

SUPPLEMENTAL MATERIAL

Cheng et al. Genome-wide association analysis of young onset stroke identifies a locus on chromosome
10q25 near *HABP2*

Table of Contents

1. Details of the Discovery and Follow-up Stages for GWAS	3
2. Study-specific GWAS of the Discovery Cohorts	3
3. Testing for association of index SNPs with FSAP levels in the SAHLSIS	4
4. Description of Stage 1 and Stage 2 Studies	4
5. Studies in METASTROKE consortium	8
6. The Sahlgrenska Study of Ischemic Stroke (SAHLSIS)	12
7. Supplemental Tables	
Supplemental Table I. Study design and characteristics of participating studies in the discovery phase	14
Supplemental Table II. Study design and characteristics of participating studies in the validation phase	16
Supplemental Table III. Genotyping and analysis methods used by each participating studies in discovery phase	17
Supplemental Table IV. Genotyping and analysis methods used in the validation studies	19
Supplemental Table V. Study design and characteristics of METASTROKE cohorts	20
Supplemental Table VI. SNPs with suggestive associations with early-onset ischemic stroke and their proxy SNPs in chr10 region	21
Supplemental Table VII. Top associated SNPs for early onset stroke	22
Supplemental Table VIII. Top associated SNPs in GWAS analysis of stroke subtypes	23
Supplemental Table IX. SNPs previously associated with ischemic stroke (and/or ischemic stroke subtypes) and their associations with early-onset stroke	24
8. Supplemental Figures	
Supplemental Figure I. QQ plot of (A) European-only meta-analysis and (B) trans- ethnic meta-analysis	27
Supplemental Figure II. European-ancestry GWAS results of stroke subtypes (A) cardioembolic, (B) large artery atherosclerotic, (C) small vessel stroke and (D) undetermined	28
Supplemental Figure III. Trans-ethnic GWAS results of stroke subtypes (A) cardioembolic, (B) large artery atherosclerotic, (C) small vessel stroke and (D) undetermined	31
9. Supplemental References	36

Details of the Discovery and Follow-up Stages for GWAS

Discovery GWAS Stage (Stage 1). The Discovery stage for the GWAS meta-analysis included a total of 4,505 early-onset cases and 21,968 controls from 6 studies: Cervical Artery Dissections and Ischemic Stroke Patients study (CADISP), Genetics of Early-onset Stroke study (GEOS), Milano study (MILANO), Stroke in Young Fabry Patients (SIFAP), Risk Assessment of Cerebrovascular Events Study (RACE) and UK samples from Wellcome Trust Case Control Consortium 2 (WTCCC2-UK). Four of these studies (CADISP, MILANO, SIFAP, WTCCC2-UK) included individuals of European ancestry only, while RACE included individuals of South-Asian ancestry from Pakistan and GEOS, which recruited participants from the Greater-Baltimore Washington Area, USA, included individuals of both European ancestry and African Americans. All studies used a case-control study design. External controls (genotyped separately from the cases) were used for three studies: SIFAP (controls from Cooperative Health Research in the Augsburg Region ["Kooperative Gesundheitsforschung in der Region Augsburg"], KORA), RACE (controls from the Pakistan Risk of Myocardial Infarction Study [PROMIS]), and WTCCC2-UK (controls from the 1958 British Birth cohort and the UK Blood Service control group). RACE cases were genotyped in two phases (RACE1 and RACE2) and thus were analyzed separately and then meta-analyzed together to provide RACE study-specific estimates. Study-specific subject inclusion/exclusion criteria are provided in the Supplementary Data, and summary of study characteristics, including their sample size, are listed in Table 1 and Supplemental Table I.

Follow-up Stage (Stage 2). The Follow-up stage included a total of 1,003 independent, early-onset cases and 7,745 controls from the Australia Stroke Genetics Collaborative (ASGC), Munich samples from WTCCC2 (Munich) and the WTCCC ImmunoChip Study (ImmunoChip). All three studies were of European ancestry and used the case-control study design, and subjects included in the Follow-up stage were independent of those included in the Discovery. The ASGC and Munich studies were genotyped using whole-genome array while WTCCC ImmunoChip Study were genotyped using Illumina ImmunoChip.

Ischemic stroke was defined according to predefined study-specific criteria, generally as a focal neurological deficit lasting greater than 24 hours or, if lasting less than 24 hours, with neuroimaging evidence. The majority of cases were first-ever strokes. Stroke subtypes were defined according to the Trial of Org 10172 in Acute Stroke Treatment (TOAST) criteria.¹

Study-specific GWAS of the Discovery Cohorts

The 6 Discovery cohorts included 3 studies with internal controls (GEOS, CADISP, and MILANO) and 3 studies with external controls (SIFAP, RACE, and WTCCC2-UK), the latter including external controls by design in order to maximize the number of cases that could be genotyped. All studies were genotyped using whole-genome SNP arrays from Illumina (San Diego, CA, USA). Extensive quality control was carried out at each site at both the sample and SNP level. Study samples were removed from analyses if they had low genotyping call rates, exhibited mismatches between genetic and reported genders or relatedness, or were outliers from the population on the basis of principal component analysis of SNPs. SNPs were excluded prior to imputation if they had high missing call rates (missing rate >5%), were monomorphic, or deviated from Hardy-Weinberg equilibrium (HWE $p < 0.0001$ for most studies). Detailed information about the genotyping platforms, genotype cleaning

procedures, imputation process, and association analysis modeling is described in Supplemental Tables III (Discovery phase) and IV (Follow-up phase).

Additional SNP level QC was performed for the studies with external controls to ensure that artifactual differences in allele frequencies between cases and controls did not arise due to differences in genotyping performance between genotyping centers. Within each study, cases and controls were merged for a second round of SNP cleaning to remove SNPs with differential missingness between cases and controls and to check for cryptic relatedness. In the GEOS, SIFAPS, and RACE I studies, genotyping concordance rates between cases and controls were evaluated by re-genotyping ~30 control samples that had been previously genotyped along with the cases, after which SNPs with ≥ 1 discordant genotype in a sample were removed from analysis. Principal components were calculated within the merged cases and control groups and outlier samples were removed from analysis. After QC, a genome-wide association study (GWAS) was performed within each study to check genomic inflation and ensure that QC had successfully removed false associations potentially resulting from the study design. Lambda values across the studies revealed no evidence for genomic inflation (lambdas: 0.98, 0.98, 1.05, 1.01, and 1.02 for GEOS-EUR, GEOS-AFR, SIFAP, RACE I, and WTCCC2-UK, respectively).

Testing for association of index SNPs with FSAP levels in the SAHLSIS

SAHLSIS participants had previously been genotyped on the Illumina Human OmniExpress 750K array at the Broad Institute at Harvard Medical School (Boston, USA). This array included two SNPs (rs7906302 and rs1338423) that were in high linkage disequilibrium (LD) with the young onset stroke-associated SNP ($r^2=0.88$; $D'=1$ for both SNPs with the index SNP). These two SNPs were in high LD with each other ($r^2 = 0.77$). Because the number of subjects with two copies of the minor allele of rs7906302 was small (2 individuals), the minor allele was treated as absent (AA) versus present (AC + CC). For rs1338423, no individuals had two copies of the minor allele (G) and comparisons were therefore made between the AA and AG genotypes.

Description of Stage 1 and Stage 2 Studies

Australian Stroke Genetics Collaborative (ASGC)

ASGC stroke cases comprised stroke patients of European ancestry who were admitted to four clinical centers across Australia (The Neurosciences Department at Gosford Hospital, Gosford; the Neurology Department at John Hunter Hospital, Newcastle; The Queen Elizabeth Hospital, Adelaide; and the Royal Perth Hospital, Perth) between 2003 and 2008. Stroke was defined by World Health Organization criteria as a sudden focal neurological deficit of vascular origin, lasting more than 24 h and confirmed by imaging, such as computerized tomography (CT) and/or magnetic resonance imaging (MRI) brain scan. Other investigative tests such as electrocardiogram, carotid Doppler and trans-esophageal echocardiogram were conducted to define ischemic stroke mechanism as clinically appropriate. Cases were excluded from participation if they were aged <18 years, were diagnosed with hemorrhagic stroke or had transient ischemic attack rather than ischemic stroke or if they were unable to undergo baseline brain imaging. On the basis of these criteria, a total of 1,230 ischemic stroke cases were included in the current study. Ischemic stroke subtypes were assigned using TOAST criteria on the basis of clinical, imaging and risk factor data.

ASGC controls were participants in the Hunter Community Study (HCS), a population-based cohort of individuals aged 55–85 years, predominantly of European ancestry and residing in the Hunter Region in New South Wales, Australia. Participants were randomly selected from the New South Wales State electoral roll and were contacted by mail between 2004 and 2007. Consenting participants completed five detailed self-report questionnaires and attended the HCS data collection center, at which time a series of clinical measures were obtained. A total of 1,280 HCS participants were genotyped for the current study.

All study participants gave informed consent for participation in genetic studies. Approval for the individual studies was obtained from the relevant institutional ethics committees.

Cervical Artery Dissection and Ischemic Stroke Patients (CADISP)

The CADISP Study aims primarily at identifying genetic and environmental risk factors for cervical artery dissection (CeAD), a major cause of ischemic stroke in young adults. We included 942 CeAD patients in the CADISP study in 2004-2009 (CADISP-1; 170 Finns and 772 non-Finnish Europeans). Additional 451 CeAD patients of European origin were recruited in 2008-2010, exclusively for the CADISP-genetics project, in some CADISP centers and additional European and US centers (CADISP-2). To assess the specificity of risk factors for CeAD, we also recruited 583 patients with an IS attributable to other causes (non-CeAD IS, 162 Finns and 421 non-Finnish Europeans from Belgium, France, Germany, Italy, and Switzerland), frequency-matched on age, sex and geographical origin with CADISP-1 CeAD patients. Of these, a total of 555 non-CeAD IS patients aged < 60 years, who were successfully genotyped and met genotyping quality control criteria, were available for the present analysis. The abstracted hospital records of cases were reviewed and adjudicated for IS subtype by a neurologist in each participating center. Each item required for the subtype classification was also recorded in a standardized fashion. Based on this, IS subtypes were then centrally readjudicated by a panel of neurologists, in agreement with the TOAST system¹, using the more detailed subtype description from Ay et al.² DNA samples were genotyped on an Illumina Human610-Quad or Human660W-Quad BeadChip® at the Centre National de Génotypage [CNG], Evry, France. The majority of controls (N=9,046, of which 74 Finns and 8,972 non-Finnish Europeans) were selected from an anonymized control genotype database at the CNG, in order to match cases for ethnic background, based on principal component analysis. European reference samples from the genotype repository at the CNG were also analyzed simultaneously to provide improved geographical resolution. Additional Finnish controls were recruited within the CADISP study, both from the general population and among spouses and unrelated friends of CADISP patients, within the Helsinki area. A total of 234 individuals were eligible for genotyping at the CNG. Of these, 213 individuals who were genotyped successfully and met quality control criteria were available for the present analysis. Individuals were excluded if they were unexpected duplicates, gender discrepancy and unexpected relatedness. After quality control, we performed genotype imputation to the nonmonomorphic SNPs described in the 1000 genomes phase I v3 reference panel.

Genetics of Early-onset Stroke Study (GEOS)

GEOS is a population-based case-control study designed to identify genes associated with early-onset ischemic stroke and to characterize interactions of identified stroke genes and/or SNPs with environmental risk factors.^{3;4} Participants were recruited from the greater Baltimore-Washington area in 4 different time periods: Stroke Prevention in Young Women-1 (SPYW-1) conducted from 1992-

1996, Stroke Prevention in Young Women-2 (SPYW-2) conducted from 2001-2003, Stroke Prevention in Young Men (SPYM) conducted from 2003-2007, and Stroke Prevention in Young Adults (SPYA) conducted in 2008. *Case participants* were hospitalized with a first cerebral infarction identified by discharge surveillance from one of the 59 hospitals in the greater Baltimore-Washington area and direct referral from regional neurologists. The abstracted hospital records of cases were reviewed and adjudicated for ischemic stroke subtype by a pair of neurologists according to previously published procedures with disagreements resolved by a third neurologist. The ischemic stroke subtype classification system retains information on all probable and possible causes, and is reducible to the more widely used TOAST system that assigns each case to a single category. *Control participants* without a history of stroke were identified by random-digit dialing and were balanced to cases by age and region of residence in each recruitment periods. Genomic DNA was isolated from a variety of sample types, including cell line, whole blood, mouth wash and buccal swab. Samples were genotyped at the Johns Hopkins Center for Inherited Disease Research (CIDR) using the Illumina HumanOmni1-Quad_v1-0_B BeadChip (Illumina, San Diego, CA, USA). Individuals were excluded if they were unexpected duplicates, gender discrepancy and unexpected relatedness. For this analysis, 5.1% of cases and 0.6% of controls that had a history of MI (based on self-report) were removed from analysis.

Milano

This study includes consecutive Italian patients referred to Besta Institute from 2000 to 2009 with stroke and included in the Besta Cerebrovascular Diseases Registry (CEDIR). Ischemic stroke cases, first ever or recurrent, confirmed on brain imaging, were selected for this study. All cases were of self-reported Caucasian ancestry and had clinically relevant diagnostic workup performed. All cases were phenotyped by an experienced stroke neurologist according to TOAST criteria, based on relevant clinical imaging and available information on cardiovascular risk factors. Controls are Italian individuals enrolled within the PROCARDIS Study, with no personal or sibling history of coronary heart disease before age 66 years.

Stroke in Young Fabry Patients (SIFAP)

SIFAP study is a multicenter study carried out to determine the frequency of Fabry disease in an unselected group of young patients (18 – 55 years of age) with acute cerebrovascular events (CVE) defined as having had an acute ischemic stroke or transient ischemic attack less than 3 months before enrollment into the study. For this project, about ischemic stroke cases recruited from 15 sites throughout Germany and determined not to have Fabry Disease were included in the analysis. All were of European ancestry and had age of first stroke of 18-55 years. The diagnosis of Fabry disease was based in males as well as in females in the first level on the sequencing data of the entire exon structure including promoter of the α -galactosidase gene. In cases where a mutation was detected biochemical analysis was done. Stroke cases from SIFAP were genotyped at CIDR using the Illumina Human Omni 2.5M-Quad array and controls free of cardiovascular diseases were selected from the KORA Study previously genotyped at CIDR in the same platform. The Cooperative Health Research in the Region of Augsburg (KORA) study is a population-based study of cardiovascular and metabolic traits carried out in the region of Augsburg, Southern Germany. Extensive care will be taken in the data cleaning process to minimize artifactual differences in allele frequencies between cases and controls that may have occurred due to cases and controls being genotyped at different times (although on the same platform and in the same laboratory).

Risk Assessment of Cerebrovascular Events (RACE) Study, Pakistan

RACE is a retrospective case-control study designed to identify and evaluate genetic, lifestyle and biomarker determinants of stroke and its subtype in Pakistan. Samples were recruited from six hospital centers in Pakistan. Cases were eligible for inclusion in the study if they: (1) are aged at least 18 years; (2) presented with a sudden onset of neurological deficit affecting a vascular territory with sustained deficit at 24 hours verified by medical attention within 72 hours after onset (onset is defined by when the patient was last seen normal and not when found with deficit); (3) the diagnosis was supported by CT/MRI; and (4) presented with a Modified Rankin Score of < 2 prior to the stroke. TOAST and Oxfordshire classification systems were used to sub-phenotype all stroke cases. Control participants were individuals enrolled in the Pakistan Risk of Myocardial Infarction Study (PROMIS), a case-control study of acute MI based in Pakistan. Controls in PROMIS were recruited following procedures and inclusion criteria as adopted for RACE cases. In order to minimize any potential selection biases, PROMIS controls selected for this stroke study were frequency matched to RACE cases based on age and gender and were recruited in the following order of priority: (1) non-blood related or blood related visitors of patients of the out-patient department; (2) non-blood related visitors of stroke patients; (3) patients of the out-patient department presenting with minor complaints.

Wellcome Trust Case-Control Consortium 2 (WTCCC2)

The WTCCC2 samples were genotyped as part of the WTCCC 2 ischemic stroke study. Stroke cases included samples recruited by investigators at St. George's University London (SGUL) and University of Oxford in the UK and the Department of Neurology, Klinikum Großhadern, Ludwig-Maximilians-University, Munich.

The SGUL collection comprised 1224 ischemic stroke samples from a hospital based setting. All cases were of self-reported Caucasian ancestry. Ischemic stroke subtypes were determined according to TOAST criteria based on relevant clinical imaging and available information on cardiovascular risk factors. The University of Oxford collection comprised 896 ischemic stroke cases, consecutively collected as part of the Oxford vascular study (OXVASC). Cases were of self-reported Caucasian ancestry, and ischemic stroke subtypes were determined according to TOAST criteria based on relevant clinical imaging. The University of Edinburgh collection comprised 727 ischemic stroke cases, consecutively collected as part of the Edinburgh Stroke Study. Cases were of self-reported Caucasian ancestry, with ischemic stroke subtypes determined according to TOAST criteria based on relevant clinical and imaging data.

The Munich samples included 1383 ischemic stroke cases. Cases were consecutive European Caucasians recruited from a single dedicated Stroke Unit at the Department of Neurology, Klinikum Großhadern, Ludwig-Maximilians-University, Munich. Ischemic stroke subtypes were determined according to TOAST criteria based on relevant clinical and imaging data. Controls for the UK samples were drawn from shared WTCCC controls obtained from the 1958 Birth Cohort. This is a prospectively collected cohort of individuals born in 1958 (<http://www.b58cgene.sgul.ac.uk/>), and ascertained as part of the national child development study (<http://www.cls.ioe.ac.uk/studies.asp?section=000100020003>). Data from this cohort are available as a common control set for a number of genetic and epidemiological studies. For the German samples controls were Caucasians of German origin participating into the population KORAgen study (www.gsf.de/kora/en/english.html). This survey represents a gender- and age stratified random sample of all German residents of the Augsburg area and consists of individuals 25 to 74 years of age, with about 300 subjects for each 10-year increment. All controls were free of a history of stroke or transient ischemic attack.

WTCCC Immunochip

The Immunochip samples consisted of 5 cohorts of patients of European ancestry with ischemic stroke. Participating centers were based in Belgium, Germany, Poland, Sweden, and the UK (2 cohorts, one from London [Imperial College; the BRAINS study] and one from Glasgow). All cohorts provided geographically and ancestry-matched controls. For the purposes of meta-analysis, the UK cohorts were treated as a single center in line with previous analyses undertaken in WTCCC2. The 5 discovery phase cohorts used the commercially available Immunochip array (Illumina, San Diego, CA), comprising ~200,000 SNPs. Genotyping for the 5 case cohorts was performed at the Sanger Centre, Hinxton, Cambridge, UK. Swedish controls were provided and genotyped by the Swedish SLE network, Uppsala, Sweden. Belgian control samples were provided through the efforts of the International Multiple Sclerosis Genetics Consortium. The Immunochip datasets were each imputed separately to the 1,000 Genomes Phase 1 integrated variant set (March 2012) using IMPUTE v2.2.2.

Studies in METASTROKE consortium (excluding studies included in either Stage 1 or Stage 2)

The Atherosclerosis Risk in Communities Study (ARIC)

The ARIC study is a prospective population-based study of atherosclerosis and clinical atherosclerotic diseases in 15,792 men and women, including 11,478 non-Hispanic white participants, drawn from 4 U.S. communities (Suburban Minneapolis, Minnesota; Washington County, Maryland; Forsyth County, North Carolina, and Jackson, Mississippi). In the first three communities, the sample reflects the demographic composition of the community. In Jackson, only black residents were enrolled. Ancestry was self-reported during an interview. Over 99% self-identified as either white or black. Only self-identified whites were included in the analyses.⁵ Hospitalized strokes that occurred by December 31, 2007 (median follow-up, 18.7 years) were included in the analyses. 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. Moreover, all local hospitals annually provided lists of stroke discharges (International Classification of Diseases, Ninth Revision, Clinical Modification codes 430 to 438), which were surveyed for ARIC participant discharges. Details on quality assurance for ascertainment and classification of stroke are described elsewhere.⁶ Briefly, the stroke diagnosis was assigned according to criteria adapted from the National Survey of Stroke.⁷ 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 are not included. Strokes were classified into hemorrhagic stroke (subarachnoid and intracerebral hemorrhage) and ischemic stroke (thrombotic and embolic brain infarction). A stroke was classified as ischemic when a brain CT or MRI revealed acute infarction and showed no evidence of hemorrhage. All definite ischemic strokes were further classified as lacunar, nonlacunar thrombotic, or cardioembolic on the basis of the recorded neuroimaging results. A stroke was classified as “lacunar” when 2 criteria were met: (1) typical location of the infarct (basal ganglia, brain stem, thalamus, internal capsule, or cerebral white matter); and (2) infarct size of ≤ 2 cm or unstated size.⁸ Definite or probable “cardioembolic” stroke required either (1) autopsy evidence of an infarcted area in the brain and a source of possible cerebral emboli in a vessel or the presence of an embolus in the brain or (2) medical record evidence of a possible noncarotid source of embolus such as moderate or greater valvular heart disease, atrial fibrillation, cardiac or arterial procedure (e.g., cardiac catheterization, open heart surgery, cerebral angiography, and carotid endarterectomy), or intracardiac thrombus. Definite or probable ischemic strokes that were not classified

as lacunar or cardioembolic, including atherothrombotic and unclassified thrombotic strokes, were labeled “nonlacunar.” Hemorrhagic strokes identified by ARIC were censored at the time of their occurrence.

Bio-Repository of DNA in Stroke (BRAINS)

The Bio-Repository of DNA in Stroke (BRAINS) is an international study recruiting highly phenotyped patients with stroke. For the purposes of the current work all patients were Caucasians. Diagnosis of stroke was confirmed using positive imaging (MRI or CT) and ischemic stroke subtypes were assigned using TOAST criteria, based on clinical, imaging and risk factor data. Controls were European-Ancestry, stroke-free participants from the shared WTCCC controls, a prospectively collected cohort of individuals born in 1958 (1958 Birth Cohort). The cohort has been described in detail elsewhere.⁹

Cardiovascular Health Study (CHS)

The CHS is a population-based cohort study of risk factors for coronary heart disease (CHD) and stroke in adults ≥ 65 years conducted across four field centers in the United States.⁶ The original predominantly white cohort of 5,201 persons (4,964 whites) was recruited in 1989-1990 from a random sample of people on Medicare eligibility lists and an additional 687 African-Americans were enrolled subsequently (1992-93) for a total sample of 5,888. Race was determined by self-identification at interview. In addition to the 5 categories used in the ARIC study, participants were also asked a second question as to whether they considered themselves to be of Hispanic origin. To reduce the possibility of confounding by population structure, these analyses were limited to participants of self-described European-ancestry. The study sample for these analyses includes participants who were free of CVD at baseline, had blood samples drawn at their baseline examination, consented to genetic testing, had DNA available, and had successful genome-wide genotyping assay. Participants were examined annually from enrollment to 1999, and since then continue to be under surveillance for stroke. Since baseline, participants have also been contacted twice a year to identify potential cardiovascular events, including stroke. In addition, all hospitalizations were screened for potential stroke events. For suspected events, information was collected from the participant or next of kin, from medical records, and, if needed, from the participant's physician. When available, CT and/or MRI scans or reports were reviewed centrally. Final adjudication of the occurrence of stroke, stroke types, and subtypes was undertaken by vascular neurologists at a consensus conference using all available information. Strokes were classified as ischemic if there was imaging (CT or MRI within 4 weeks), surgical or autopsy evidence excluding a hemorrhage, or in the absence of such direct evidence (in $<10\%$ of cases in FHS and Rotterdam, none in CHS) if the preponderance of indirect evidence (e.g. deficit limited to one limb or completely resolved within 72 hours, atrial fibrillation in persons not on anticoagulants) suggested the event was an ischemic rather than a hemorrhagic stroke. A stroke was classified as hemorrhagic if there was imaging, surgical, lumbar puncture or autopsy evidence of hemorrhage, and in the absence of direct evidence to the contrary, when the participant lost consciousness permanently or died within hours after onset of focal signs. The stroke type was defined as unknown if there was insufficient information available to categorize the event as ischemic or hemorrhagic. All ischemic and hemorrhagic strokes and strokes of unknown type were included in the analyses of total stroke with one exception: subarachnoid hemorrhages ($n=28$ across all studies) were excluded from these analyses since the heritability, risk factors and pathophysiologic mechanisms underlying subarachnoid hemorrhages are distinctly different from other stroke subtypes. Persons with a subarachnoid hemorrhage were censored at the time of the event. Only known ischemic strokes were included in the analysis of ischemic stroke. In secondary analyses we related those SNPs which reached genome-wide significance in our initial GWAS to the specific ischemic stroke subtype of

atherothrombotic stroke, also called atherothrombotic brain infarction (ABI). We used the best available definitions of definite and possible ABI in each cohort; both large artery atherosclerotic strokes and small-vessel or lacunar strokes were included in this phenotype, events known to be cardioembolic were excluded. For the analysis of ABI, participants were censored when they developed an alternative type of stroke.

deCODE Genetics

Cases, irrespective of age, were identified from a registry of individuals diagnosed with ischemic stroke or TIA at Landspítali University Hospital in Reykjavik, the only tertiary referral centre in Iceland, during the years 1993 to 2006. The ischemic stroke or TIA diagnoses were based on standard WHO criteria and imaging evidence (either CT or MRI), and were clinically confirmed by neurologists. Eligible patients who survived the stroke were invited to participate the genetic study, either by attending a recruitment center for deCODE's genetic studies, or they were visited at their home by a study nurse. Control subjects were participants from a large variety of genetic programs at deCODE. Individuals with confirmed stroke (identified by cross-matching with hospital lists), who had participated in genetic studies other than those of cardiovascular diseases (CVD) (but not participated in CVD studies) were excluded as controls.

Heart Protection Study (HPS)

The Heart Protection Study (HPS) was a large randomized trial involving individuals at increased risk of vascular events. Between 1994-1997 20,536 men and women aged 40-80 years were recruited from 69 collaborating hospitals in the United Kingdom (with ethics committee approval). Participants were eligible for inclusion provided they had non-fasting blood total cholesterol concentrations of at least 135 mg/dL (3.5 mmol/L) and either a previous diagnosis of coronary disease, ischemic stroke, other occlusive disease of non-coronary arteries, diabetes mellitus, or (if were men 65 years or older) treated hypertension. None of them was on statin therapy. At the initial screening visit, all participants provided written consent and began a "run-in" phase involving 4 weeks of placebo followed by 4 to 6 weeks of 40 mg simvastatin daily, after which compliant and eligible individuals were randomly allocated 40 mg simvastatin daily or matching placebo for approximately 5 years. Individuals entering HPS with a clinical diagnosis of ischemic stroke were used as cases in the METASTROKE study. Individuals entering HPS with pre-existing diabetes but no history of cerebrovascular disease, coronary heart disease or peripheral vascular disease were used as controls.

The Heart and Vascular Health Study (HVH)

The setting for this study was Group Health (GH), a large integrated health care system in western Washington State. Data were utilized from an ongoing case-control study of incident myocardial infarction (MI) and stroke cases with a shared common control group. Methods for the study have been described previously¹⁰⁻¹² and are briefly summarized below. The study was approved by the human subjects committee at GH, and written informed consent was provided by all study participants. All study participants were GH members and aged 30-79 years. MI and stroke cases were identified from hospital discharge diagnosis codes and were validated by medical record review. Controls were a random sample of GH members frequency matched to MI cases on age (within decade), sex, treated hypertension, and calendar year of identification. The index date for controls was a computer-generated random date within the calendar year for which they had been selected. For stroke cases, the index date was the date of admission for the first acute stroke. Participants were excluded if they were recent enrollees at GH, had a history of prior stroke, or if the incident event was a complication of a procedure

or surgery. Trained medical record abstractors collected eligibility and risk factor information from a review of the GH medical record using only data available prior to the index date and through a telephone interview. Medication use was ascertained using computerized GH pharmacy records. A venous blood sample was collected from all consenting subjects, and DNA was extracted from white blood cells using standard procedures. Diagnostic criteria for ischemic stroke were adopted from the Cardiovascular Health Study¹³ These criteria included (1) rapid onset of neurologic deficit or subarachnoid hemorrhage, (2) deficit persisting for longer than 24 hours unless computed tomography or magnetic resonance imaging show evidence of permanent damage, and (3) no underlying brain trauma, tumor, or infection to cause symptoms. These analyses were limited to ischemic stroke cases, namely those satisfying 1 or more of the following criteria: (a) Focal deficit, without evidence of blood on CT or MRI, (b) Focal deficit, with mottled appearance in the appropriate location on CT, or (c) surgery or autopsy evidence of infarction. Among ischemic strokes, the subtypes were defined as follows: Lacunar stroke (“SVD”) required either: (a) CT/MRI demonstrates a deep area of infarction (decreased density) less than 2 cm. across, or (b) A normal CT, but the clinical syndrome is typical of a lacunar infarction, that is: a pure motor stroke, a pure sensory stroke, hemiparesis plus ataxia, or dysarthria plus a clumsy hand. Embolic stroke (“CE”) required either (a) a recognized source of emboli such as atrial fibrillation, endocarditis, mitral stenosis, thrombus in heart, recent MI or cardiac surgery, or (b) a mottled appearance consistent with infarction on the CT. Atherosclerotic infarction (“LAA”) when there is no apparent source of emboli or evidence of lacunar infarction and there is evidence of large vessel atherosclerosis by carotid ultrasound or angiography.

The Ischemic Stroke Genetics Study (ISGS)/ Siblings With Ischemic Stroke Study (SWISS)

The Siblings with Ischemic Stroke Study (SWISS) is a multicenter affected sibling pair study enrolling probands with ischemic stroke at 66 US medical centers and 4 Canadian medical centers.¹⁴ All probands are adult men and women over the age of 18 years diagnosed with ischemic stroke confirmed by a study neurologist on the basis of history, physical examination and CT or MR imaging of the brain. Additionally all probands were required to have at least one living sibling with a history of stroke. Siblings were enrolled using proband-initiated contact or direct contact when permitted by Institutional Review Boards.¹⁵ Clinical exclusion criteria mirrored that in ISGS. Concordant siblings had their diagnosis of ischemic stroke confirmed by review of medical records by a central vascular neurology committee. Subtype diagnoses were assigned to the index strokes of probands and concordant siblings according to TOAST criteria.¹⁶ Readily available US controls were utilized, including stroke-free participants from the Baltimore Longitudinal Study of Aging and the National Institute of Neurological Diseases and Stroke neurologically normal control series taken from the Coriell Cell Repositories. All controls had been previously genotyped and described in detail elsewhere.

The Ischemic Stroke Genetic Study (ISGS) is a multicenter study where inpatient cases were recruited from five United States academic medical centers.²⁰ Cases are adult men and women over the age of 18 years diagnosed with first-ever ischemic stroke confirmed by a study neurologist on the basis of history, physical examination and CT or MR imaging of the brain who were enrolled within 30 days of onset of stroke symptoms. Cases exclusion criteria include: a mechanical aortic or mitral valve at the time of the index ischemic stroke, central nervous system vasculitis, bacterial endocarditis, cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL), Fabry disease, homocystinuria, mitochondrial encephalopathy with lactic acidosis and stroke-like episodes (MELAS), or sickle cell anemia. Stroke severity at enrollment was assessed using the NIH Stroke Scale with the diagnostic evaluation including head CT (95%) or MRI (83%), electrocardiography (92%), cervical

arterial imaging (86%), and echocardiography (74%). Medical records from all cases were centrally reviewed by a vascular neurology committee and assigned ischemic stroke subtype diagnoses according to criteria from the Trial of ORG10172 (TOAST), the Oxfordshire Community Stroke Project, and the Baltimore-Washington Young Stroke Study.

Rotterdam

The Rotterdam Study is a population-based cohort study among inhabitants of a district of Rotterdam (Ommoord), The Netherlands, and aims to examine the determinants of disease and health in the elderly with a focus on neurogeriatric, cardiovascular, bone, and eye disease.²¹ In 1990-1993, 7,983 persons participated and were re-examined every 3 to 4 years. After enrollment in the Rotterdam Study, participants are continuously monitored for incident stroke through automated linkage of the study database with files from general practitioners. Nursing home physicians' files and files from general practitioners of participants who moved out of the district are scrutinized as well. Additional information is obtained from hospital records. Potential strokes are reviewed by research physicians and verified by an experienced stroke neurologist. Stroke is defined as rapidly developing clinical signs of focal or global disturbance of cerebral function with no apparent cause other than a vascular origin. History of stroke at baseline was assessed during the baseline interview and verified in medical records. Strokes are further classified as cerebral infarction or intracerebral hemorrhage based on neuroimaging reports. If neuroimaging is lacking, a stroke is classified as unspecified. Subarachnoid hemorrhages were excluded. Ischemic strokes are subtyped into large-vessel disease, small vessel disease, or cardio-embolic based on all available clinical and imaging data. If insufficient information is present to reliably subtype into one of three categories, the ischemic stroke is classified as unspecified ischemic stroke. For the current study, too few cases of ischemic stroke subtypes were present to reliably perform statistical analyses. Participants were followed from baseline to stroke, death, last health status update when they were known to be stroke-free, or January 1, 2005, whichever came first. Follow-up was complete up to January 1, 2005, for 99.1% of potential person-years.

The Sahlgrenska Study of Ischemic Stroke (SAHLSIS)

SAHLSIS has been described in more detail previously.¹⁷ In brief, Caucasian patients (n=1088) under 70 years of age who presented with first-ever or recurrent acute ischemic stroke were recruited consecutively at four stroke units in western Sweden. Healthy Caucasian community controls (n=600) matched for age and sex from the same geographic area were randomly selected from participants in a population-based survey or the Swedish Population Register. All patients were examined by a physician trained in stroke medicine, at baseline and at a follow-up three months after inclusion. All patients underwent ECG and neuroimaging with computed tomography (CT) and/or magnetic resonance imaging (MRI). Extracranial carotid and vertebral duplex ultrasound, MR angiography, catheter angiogram, transcranial Doppler ultrasound, transthoracic and/or transesophageal echo-cardiography were performed when clinically indicated. IS was defined as an episode of focal neurological deficits with acute onset and lasting >24 hours or until death, with no apparent non-vascular cause, and no signs of primary hemorrhage on brain imaging. All participants provided written informed consent prior to enrolment. For participants who were unable to communicate, consent was obtained from their next-of-kin. This study was approved by the Ethics Committee of the University of Gothenburg.

Whole blood was isolated from all individuals for genotyping. Plasma was additionally isolated from the first 600 patients and controls for protein measurement. Genotyping for 600 controls was performed using the Illumina Human OmniExpress 750K array at the Broad Institute at Harvard

Medical School (Cambridge USA). Genotype data for rs7906302 and rs1338423 were extracted from the above array for all genotyped controls in SAHLSIS. Genotyping was performed blinded to case/control status.

Web Resources

The URLs for data presented herein are as follows:

The 1000 Genomes Project: <http://www.1000genomes.org/>

GWAMA software: <http://www.well.ox.ac.uk/gwama/>

METASOFT software: <http://genetics.cs.ucla.edu/meta/>

QUANTO software: <http://biostats.usc.edu/software.html>

Ensemble Variant Effect Predictor tool: <http://www.ensembl.org/info/docs/tools/vep/index.html>

SNP Annotation and Proxy Search (SNAP): <http://www.broadinstitute.org/mpg/snap/>

The Genotype-Tissue Expression project (*GTE*_x) project: <http://www.gtexportal.org/home/>

The Wellcome Trust Sanger Institute Genevar (GENe Expression VARiation) tool:

<http://www.sanger.ac.uk/resources/software/genevar/>

LocusZoom software: <http://csg.sph.umich.edu/locuszoom/>

Accession Numbers

The dbGaP accession number for GEOS data reported in this paper is **phs000292.v1.p1**.

Supplemental Tables

Supplemental Table I. Study design and characteristics of participating studies in the discovery phase.

	CADISP	GEOS EA	GEOS AA	MILANO	RACE 1	RACE 2	SIFAP	WTCCC-UK
Cases	555	448	381	201	1218	339	981	382
Stroke Subtype*								
Cardioembolic (CE)	211	90	74	37	226	67	170	44
Large Artery Atherosclerosis (LA)	67	37	23	34	201	56	184	79
Small Vessel (SV)	31	54	77	12	190	42	104	91
Other known causes	18	29	24	N/A	15	6	164	NA
Undetermined causes	228	238	183	67	586	168	331	158
First-ever ischemic stroke (%)	~100	100	100	100	100	100	N/A	100
CT/MRI Brain Imaging (%)	100	97.80	96.3	100	100	100	N/A	100
Age, mean (SD)	43.73 (9.9)	41.0 (7.0)	275 (61.4)	45.0(10.4)	50.1 (9.9)	50.2 (9.2)	41.7 (7.4)	51.9 (7.3)
Male, n (%)	339 (61.1)	41.9 (6.8)	207 (54.3)	120 (60.9)	638 (52.4)	272 (80.4)	599 (61.1)	228 (59.7)
Controls	9259	498	352	407	1158	3295	1825	5175
Age, mean (SD)	N/A	39.5 (6.7)	40.0 (6.8)	50.8 (8.1)	51.9 (7.9)	60.9 (13.2)	55.2 (11.6)	52 (?)
Male, n (%)	N/A	282 (56.6)	196 (55.7)	357 (87.8)	613 (53.0)	1838 (55.8)	899 (49.3)	2611 (50.5)
Control Source	Internal	Internal	Internal	Internal	External	External	External (KORA study)	External (1958 British Birth

N/A: not available
* defined by TOAST criteria

Supplemental Table II. Study design and characteristics of participating studies in the validation phase.

	ASGC	Munich	WTCCC2-immunochip-Belgium	WTCCC2-immunochip-Germany	WTCCC2-immunochip-Krakow	WTCCC2-immunochip-Sweden	WTCCC2-immunochip-UK
Cases	227	275	97	96	93	96	119
First-ever ischemic stroke (%)	N/A	94.2%	N/A	N/A	N/A	N/A	N/A
CT/MRI Brain Imaging (%)	97.4	100	100	N/A	100	N/A	100
Age (mean \pm SD)	52.3 (8.7)	48.8 (8.51)	49.9 \pm 10.19	48.7 \pm 10.07	51.0 \pm 8.54	53.4 \pm 6.49	49.8 \pm 9.9
Male Participants, n (%)	151 (66.5)	188 (68.4)	67 (71.1)	65 (67.7)	58 (62.3)	69 (71.9)	73 (61.3)
Controls	1244	797	319	2344	255	996	1790
Age	66.3 (7.5)	62.7(10.9)	N/A	N/A	N/A	N/A	N/A
% Males	625 (50.2)	410 (51.4)	152 (47.6)	1186 (50.3)	94 (36.9)	269 (27.0)	741 (41.4)
Control Source	Internal	External	Internal	Internal	Internal	Internal	Internal

N/A: not available

Supplemental Table III. Genotyping and analysis methods used by each participating studies in discovery phase

	CADISP	GEOS	MILANO	SIFAP ¹	RACE 1 ¹	RACE 2	WTCCC2-UK
Genotyping Platform	Human610-Quad BeadChip (Illumina, San Diego, USA)	Illumina HumanOmni1-Quad_v1-0_B	Cases: Illumina Human610-Quad v1_B or Human660W-Quad v1_A ; Controls: Illumina HumanHap610-Quad	Illumina Human Omni 2.5M-Quad array	Illumina Human660W-Quad	Illumina HumanOmni express	Cases: Illumina Human660W-Quad ; Controls: Illumina Human1.2M-Duo
QC filters	Individuals with a call rate <0.95, sex mismatch and duplicates were excluded.	Individuals with unexpected relatedness, gender mismatch, population outliers were excluded; SNPs with missing call rate >=5%, HWE P<0.0001, monomorphic SNP were excluded	Exclude individuals with unexpected relatedness, gender mismatch, population outliers; Exclude SNPs genotyped in <99% of individuals, HWE p<0.0001 or MAF<0.01	Exclude SNPs with missing call rate >=5%, HW P<0.0001, monomorphic SNP	Exclude SNPs with missing call rate >=5%, HW P<0.0001, monomorphic SNP	Exclude SNPs with missing call rate >=5%, HW P<0.0001, monomorphic SNP	Exclude Samples as outliers on call rate, ancestry, heterozygosity and average probe intensity using a Bayesian clustering algorithm. Samples were also removed if they exhibited gender or cryptic relatedness with other WTCCC2 samples; Exclude SNPs genotyped in <99% of individuals, HWE p<0.0001 or MAF<0.01
Pre-phasing software	SHAPE-IT	IMPUTE2	no prephasing	IMPUTE2	IMPUTE2	IMPUTE2	IMPUTE2
Imputation software	IMPUTE2	IMPUTE2	IMPUTE2	IMPUTE2	IMPUTE2	IMPUTE2	IMPUTE2
1KG reference panel used (website link)	1000G Phase 1 integrated variant set (Mar 2012)	1000G phase 1 interim set (June 2011)	1000G Phase 1 integrated variant set (Mar 2012)	1000G phase 1 interim set (June 2011)	1000G phase 1 interim set (June 2011)	1000G phase 1 interim set (June 2011)	1000G Phase 1 integrated variant set (Mar 2012)
Model adjustments	sex, PC1-10	age, study phase, PC1	sex, PC1 & PC5	Age, sex, PC1-5	adjusted for age, sex, PCs1-5	adjusted for age, sex, PCs1-10	sex

¹ For RACE and SIFAP studies where controls were genotyped separately from cases, raw scan data from cases and controls were clustered together to obtain a combined set of genotypes. In addition, a subset of control subjects (28 control subjects for SIFAP and 95 control subjects for RACE) were re-genotyped together with cases to provide cross-set duplicates. This joint clustering was used to minimize possible artifactual differences in allelic frequency between cases and controls due to the fact that they were genotyped at different times and the cross-set duplicates were used to detect such artifacts that may have occurred.

Supplemental Table IV. Genotyping and analysis methods used in the validation studies.

	ASGC	Munich	WTCCC2-immuno-chip-Belgium	WTCCC2-immuno-chip-Germany	WTCCC2-immuno-chip-Krakow	WTCCC2-immuno-chip-Sweden	WTCCC2-immuno-chip-UK
Genotyping Platform	Human610-Quad BeadChip (Illumina, San Diego, USA)	Human660W-Quad (cases) / Human 550 (controls)	Illumina Immuno-chip	Illumina Immuno-chip	Illumina Immuno-chip	Illumina Immuno-chip	Illumina Immuno-chip
QC filters	Call rate <97%, MAF <0.5%, P-HWE <1x10 ⁻⁴ , ambiguous SNPs, SNPs not in 1000G, SNPs that couldn't be lifted to B37	Bayesian clustering approach	Exclude individuals with high heterozygosity, high missingness, relatedness to other individuals or if they did not segregate with HapMap II CEU individuals based on PCA analysis using EIGENSTRAT. Exclude SNPs with high missingness, deviation from HWE, low MAF (<0.01)	Exclude individuals with high heterozygosity, high missingness, relatedness to other individuals or if they did not segregate with HapMap II CEU individuals based on PCA analysis using EIGENSTRAT. Exclude SNPs with high missingness, deviation from HWE, low MAF (<0.01)	Exclude individuals with high heterozygosity, high missingness, relatedness to other individuals or if they did not segregate with HapMap II CEU individuals based on PCA analysis using EIGENSTRAT. Exclude SNPs with high missingness, deviation from HWE, low MAF (<0.01)	Exclude individuals with high heterozygosity, high missingness, relatedness to other individuals or if they did not segregate with HapMap II CEU individuals based on PCA analysis using EIGENSTRAT. Exclude SNPs with high missingness, deviation from HWE, low MAF (<0.01)	Exclude individuals with high heterozygosity, high missingness, relatedness to other individuals or if they did not segregate with HapMap II CEU individuals based on PCA analysis using EIGENSTRAT. Exclude SNPs with high missingness, deviation from HWE, low MAF (<0.01)
Pre-phasing software	Minimac	MACH	IMPUTE2	IMPUTE2	IMPUTE2	IMPUTE2	IMPUTE2
Imputation software	Mach	MACH	IMPUTE2	IMPUTE2	IMPUTE2	IMPUTE2	IMPUTE2
1KG reference panel used (website link)	Phase 1 (v3)	1000G Phase 1 integrated variant set	1000G Phase 1 integrated variant set (Mar 2012);	1000G Phase 1 integrated variant set (Mar 2012);	1000G Phase 1 integrated variant set (Mar 2012);	1000G Phase 1 integrated variant set (Mar 2012);	1000G Phase 1 integrated variant set (Mar 2012);
Model adjustments	age, sex and the first 4 principal components of ancestry	No covariates	gender, 5 PCs	gender, 10 PCs	gender, 5 PCs	gender, 5 PCs	gender, 10 PCs

Supplemental Table V. Study design and characteristics of METASTROKE cohorts.

Cohorts	Number of cases	Number of controls	Number (%) of cases < 60 years	Mean Age (yrs)	Study design
ARIC	385	8803	42 (10.9)	57.3 ± 5.3	Population-based, prospective
BRAINS	361	444	90 (24.9)	74.4 ± 14.2	Cross-sectional
CHS	454	2817	0 (0)	81.6 ± 6.11	Population-based, prospective
deCODE	2391	26970	N/A	73.3 ± 12.1	Cross-sectional
HPS	578	468	231 (40.0)	64.9 ± 7.5	Cross-sectional
HVH	566	1290	97 (17.1)	69.2 ± 8.6	Cross-sectional
ISGS/SWISS	1070	2329	N/A	66.5 ± 13.6	Cross-sectional
Rotterdam	367	5396	N/A	70.8 ± 7.5	Population-based, prospective

N/A: information not available

Supplemental Table VI. SNPs with suggestive associations with early-onset ischemic stroke and their proxy SNPs in chr10 region.

SNP ¹	Position (hg18)	variant	Association p-value for young-onset stroke in the Discovery Phase	AFR_MAF	ASN_MAF	EUR_MAF	LD with rs11196288 (r ² /D')			Functional Prediction
							CEU	YRI	CHB/JPT	
rs4075717	115037998	G/A	4.15E-05	0.02	0.21	0.04	0.618/0.844	NA	0.117/0.452	
rs12219131	115038675	C/T	1.65E-05	0.23	0.21	0.05	0.618/0.844	0.011/0.847	0.117/0.452	
rs11196283	115039246	C/T	1.46E-05	0.22	0.21	0.04	0.529/0.727	0.002/0.125	0.103/0.412	
rs4917648	115042323	A/G	3.50E-06	0.23	0.2	0.06	0.618/0.844	0.011/1.000	0.196/0.600	
rs11196284	115045236	C/T	1.29E-05	0.01	0.19	0.04	0.737/1	NA	0.173/0.616	
rs11196287	115052431	A/C	6.25E-07	0.04	0.38	0.04	1/1	0.330/0.574	0.765/1.000	
rs12218673	115053330	C/T	1.36E-05	0.02	0.34	0.04	0.737/1	0.001/1.000	0.960/1.000	
rs61872854	115055671	A/G	2.85E-07	0.05	0.35	0.04	1/1	NA	1/1	
rs7906302	115056097	G/T	1.09E-06	0.15	0.41	0.05	0.881/1	0.231/1.000	0.765/1.000	
rs4918806	115057210	C/G	4.44E-07	0.22	0.41	0.05	1/1	0.193/1.000	0.765/1.000	Regulatory
rs11196288	115057443	A/G	1.24E-07	0.07	0.35	0.04	--	--	--	
rs11196289	115059235	C/T	5.04E-07	0.04	0.35	0.04	0.881/1	0.590/1.000	0.961/1.000	
rs1338423	115060426	G/A	6.52E-07	0.07	0.36	0.04	0.881/1	0.793/1.000	0.961/1.000	Regulatory
rs10885429	115061139	A/T	2.74E-06	0.14	0.42	0.06	0.881/1	0.231/1.000	0.737/1.000	
rs4918807	115062997	A/G	1.22E-06	0.07	0.36	0.04	0.881/1	1/1	0.961/1.000	
rs4918809	115063701	C/T	3.82E-06	0.18	0.41	0.06	0.786/1	0.217/1.000	0.737/1.000	
rs4918810	115064197	C/T	5.17E-07	0.07	0.35	0.04	0.881/1	1/1	1/1	

¹ Twelve SNPs with suggestive associations ($P < 5.0 \times 10^{-6}$ in the Discovery Phase, trans-ethnic meta-analysis) are in bold; proxy SNPs (unbold) were identified as having $r^2 > 0.8$ with any of the 12 stroke-associated SNPs in ≥ 2 reference populations (CEU, CHB/JPT or YRI) based on 1000 Genomes Pilot data.

Supplemental Table VII. Top association signals for early-onset ischemic stroke

Chr	SNP	Position	Candidate Gene(s)	Effect allele	Other allele	EAF	Discovery Stage		Follow-up Stage		Discovery + Follow-up ^a	
							OR (95% CI)	P-value	OR (95% CI)	P-value	OR (95% CI)	P
1. European-only GWA meta-analysis												
1q24.2	rs1800594	169510524	F5	A	G	0.69	1.23 (1.13-1.34)	1.39E-06	1.08 (0.94 – 1.24)	0.30	1.18 (1.11-1.27)	2.26E-06
	rs9332599	169512980		G	A	0.67	1.20 (1.11-1.30)	3.90E-06	1.05 (0.93 – 1.19)	0.39	1.16 (1.08-1.23)	1.16E-05
2p15	rs74739745	61313269	KIAA1841-	G	A	0.98	1.88 (1.49-2.37)	8.52E-08	0.77 (0.51 – 1.17)	0.22	1.53 (1.25-1.87)	2.78E-05
	rs140510319	61367943	LOC39803	C	A	0.97	1.75 (1.40-2.19)	8.42E-07	0.82 (0.54 - 1.25)	0.35	1.49 (1.23-1.81)	6.13E-05
14q23.3	rs12434854	65896234	FUT8	G	A	0.93	1.38 (1.21-1.58)	2.07E-06	1.08 (0.87 – 1.34)	0.50	1.29 (1.15-1.45)	7.98E-06
	rs12435908	66121468		A	C	0.93	1.37 (1.20-1.56)	3.25E-06	1.10 (0.89 – 1.36)	0.39	1.29 (1.15-1.44)	7.19E-06
17q25.3	rs6565653	78201783	SLC26A11	T	G	0.46	1.21 (1.12-1.31)	1.08E-06	1.12 (0.87 – 1.44)	0.39	1.20 (1.12-1.29)	5.51E-07
	rs12601526	78202358		C	T	0.50	1.21 (1.12 -1.31)	1.08E-06	1.05 (0.81 – 1.36)	0.71	1.20 (1.11-1.29)	1.14E-06
2. Trans-ethnic GWA meta-analysis												
10q25.3	rs61872854	115055671	Intergenic (TCF7L2 – HABP2)	A	G	0.05	1.39 (1.22-1.57)	2.85E-07	1.80 (1.06-3.05)	0.03	1.40 (1.25-1.59)	2.21E-08
	rs11196288	115057443		G	A	0.05	1.39 (1.23-1.58)	1.24E-07	1.80 (1.06-3.05)	0.03	1.41 (1.25-1.59)	9.50E-09

^a Study-specific estimates were meta-analyzed using fixed effect model

Supplemental Table VIII. Top associated SNPs in GWAS analysis of stroke subtypes

Chr	Position	SNP	Candidate Gene(s)	Subtype	Effect allele	Other allele	EAF	Association with stroke subtype		
								OR (95% CI)	P-value	Heterogeneity I ²
1. European-only										
12	21323611	rs2417957	SLCO1B1	LA	C	T	0.97	0.36 (0.25-0.51)	2.64E-08	34%
10	18729855	rs72786098	CACNB2	SV	G	A	0.97	0.28 (0.18 – 0.44)	2.20E-08	60%
2. Trans-ethnic association										
1	51298205	rs138364069	FAF1	LA	C	T	0.98	0.28 (0.18 – 0.42)	6.81E-09	36%
1	51439456	rs3176471	CDKN2C	LA	A	T	0.98	0.27 (0.17 – 0.43)	4.09E-08	57%

Supplemental Table IX. SNPs previously associated with ischemic stroke (and/or ischemic stroke subtypes) and their associations with early-onset stroke

Gene (ref)	Chr	Pos	SNP (ref)	RA	RAF	Previously reported OR (trait)	Association with early-onset ischemic stroke ^a				
							European ancestry ^b			Trans-ethnic ^c	
							Trait	OR (95% CI)	P	OR (95% CI)	P
<i>NINJ2</i> ^{12, 20}	12	775199	rs11833579	A	0.23	1.06 (IS)	IS	1.00 (0.91 – 1.09)	0.96	1.0 (0.94 – 1.07)	0.89
							CE	1.00 (0.86 – 1.17)	0.99	1.10 (0.98 – 1.24)	0.12
							LAA	1.05 (0.88 – 1.26)	0.56	1.05 (0.92 – 1.20)	0.46
							SV	1.06 (0.87 – 1.31)	0.55	0.96 (0.83 – 1.11)	0.60
							UE	0.93 (0.82 – 1.05)	0.23	0.99 (0.91 – 1.09)	0.91
<i>ALDH2</i> ²¹	12	112233018	rs10744777	T	0.68	1.10 (IS)	IS	1.08 (1.00 – 1.17)	0.05	1.05 (0.97 – 1.12)	0.21 ^d
							CE	1.03 (0.89 – 1.19)	0.68	1.03 (0.92 – 1.15)	0.64
							LAA	1.06 (0.90 – 1.25)	0.46	1.10 (0.97 – 1.26)	0.13
							SV	1.06 (0.86 – 1.29)	0.57	1.03 (0.82 – 1.30)	0.79 ^d
							UE	1.12 (1.00 – 1.25)	0.05	1.06 (0.92 – 1.21)	0.44 ^d
<i>PITX2</i> ^{12, 22}	4	111710169	rs2200733	T	0.12	1.36 (CE)	IS	1.15 (1.03 – 1.28)	0.01	1.11 (1.02 – 1.20)	0.02
							CE	1.25 (1.03 – 1.52)	0.02	1.20 (1.03 -1.39)	0.02
							LAA	1.20 (0.95 – 1.53)	0.13	1.15 (0.84 – 1.58)	0.37 ^d
							SV	1.29 (0.99 – 1.59)	0.06	1.12 (0.88 – 1.43)	0.34 ^d
							UE	1.15 (0.98 – 1.34)	0.06	1.14 (1.02 – 1.28)	0.03

<i>ZFHX3</i> ^{12, 23}	16	73029160	rs7193343	T	0.23	1.25 (CE)	IS	0.99 (0.90 – 1.09)	0.83	1.01 (0.94 – 1.08)	0.84
							CE	1.04 (0.87 – 1.24)	0.68	1.02 (0.90 – 1.17)	0.68
							LAA	1.06 (0.86 – 1.31)	0.57	1.10 (0.95 – 1.28)	0.19
							SV	0.91 (0.71 – 1.15)	0.42	0.97 (0.83 – 1.13)	0.67
							UE	1.06 (0.92 – 1.21)	0.42	1.04 (0.95 – 1.14)	0.42
<i>CDKN2B-AS1</i> ^{12, 24}	9	22115959	rs2383207	G	0.57	1.15 (LA)	IS	1.10 (1.02-1.19)	0.01	1.08 (0.98 – 1.19)	0.13 ^d
							CE	1.08 (0.94 – 1.23)	0.29	1.06 (0.91 - 1.22)	0.45 ^d
							LAA	1.20 (1.03 - 1.40)	0.02	1.15 (0.96 – 1.38)	0.12 ^d
							SV	1.18 (0.98 – 1.41)	0.08	1.11 (0.94 – 1.32)	0.20 ^d
							UE	1.05 (0.95 – 1.17)	0.34	1.06 (0.94 – 1.20)	0.33 ^d
<i>HDAC9</i> ^{12, 25}	7	19031935	rs11984041	A	0.09	1.39 (LA)	IS	1.05 (0.93 – 1.19)	0.40	1.08 (0.99 – 1.11)	0.10
							CE	0.96 (0.76 – 1.21)	0.71	1.04 (0.77 – 1.40)	0.80 ^d
							LAA	1.23 (0.95 – 1.59)	0.12	1.19 (0.91 – 1.54)	0.19
							SV	1.14 (0.85 – 1.52)	0.38	1.05 (0.86 – 1.28)	0.62
							UE	1.00 (0.84 – 1.19)	0.99	1.05 (0.92 – 1.19)	0.47
<i>6p21</i> ⁵	6	44594159	rs556621	A	0.33	1.21 (LA)	IS	0.95 (0.88 – 1.03)	0.20	0.97 (0.91 – 1.05)	0.54 ^d
							CE	1.00 (0.87 – 1.15)	0.98	1.02 (0.86 – 1.20)	0.82 ^d
							LAA	0.92 (0.78 – 1.09)	0.32	0.95 (0.83 – 1.08)	0.41
							SV	1.00 (0.83 – 1.21)	0.97	1.02 (0.89 – 1.17)	0.80
							UE	0.92 (0.83 – 1.03)	0.15	0.95 (0.85 – 1.07)	0.42 ^d

<i>ABO</i> ²⁶	9	136149229	rs505922	C	0.32	1.07 (IS)	IS	1.15 (1.07- 1.25)	2.4x10⁻⁴	1.10 (1.04 – 1.17)	6.9x10⁻⁴
						1.13 (CE)	CE	1.22 (1.06-1.40)	0.005	1.18 (1.06 – 1.31)	0.002
						1.23 (LAA)	LAA	1.21 (1.04 – 1.43)	0.02	1.18 (1.04 – 1.33)	0.008
							SV	0.97 (0.80 – 1.17)	0.74	0.94 (0.82 – 1.07)	0.36
							UE	1.18 (1.06 – 1.32)	0.002	1.10 (1.02 – 1.20)	0.02

Abbreviations: Chr=chromosome, pos=position, RA=risk allele, RAF=risk allele frequency, IS=ischemic stroke, CE=cardioembolic stroke, LAA=large artery atherosclerotic stroke, SV=small vessel stroke, UE=undetermined etiology stroke

^a All association results were obtained by meta-analyzing study-specific estimates. In trans-ethnic meta-analysis, odds ratios and p-values taken from the random effects model if evidence for between-study heterogeneity (i.e., $I^2 > 20\%$); otherwise, odds ratios and p-values based on fixed effect model.

^b A total of 2,567 total ischemic stroke cases, including 546 CE, 403 LAA, 294 SV and 1035 UE subtype cases.

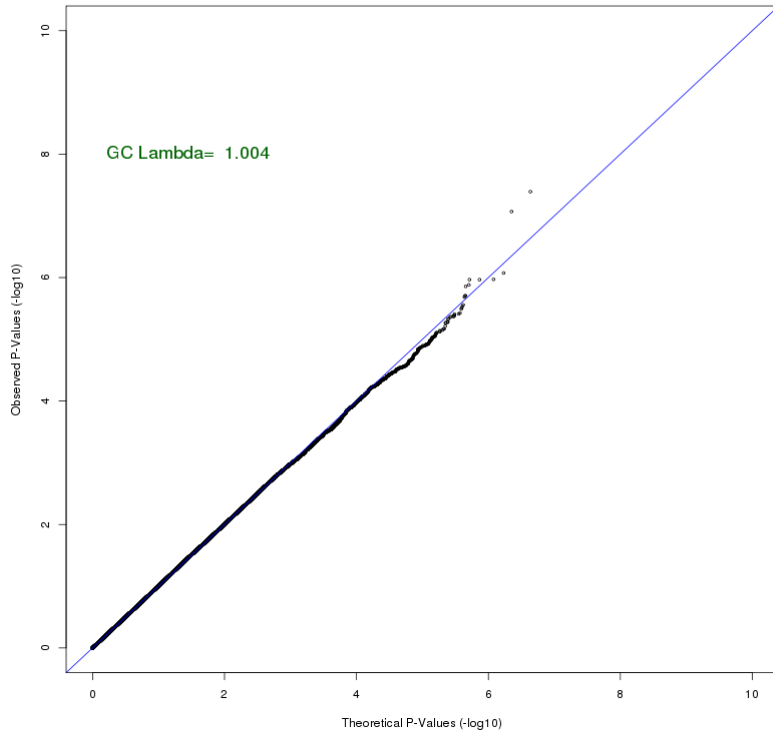
^c A total of 4,501 total ischemic stroke cases, including 882 CE, 657 LAA, 591 SV and 1854 UE subtype cases.

^d SNPs with evidence for between-study heterogeneity

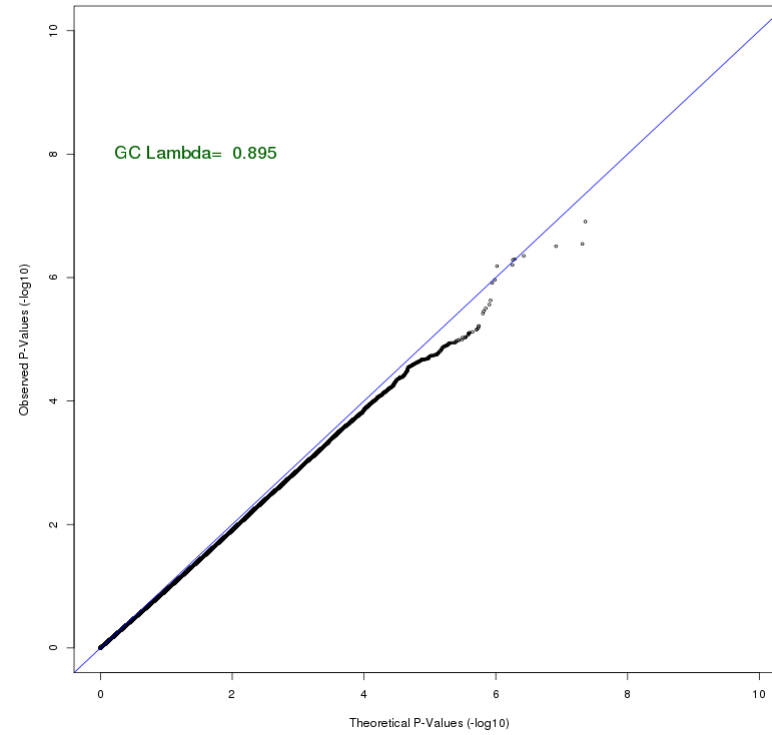
Supplemental Figures

Supplemental Figure I. QQ plot of (A) European-only meta-analysis and (B) trans-ethnic meta-analysis.

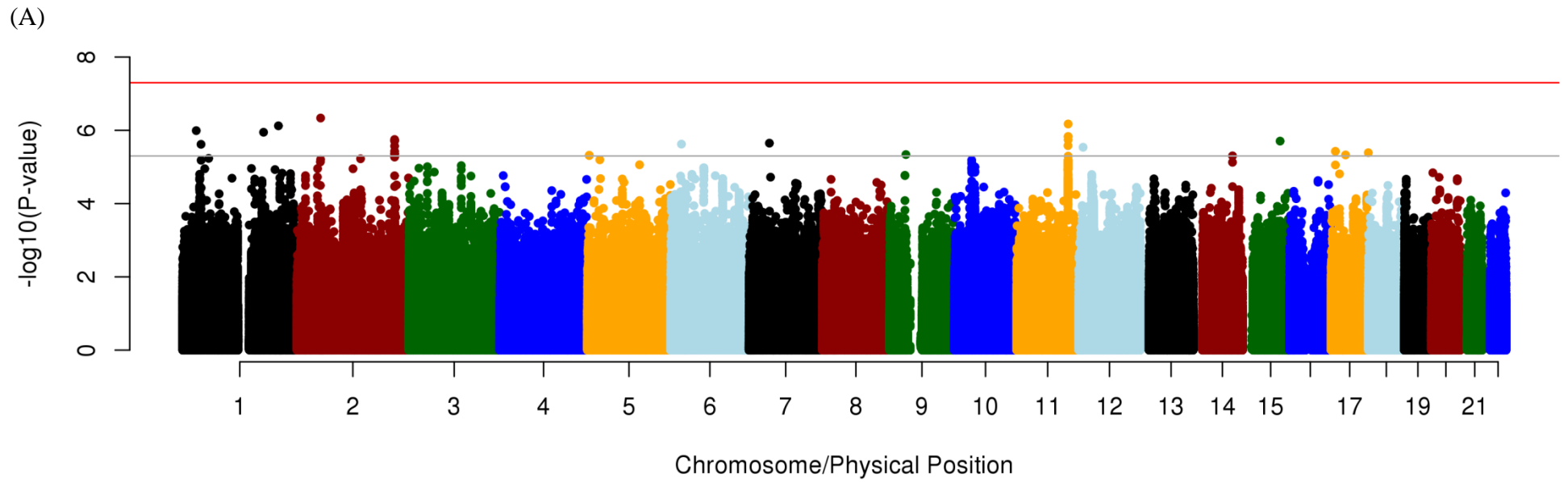
(A)



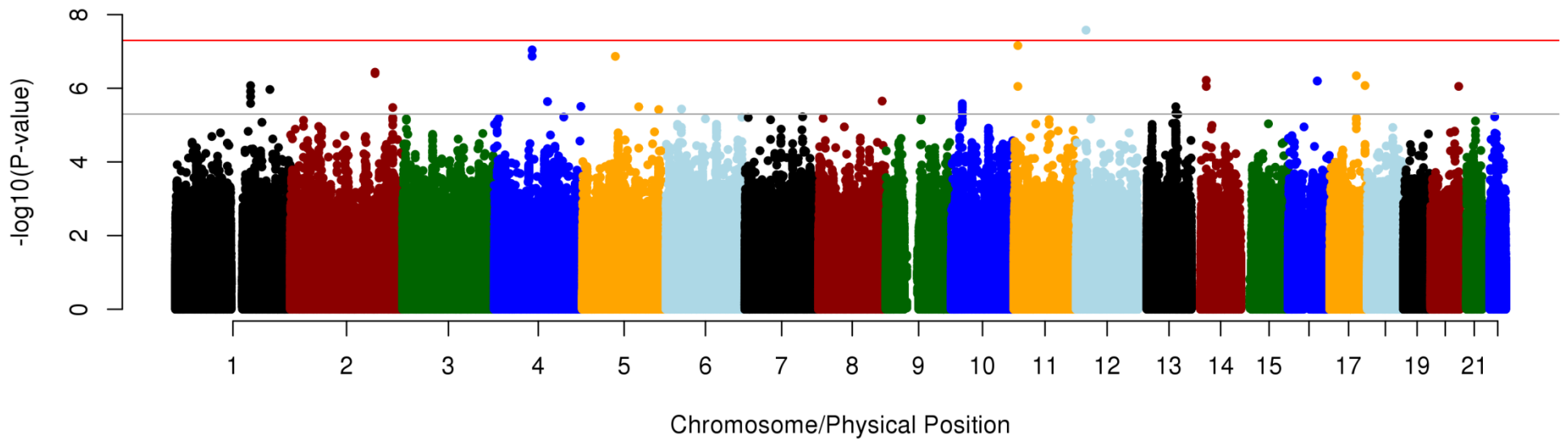
(B)



Supplemental Figure II. European-ancestry GWAS results of stroke subtypes (A) cardioembolic, (B) large artery atherosclerotic, (C) small vessel stroke and (D) undetermined.

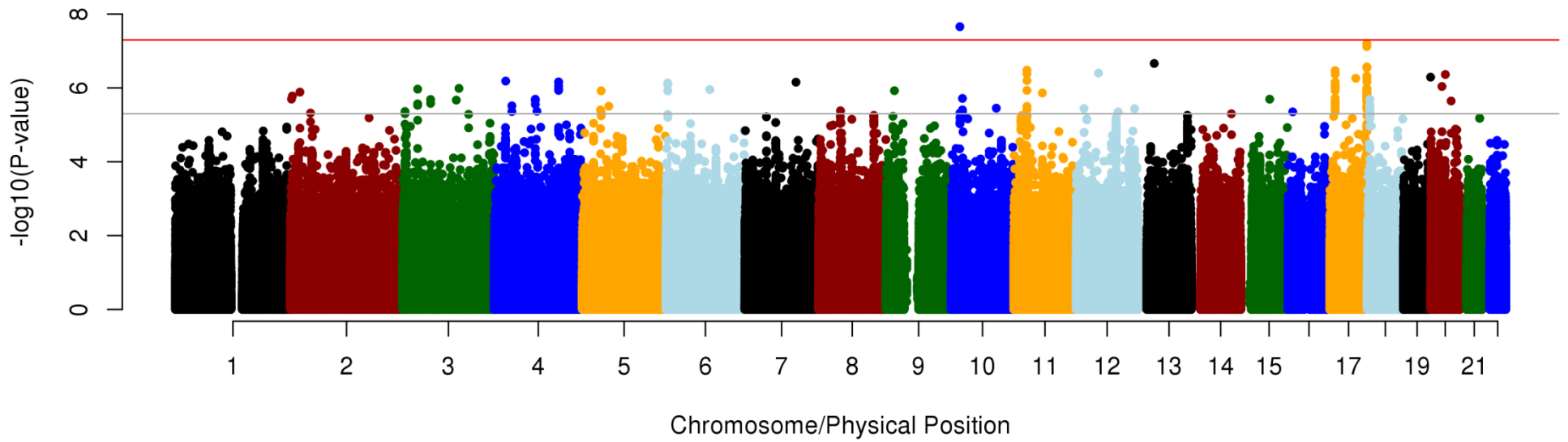


(B)

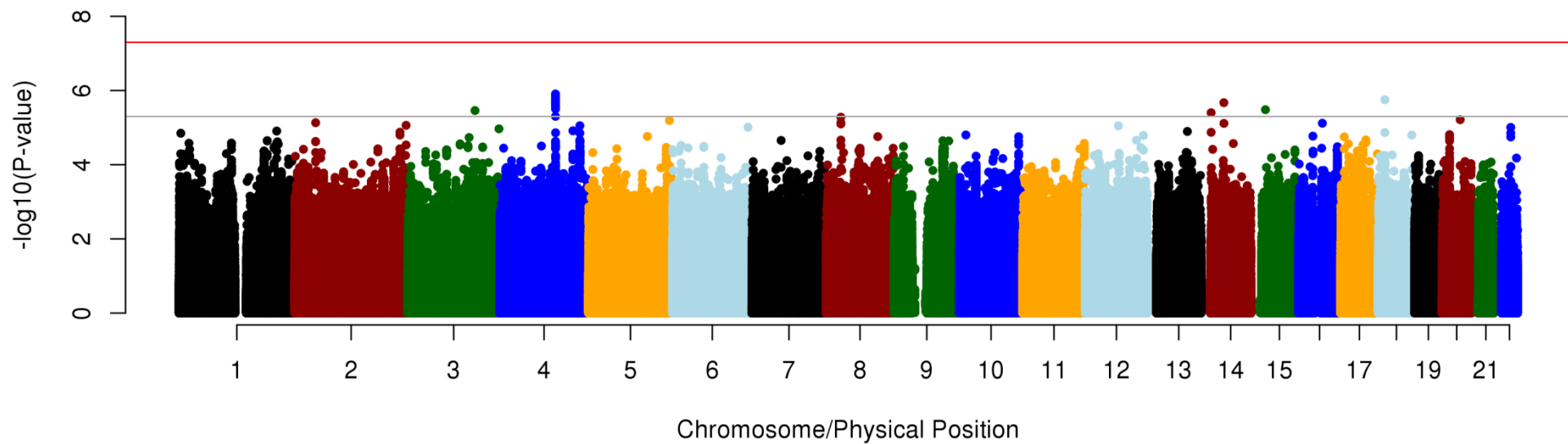


Supplemental Figure II. (continue)

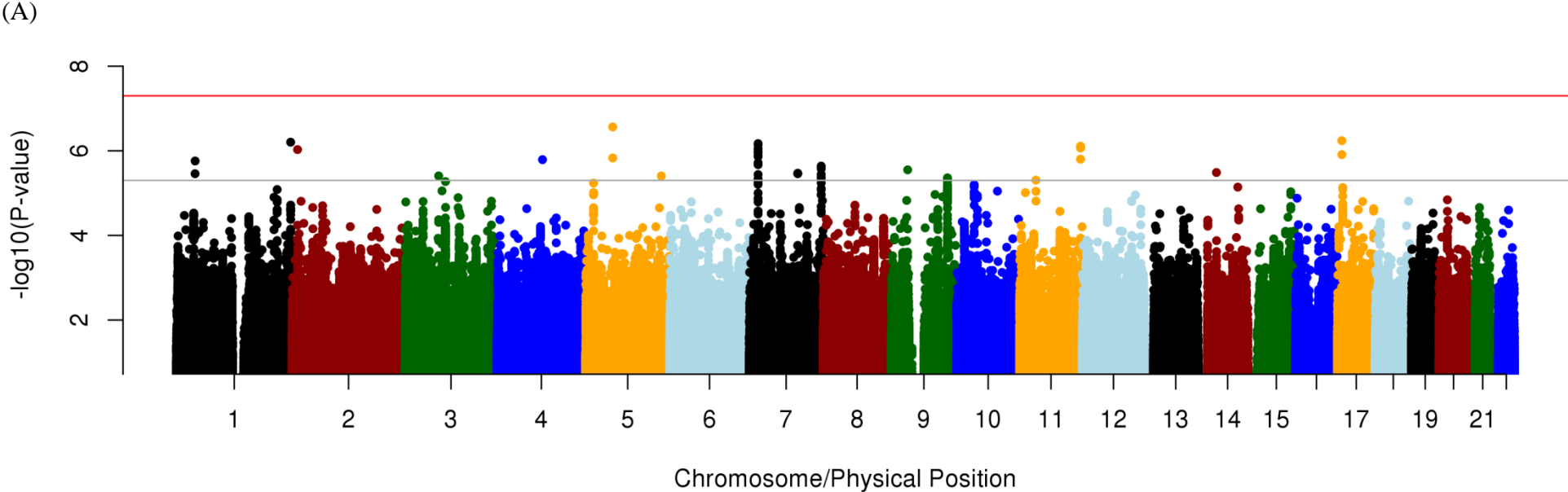
(C)



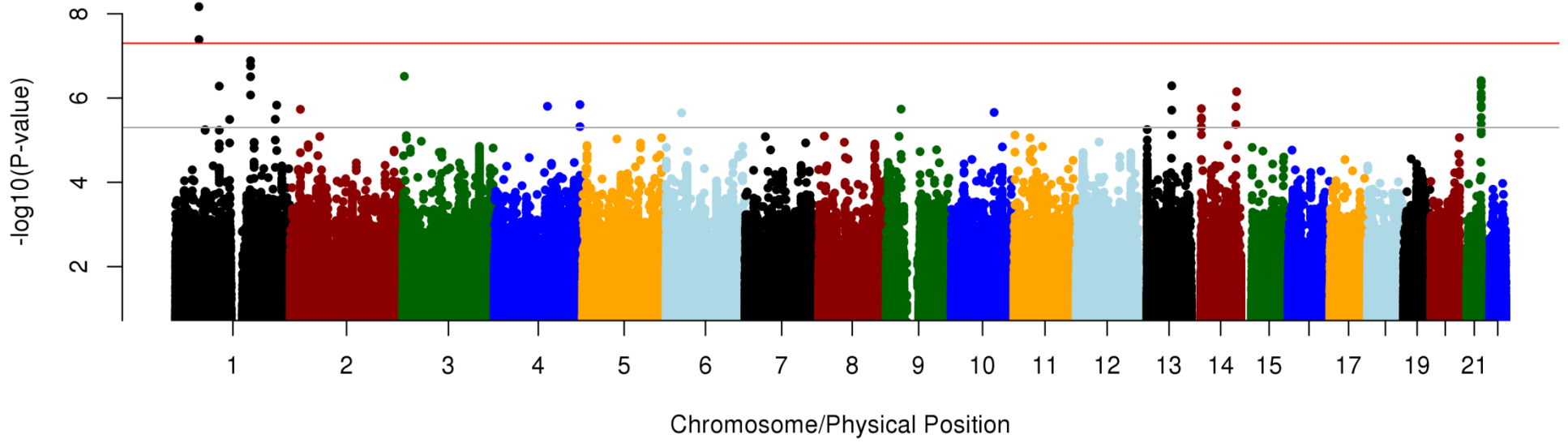
D)



Supplemental Figure III Trans-ethnic GWAS results of stroke subtypes (A) cardioembolic, (B) large artery atherosclerotic, (C) small vessel stroke and (D) undetermined.

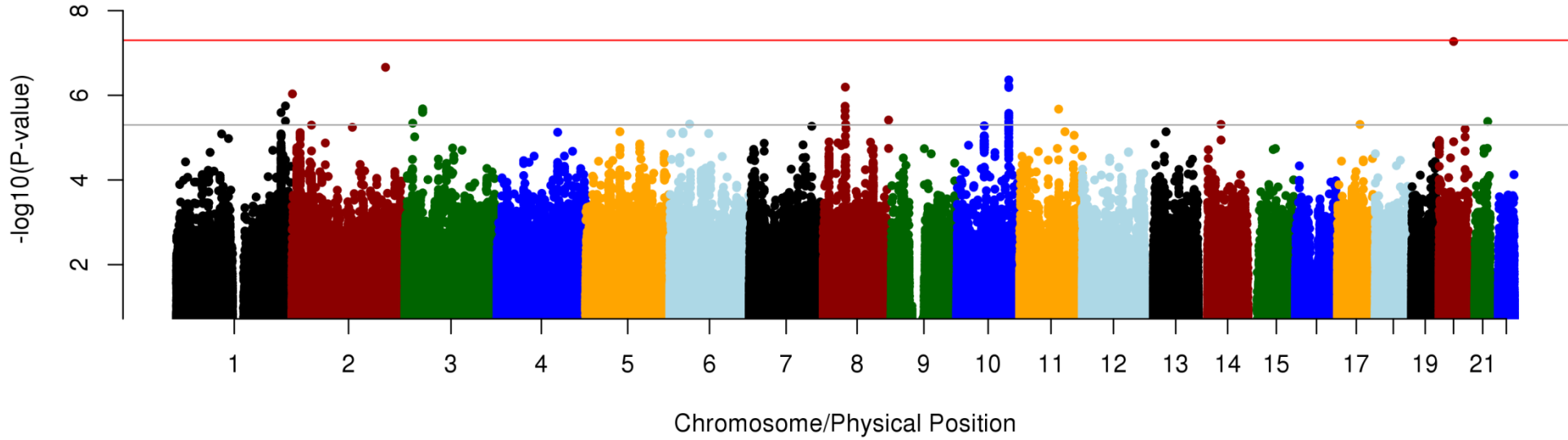


(B)

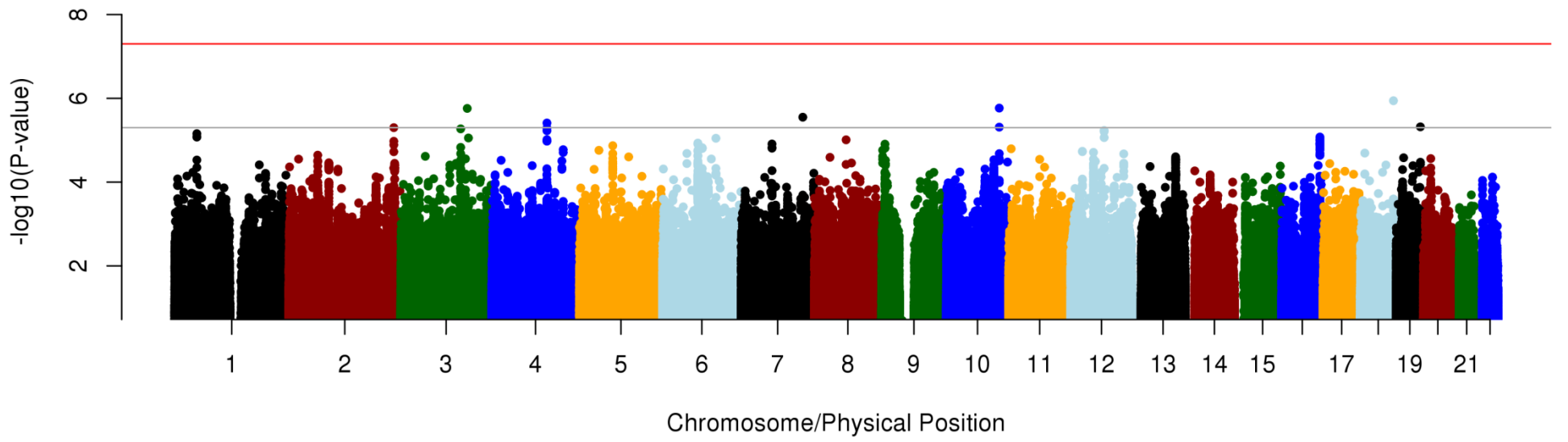


Supplemental Figure III. (continued)

(C)



(D)



Supplemental References

1. Adams Jr HP, Bendixen BH, Kappelle LJ, Biller J, Love BB, Gordon DL, et al. Classification of subtype of acute ischemic stroke. Definitions for use in a multicenter clinical trial. TOAST. Trial of Org 10172 in Acute Stroke Treatment. *Stroke*. 1993;24:35-41
2. Ay H, Benner T, Arsava EM, Furie KL, Singhal AB, Jensen MB, et al. A computerized algorithm for etiologic classification of ischemic stroke: the Causative Classification of Stroke System. *Stroke*. 2007;38:2979-2984
3. MacClellan LR, Mitchell BD, Cole JW, Wozniak MA, Stern BJ, Gile WH, et al. Familial aggregation of ischemic stroke in young women: the Stroke Prevention in Young Women Study. *Genet Epidemiol*. 2006;30:602-608
4. Kittner SJ, Stern BJ, Wozniak M., Buchholz DW, Earley CJ, Feeser BR, et al. Cerebral infarction in young adults: the Baltimore-Washington Cooperative Young Stroke Study. *Neurology*. 1998;50: 890-894
5. The ARIC Investigators. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. *Am J Epidemiol*. 1989;129:687-702
6. Rosamond,WD, Folsom AR, Chambless LE, Wang CH, McGovern PG, Howard G, et al. Stroke incidence and survival among middle-aged adults: 9-year follow-up of the Atherosclerosis Risk in Communities (ARIC) cohort. *Stroke*. 1999;30:736-743
7. The National Survey of Stroke. National Institute of Neurological and Communicative Disorders and Stroke. *Stroke*. 1981;2 Pt 2 Suppl 1:11-91
8. Ay H, Furie KL, Singhal A, Smith WS, Sorensen AG, Koroshetz WJ. (2005). An evidence-based causative classification system for acute ischemic stroke. *Annals Neurol*. 2005;58: 688-697
9. Yadav S, Schanz R, Maheshwari A, Khan MS, Slark J, de Silva R, et al. Bio-Repository of DNA in stroke (BRAINS): a study protocol. *BMC Med Genet*. 2011;12:34
10. Psaty BM, Heckbert SR, Atkins D, Lemaitre R, Koepsell TD, Wahl PW, et al. The risk of myocardial infarction associated with the combined use of estrogens and progestins in postmenopausal women. *Arch Intern Med*. 1994;154:1333-1339
11. Psaty BM, Heckbert SR, Koepsell TD, Siscovick DS, Raghunathan TE, Weiss NS, et al. The risk of myocardial infarction associated with antihypertensive drug therapies. *JAMA*. 1995;274:620-625
12. Klungel OH, Heckbert SR, Longstreth WT, Jr, Furberg CD, Kaplan RC, Smith NL, et al. Antihypertensive drug therapies and the risk of ischemic stroke. *Arch Intern Med*. 2001;161:37-43
13. Price TR, Psaty B, O'Leary D, Burke G, Gardin J. Assessment of cerebrovascular disease in the Cardiovascular Health Study. *Ann Epidemiol*. 1993;3:504-507
14. Meschia JF, Kissela BM, Brott TG, Brown RD Jr., Worrall BB, Beck J, Skarp AN The Siblings With Ischemic Stroke Study (SWISS): a progress report. *Clin Med & Res*. 2006;4:12-21

15. Chen DT, Worrall BB, Brown RD, Jr., Brott TG, Kissela BM, Olson TS, et al. The impact of privacy protections on recruitment in a multicenter stroke genetics study. *Neurology*. 2005;64:721-724
16. Meschia JF, Barrett KM, Chukwudelunzu F, Brown WM, Case LD, Kissela BM, et al. Interobserver agreement in the trial of org 10172 in acute stroke treatment classification of stroke based on retrospective medical record review. *J Stroke Cerebrovasc Dis*. 2006;15:266-272
17. Jood K, Ladenvall C, Rosengren A, Blomstrand C, Jern C. Family history in ischemic stroke before 70 years of age: the Sahlgrenska Academy Study on Ischemic Stroke. *Stroke*. 2005;36:1383-1387