Supporting Information

Malik et al. 10.1073/pnas.1616301114

SI Materials and Methods

Genotyping. Study case samples were genotyped at the Helmholtz Zentrum München (Deutsches Forschungszentrum für Gesundheit und Umwelt, Institute of Human Genetics). Genotype calling was carried out according to best practices from the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium (48). For the stroke-free controls, raw genotyping intensities were not available, thus prohibiting reclustering of variants. Genotypes were converted to PLINK (42) format, TOP-oriented. The strand orientation and the same coding for minor alleles were precisely controlled to match between cases and controls both within and between studies. Allele frequencies of all variants were compared with reported allele frequencies in the 1,000 Genomes phase I reference panel.

QC was performed separately for cases and controls according to best practices, merged, and aggregated into three files. Australian and UK cases were pooled because principal component (PC) analysis showed no significant difference in population structure. Samples with a call rate <95% per variant and one individual per pair of close relatives or duplicates (Pi-Hat > 0.5, calculated with PLINK on identity by descent markers) were excluded from analysis. Variants with a call rate <98%, Hardy Weinberg Equilibrium (HWE) P value < 10-6, MAF < 0.01%, mean heterozygosity > ±1 SD difference from mean in cases, and mean heterozygosity > ±3 SD in controls were excluded.

Ancestry-informative autosomal markers were LD-pruned (r^2 > 0.5) and carried forward to PC analysis to check the population genetic structure and identify outliers in cases and controls using R. Genetic outliers were identified by visual inspection of the PC plots (Fig. S1).

QC Characteristics. PC analysis partitioned the Caucasian samples into groups consistent with their geographical location. Thirteen German cases, 62 UK/Australian cases, 29 German controls, and 55 UK controls were removed during QC or after visual inspection of plotting the first two PCs (Fig. S1).

Association analyses were performed separately for the Caucasian and South Asian samples using an additive logistic model with the first two PCs as covariates to correct for population stratification. The results were then entered into a transethnic metaanalysis (Fig. 1). This strategy reduced the inflation factor to nominal levels without any sign of systemic inflation (Fig. S2). Inspection of the distribution of χ^2 (score) tests showed an overdispersion of test statistics (genomic control λ estimate of ∼1.21). However, after removal of very rare SNPs (MAF < 0.1%, or fewer than three alleles in cases), Wald statistics for the remaining SNPs appeared to be sampled from an overall central χ^2 distribution (λ = 1.06). For common (λ = 1.11 after removal of significant loci from this analysis), low-frequency (λ = 1.01), and rare (λ = 0.97) variants, we did not observe significant inflation of test statistics (Fig. S2).

The single-variant analysis covered 229,504 exonic variants, 2,262 splice site variants, 6,209 intronic variants, 78 5′-UTR variants, and 525 3′-UTR variants. A total of 92,772 variants passed QC procedures. The remaining variants were either monomorphic or had an allele frequency that was too low to yield meaningful test statistics (MAF < 0.1% or fewer than three minor alleles).

In the final transethnic metaanalysis, 66,229 SNPs had a MAF < 5% (Fig. 1). MAF distributions were similar across the German and UK/Australian controls and showed a correlation of R^2 = 0.98 compared with the 1000 Genomes phase 1 reference panel.

Gene-based association tests were performed using SKAT and SKAT-O (49, 50). In total, 17,168 genes harboring at least two rare variants within ±1 kb of gene boundaries were analyzed for association. Studies were analyzed separately and metaanalyzed using the R package metaSKAT (51).

For the SKAT and SKAT-O analyses, the total number of genes having at least two polymorphic variants passing QC was 17,168.

Association with Carotid Plaque Characteristics in Athero-Express. Human carotid atherosclerotic plaques (n = 1,414) were obtained from the Athero-Express study, an ongoing, longitudinal, multicenter study collecting carotid atherosclerotic plaques from patients with significant (>70%) stenosis who undergo carotid endarterectomy (52, 53). The medical ethics committee of the participating centers approved the study, and written informed consent was obtained from all patients.

Immunohistochemical plaque phenotyping was performed as described elsewhere (52, 53). In brief, carotid plaques were divided into segments of 5-mm thickness. The culprit lesion was defined and used for immunohistochemical staining. Calcification (hematoxylin and eosin) and collagen content (picrosirius red) were semiquantitatively defined as no or minor versus moderate/heavy staining as previously described (52, 53). Atheroma size was defined as less than or more than 10% fat content (hematoxylin and eosin and picrosirius red). We quantitatively scored macrophages (CD68) and smooth muscle cells (α-actin) as the percentage of plaque area. We also determined the presence of intraplaque hemorrhage (hematoxylin and eosin) and counted the number of vessels per three to four hotspots (CD34). Genotyping was done in two batches using Affymetrix arrays. Using phased data from HapMap 2 (release 22, b36) as a reference, rs6647 was imputed. QC was performed according to standard procedures. We adopted a significance level of P < 0.01 to account for testing multiple plaque phenotypes.

Expression and Purification of Recombinant Proteins. The cDNA modification of the human M1(A213) variant was introduced by PCR from a cDNA encoding the human M1(V213) variant (54) using the forward primer DJ3689 (5′-GACTTCCACGTGGACCAGGcGACCACCGTGAA-3′) and reverse primer DJ3613 (5′-GATGACCGGT TTTTTGGGTGGGATTCACCACTT-3′). Following digestion of the PCR product with PmlI and AgeI, the insert was cloned into the respective sites of the wild-type AAT pTT5 plasmid. Recombinant clones were identified by BstEII digestion because this site is lost by the V (GTG)-to-A (GCG) replacement. The cDNA constructs were expressed under identical conditions in HEK293 cells stably transfected with EBNA-1 cells (Yves Durocher, National Research Council Canada, Montreal) cultured in serum-free Free-Style 293 expression medium (Thermo Fisher Scientific, Inc.), 1% Pluronic, and G416 (25 μg/mL) at 37 °C and 8% CO_2. Harvesting of supernatants and purification of recombinant AAT variants were performed as described (54).

As a binding partner for AAT, we used a catalytically inactive NE (S195A variant). We changed Ser195 to Ala by inserting an oligoduplex [DJ3532 (5′-GTGAACGTATGCACTCTGGTGCCACGTCGGCAGGCAGGCATCTGCTTCGGGGACGCT-3′) and DJ3533 (5′-CGTCCCCGAAGCAGATGCCTGCCTGCCGACGTGGCACCAGAGTGCATACGTTCACAC-3′)] into the AlfI site of the previously described wild-type mouse NE construct in pTT5 (55). A cysteine tag (DDDCDDD) was added by insertion of an oligoduplex DJ3632 (5′-CTAGCGACGACGATTGCGACGATGATC-3′) and DJ3633 (5′-CTAGGATCATCGTCGCAATCGTCGTCG-3′) into the AvrII site.

Labeling of NE. To reduce all cysteine tags, we incubated recombinant NE in phosphate buffer [20 mM Na_2HPO_4, 300 mM NaCl (pH 7.4)] and 1 mM DTT for 2 h at room temperature. DTT was removed by precipitating NE with 75% ammonium sulfate. This reduction and precipitation step was repeated once. NE was dissolved in labeling buffer 1 [MO-L004 Monolith Protein Labeling Kit RED-MALEIMIDE (Cysteine Reactive); NanoTemper Technologies] and incubated with fivefold molar excess of dye [Alexa Fluor 647 NHS Ester (Succinimidyl Ester); Thermo Fisher Scientific] at room temperature for 1 h. The excess dye was removed with a PD MiniTrap G-10 column (GE Healthcare), used according to the manufacturer's instructions.

Determination of the Concentration of Recombinant AAT. Concentrations of active AAT were determined using active-site titrated human NE. A dilution series of AAT was incubated with 2.8 nM NE (Elastin Products) in 150 mM NaCl, 50 mM Tris, and 0.01% Triton X-100 (pH 7.4) at 4 °C for 1 h. Subsequently, the residual activity was determined by adding MCA-GEAIPSSIPPEVK(Dnp)-rr (EMC Microcollections) and measuring the fluorescence progression curve (excitation = 320 nm, emission = 405 nm). The initial cleavage rate was plotted against increasing AAT concentrations and decreased linearly with the amount of AAT added. By extrapolating this straight line, the molar concentration of functional AAT was determined at which the cleavage rate was zero. Each measurement was performed in duplicate, and the whole experiment was repeated twice.

Formula for Determination of Dissociation Constants.

$$FB = \frac{[AB]}{[B]} = \frac{[A] + [B] + K_D - \sqrt{([A] + [B] + K_D)^2 - 4[A][B]}}{2[B]},$$

where FB is fraction-bound, $[A]$ is the concentration of AAT, $[B]$ is the concentration of labeled catalytically inactive NE, and K_D is the dissociation constant.

Description of Individual Study Samples.

Australian Stroke Genetics Collaborative. 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 WHO 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 transesophageal echocardiogram, were conducted to define ischemic stroke (IS) mechanism as clinically appropriate. Cases were excluded from participation if patients were aged <18 y, diagnosed with hemorrhagic stroke or transient ischemic attack rather than IS, or were unable to undergo baseline brain imaging. IS subtypes were assigned using Trial of Org 10172 in Acute Stroke Treatment (TOAST) criteria, based on clinical, imaging, and risk factor data. All study participants provided informed consent for participation in genetic studies. Approval for the individual studies was obtained from relevant institutional ethics committees.

Wellcome Trust Case Control Consortium 2. Stroke cases included samples recruited by investigators at St. George's University London and the University of Oxford, the University of Edinburgh, and Ludwig Maximilians University Munich. The London collection comprised 1,224 IS samples from a hospital-based setting. All cases were of self-reported Caucasian ancestry. IS 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 IS cases, consecutively collected as part of the Oxford Vascular Study. Patients were of self-reported Caucasian ancestry, and IS subtypes were determined according to TOAST criteria based on relevant clinical imaging. The University of Edinburgh collection comprised 727 IS cases, consecutively collected as part of the Edinburgh Stroke Study. Patients were of self-reported Caucasian ancestry, with IS subtypes determined according to TOAST criteria based on relevant clinical and imaging data. The Munich samples included 1,383 IS cases. Patients were consecutive European Caucasians recruited from a single dedicated stroke unit at the Department of Neurology, Klinikum Grosshadern, Ludwig Maximilians University. IS subtypes were determined according to TOAST criteria based on relevant clinical and imaging data. The study was approved by the respective local ethics committees. All participants gave their informed consent.

Muenster (Westphalian stroke cases and controls from the Dortmund Health Study, Germany). Cases were recruited through hospitals participating in the Westphalian Stroke Registry, located in the west of the country. For the current analysis, patients recruited during the period 2000–2005 were included. The register's standardized patient documentation form included major stroke type and severity, comorbidities, and diagnostic and therapeutic details of the treatment process. IS was further subtyped according to the TOAST classification by the documenting physician. Patients who had experienced a transient ischemic attack or a hemorrhagic stroke were excluded from this analysis. The study was approved by the ethics committee of the University of Münster. All participants gave their informed consent.

Risk Assessment of Cerebrovascular Events Study, Pakistan. The Risk Assessment of Cerebrovascular Events (RACE) is a retrospective case-control study designed to identify and evaluate genetic, lifestyle, and biomarker determinants of stroke and its subtypes in Pakistan. Samples were recruited from six hospital centers in Pakistan. Patients were eligible for study inclusion if they met the following criteria: (*i*) aged at least 18 y; (*ii*) presented with a sudden onset of neurological deficit affecting a vascular territory, with sustained deficit at 24 h verified by medical attention within 72 h after onset (onset is defined by when the patient was last seen normal and not when found with deficit); (*iii*) the diagnosis was supported by CT/MRI; and (*iv*) presented with a modified Rankin score of <2 before the stroke. TOAST and Oxfordshire classification systems were used to subphenotype all stroke cases. Control participants were individuals enrolled in the Pakistan Risk of Myocardial Infarction Study (PROMIS), a case-control study of acute myocardial infarction based in Pakistan. Controls in the PROMIS were recruited following procedures, and inclusion criteria were as adopted for RACE cases. 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: (*i*) non–blood-related or blood-related visitors of patients of the outpatient department, (*ii*) non–blood-related visitors of stroke patients, and (*iii*) patients of the outpatient department presenting with minor complaints.

Control samples. Controls for the UK samples were drawn from shared Wellcome Trust Case Control Consortium controls obtained from the 1958 Birth Cohort. This cohort is a prospectively collected cohort of individuals born in 1958 and ascertained as part of the national child development study (www.cls.ioe.ac.uk/ncds). 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 Kooperative Gesundheitsforschung in der Region Augsburg (KORAgen) study (https://www.helmholtz-muenchen.de/en/kora/). This survey represents a gender- and age-stratified random sample of all German residents of the Augsburg area and consists of individuals 25–74 y of age, with about 300 subjects for each 10-y increment. All controls were free of a history of stroke or transient ischemic attack.

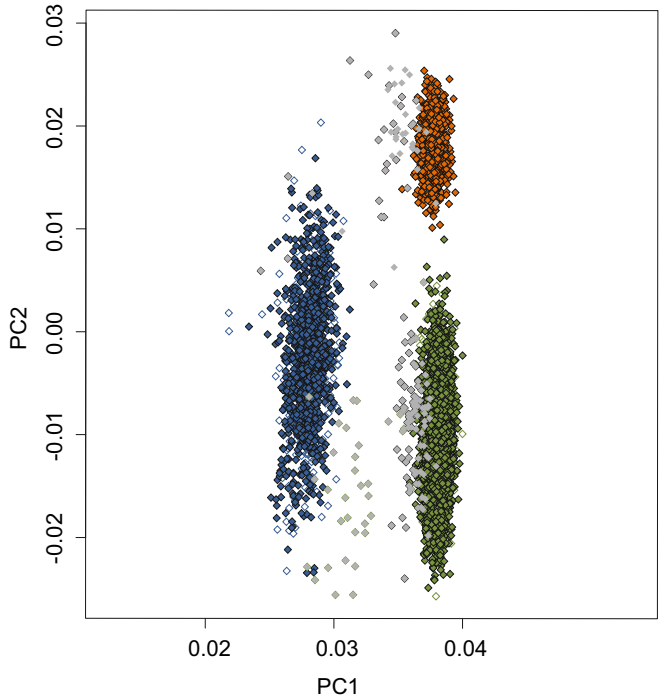

Fig. S1. Outlier analysis by PC analysis. German, UK, Australian, and South Asian samples were checked to remove potential outliers with respect to genetic background. Green with black outline, German cases; white with green outline, Kooperative Gesundheitsforschung in der Region Augsburg (KORA) controls; blue with black outline, UK/Australian cases; white with blue outline, Wellcome Trust Case Control Consortium 2 controls; orange with black outline, Pakistan cases; white with orange outline, Pakistan controls. Samples in gray were removed from the analysis after visual inspection. PC1 and PC2 are displayed on the *x* axis and *y* axis, respectively.

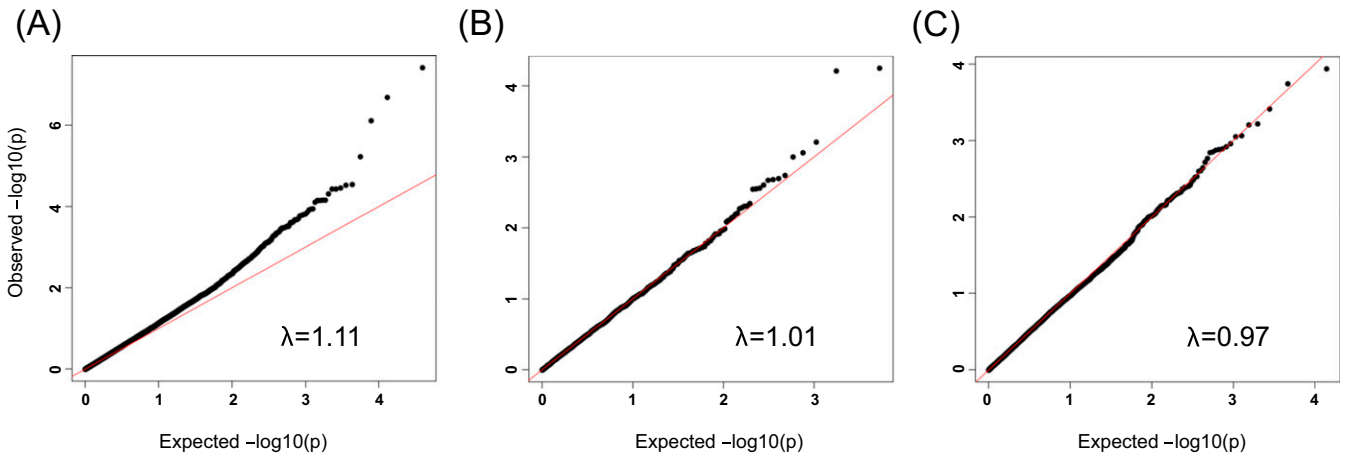

Fig. S2. Quantile–quantile (QQ) plots for single-variant analysis. Shown are the observed versus expected $-\log_{10}$ *P* value distributions for common (*A*), low-frequency (*B*), and rare (*C*) variants, along with the genomic inflation factors (λ). The red line shows the expected (null) distribution of the statistic.

Table S1. Overview of case and control samples

Study	No. of subjects	Age, mean ± SD	Male, %	Ancestry
Cases				
Munich	597	72.0 ± 9.4	54.0	Caucasian
Muenster	879	69.9 ± 6.4	67.1	Caucasian
Immunochip London	466	66.9 ± 8.3	67.5	Caucasian
Edinburgh	378	73.5 ± 9.7	57.4	Caucasian
Oxford	348	70.2 ± 6.5	57.4	Caucasian
ASGC	158	70.0 ± 13.0	69.6	Caucasian
RACE	376	48.2 ± 9.7	60.2	South Asian
Controls				
1958BC (WTCCC2)	5,963	NA	56.1	Caucasian
KORA	2,921	55.6 ± 13.2	52.0	Caucasian
PROMIS	978	NA	NA	South Asian

Shown are the LAS case and stroke-free control studies used for the current analysis. ASGC, Australian Stroke Genetics Consortium; KORA, Kooperative Gesundheitsforschung in der Region Augsburg; NA, information not available; WTCCC2, Wellcome Trust Case Control Consortium 2.

Table S2. Top 10 results of the gene-based test

Gene name	No. of SNPs in test	*P* value of metaSKAT test
B4GALNT4	2	2.28E-05
TTC29	3	4.08E-05
KANK4	2	4.38E-05
RSPO3	4	6.39E-05
SLC25A17	2	7.43E-05
CALM2	2	9.68E-05
TRPM4	3	0.00010258
SPTBN5	16	0.00014232
GSTA5	2	0.00017318
RPS6KL1	3	0.00020523

P values were derived from metaSKAT tests. The number of SNPs refers to the number of rare input SNPs into the individual SKAT tests before meta-SKAT.