b	Describe any efforts to address potential sources of bias	NA
Study size			
	#10	Explain how the study size was arrived at	9
Quantitative variables			
	#11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen, and why. If applicable, describe how effects of treatment were dealt with.	10-12
Statistical methods			
	#12a	Describe all statistical methods, including those used to control for confounding. State software version used and options (or settings) chosen.	10-12
	#12b	Describe any methods used to examine subgroups and interactions	10

#12c	Explain how missing data were addressed	NA
#12d	If applicable, explain how loss to follow-up was addressed	NA
#12e	Describe any sensitivity analyses	NA
#12f	State whether Hardy-Weinberg equilibrium was considered and, if so, how.	10
#12g	Describe any methods used for inferring genotypes or haplotypes	10
#12h	Describe any methods used to assess or address population stratification.	10-11
#12i	Describe any methods used to address multiple comparisons or to control risk of false positive findings.	11
#12j	Describe any methods used to address and correct for relatedness among subjects	10-11
Participants		
#13a	Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed. Give information separately for for exposed and unexposed groups if applicable. Report numbers of individuals in whom genotyping was attempted and numbers of individuals in whom genotyping was successful.	12-13
#13b	Give reasons for non-participation at each stage	NA
#13c	Consider use of a flow diagram	34
Descriptive data		
#14a	Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders. Give information separately for exposed and unexposed groups if applicable. Consider giving information by genotype	12-13

	#14b	Indicate number of participants with missing data for each variable of interest	NA
	#14c	Cohort study – Summarize follow-up time, e.g. average and total amount.	NA
Outcome data			
	#15	Cohort study Report numbers of outcome events or summary measures over time.Give information separately for exposed and unexposed groups if applicable. Report outcomes (phenotypes) for each genotype category over time Case-control study – Report numbers in each exposure category, or summary measures of exposure.Give information separately for cases and controls . Report numbers in each genotype category. Cross-sectional study – Report numbers of outcome events or summary measures. Give information separately for exposed and unexposed groups if applicable. Report outcomes (phenotypes) for each genotype category	31
Main results			
	#16a	Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	13,14,33
	#16b	Report category boundaries when continuous variables were categorized	NA
	#16c	If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	NA
	#16d	Report results of any adjustments for multiple comparisons	13-15
Other analyses			
	#17a	Report other analyses done—e.g., analyses of subgroups and interactions, and sensitivity analyses	13-14

	#17b	Report other analyses done—e.g., analyses of subgroups and interactions, and sensitivity analyses	NA
	#17c	Report other analyses done—e.g., analyses of subgroups and interactions, and sensitivity analyses	NA
Key results			
	#18	Summarise key results with reference to study objectives	16
Limitations			
	#19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias.	18
Interpretation			
	#20	Give a cautious overall interpretation considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence.	18-19
Generalisability			
	#21	Discuss the generalisability (external validity) of the study results	19
Funding			
	#22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	19-22