

Supplementary Appendix

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

Supplement to: Ikram MA, Seshadri S, Bis JC, et al. Genomewide association studies of stroke. *N Engl J Med* 2009;360:1718-28.

List of Contents

The Supplementary Appendix has the following sections in order:

1. Abbreviations
2. Consortium organization and Sample selection
3. Stroke surveillance and classification
4. Genotyping quality control filters and imputation
5. Screening for latent population substructure
6. Meta-analysis techniques
7. Replication Samples
8. Candidate gene selection and results
9. Population attributable risk
10. Reference list

Figures

Figure 1A and B: Quantile-quantile (QQ)-plot showing the observed versus the expected p-values after meta-analysis for total stroke and ischemic stroke respectively

Figure 2: Forest plots showing associations of rs11833579 (A, C) and rs12425791 (B, D) with total and ischemic stroke based on genotyped and imputed data from GWAS.

Figure 3: Linkage-disequilibrium plot of SNPs adjacent to the *NINJ2* gene that were significantly associated with total stroke and ischemic stroke and SNPs at the 5' end of the *NINJ2* gene

Tables

Table 1: Results for GWAS associations with total stroke at a meta-analysis p-value lower than 10^{-4}

Table 2: Results for GWAS associations with ischemic stroke at a meta-analysis p-value lower than 10^{-4}

Table 3: Results of NINJ2 associations with total and ischemic stroke adjusted for age-, sex- and various stroke risk factors

Table 4: Characteristics of Participants in Replication Cohorts

Table 5: Association of rs12425791 and rs11833579 with various stroke related phenotypes in discovery and replication cohorts

Table 6: Associations with Putative Candidate Stroke SNPs in the CHARGE meta-analysis

1. Abbreviations

AGES	Aging Gene-Environment Study
ARFGAP2	ADP-ribosylation factor GTPase activating protein 2
API5	Apoptosis Inhibitor 5
ARIC	Atherosclerosis Risk in Communities
AVEN	apoptosis caspase activation inhibitor
BBS10	Bardet-Biedl syndrome 10
C1QTNF7	C1q and tumor necrosis factor related protein 7
C5orf23	chromosome 5 open reading frame 23
C8orf79	chromosome 8 open reading frame 79
C10orf88	chromosome 10 open reading frame 88
C11orf49	chromosome 11 open reading frame 49
CD36	CD36 molecule (thrombospondin receptor)
CHARGE	Cohorts for Heart and Aging Research in Genomic Epidemiology
CHS	Cardiovascular Health Study
CI	Confidence Interval
CKAP5	cytoskeleton associated protein 5
CPEB2	cytoplasmic polyadenylation element binding protein 2
DDB2	Damage-specific DNA binding protein 2
DLC1	deleted in liver cancer 1
EFCAB4B	EF-hand calcium binding domain 4B
F2	coagulation factor II (thrombin)
FAM24A	family with sequence similarity 24, member A
FHS	Framingham Heart Study
FIP1L1	FIP1 like 1 (<i>S. cerevisiae</i>)
FUT8	fucosyltransferase 8
GWAS	Genome Wide Association Study
GPHN	gephyrin
HNT	neurotrimin
HR	Hazard Ratio
LD	Linkage Disequilibrium
LNX1	ligand of numb-protein X 1
LRP4	Low density lipoprotein receptor-related protein
LRRC4C	leucine rich repeat containing 4C
MA	Minor Allele
MAF	Minor Allele Frequency
MAF	v-maf musculoaponeurotic fibrosarcoma oncogene homolog (avian)
MPDZ	multiple PDZ domain protein
NFIB	nuclear factor I/B
NINJ2	ninjurin 2
NPR3	natriuretic peptide receptor C/guanylate cyclase C (atrionatriuretic

	peptide receptor C)
OPCML	opioid binding protein/cell adhesion molecule-like
OSBPL8	oxysterol binding protein-like 8
P2RY5	purinergic receptor P2Y, G-protein coupled, 5
PACSIN3	Protein kinase C and casein kinase substrate in neurons 3
PAR	Population Attributable Risk
PARP11	poly (ADP-ribose) polymerase family, member 11
QC	Quality Control
RB1	retinoblastoma 1
RYR3	ryanodine receptor 3
SD	Standard Deviation
SEMA3C	sema domain, immunoglobulin domain (Ig), short basic domain, secreted, (semaphorin) 3C
SNP	Single Nucleotide Polymorphism
TP63	tumor protein p63
TPRG1	tumor protein p63 regulated 1
WNK1	WNK lysine deficient protein kinase 1

2. Consortium organization and Sample selection

We combined data from white participants in four large, prospective population-based cohort studies: the Atherosclerosis Risk in Communities (ARIC) Study,¹ the Cardiovascular Health Study (CHS),² the Framingham Heart Study (FHS),^{3,4} and the Rotterdam Study⁵ to study the genetics of stroke. Together with the Aging Gene-Environment Susceptibility- Reykjavik Study (AGES-RS)⁶ these studies form the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium.⁷ However, the AGES-RS does not have data on incident stroke and hence did not contribute to this meta-analysis.

All participating studies approved guidelines for collaboration, and a neurology working-group arrived at a consensus on phenotype harmonization, covariate selection and analytic plans for within-study analyses and meta-analysis of results. Each study has an Institutional Review Board that approved the consent procedures, examination and surveillance components, data security processes, genotyping protocols and current study design.

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. Participants were handed a card and asked to tell the interviewer which best described his or her race. Choices offered were: white, black, american indian/ alaskan native, asian/pacific islander, other: specify. Over 99% identified as either white or black. Only self-identified whites were included in the discovery cohort. 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.¹ Genotyping was performed at the Broad Institute using the Affymetrix[®] Genome-Wide Human SNP Array 6.0. As of June 2008, genotyping had been completed on 8,861 white participants. Individuals with and without genotype data did not significantly differ with regards to baseline CVD risk factors (not shown). Only individuals free of stroke or TIA at baseline were included (n=8,377). We excluded individuals with a sex mismatch (n=40), those who were discordant for more than 5% of genotypes among 47 previously genotyped overlapping SNPs (n=157), persons who were 1st degree relatives (n=314), those who were outliers based on average identity by state (n=131) or based on Eigenstrat clustering (n=250), or persons who had an incident SAH (n=18); some individuals met more than one exclusion criterion. In total, 7,686 individuals were included in the present analyses.

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.² 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 blacks were enrolled subsequently (1992-93) for a total sample of 5,888. As in the ARIC cohort, 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. Only self-identified blacks were excluded from the discovery cohort. DNA was extracted from blood samples drawn on all participants who consented to genetic testing at their baseline examination. In 2007-2008, genotyping was performed at the General Clinical Research Center's Phenotyping/Genotyping Laboratory at Cedars-Sinai using the Illumina 370CNV Duo[®] BeadChip system on the first 2,427 of 3,980 CHS participants who were free of CVD at baseline. The 1,908 persons excluded for prevalent CVD had prevalent coronary heart disease (n=1195), congestive heart failure (n=86), peripheral vascular disease (n=93), valvular heart disease (n=20), stroke (n=166) or transient ischemic attack (n=56). Some persons had more than one reason to be excluded and for these individuals only the initial exclusionary event is listed. The 2,427 participants who had been genotyped by July 2008 (time of the meta-analysis) were a stratified probability sample that included all cases of myocardial infarction (MI), stroke, atrial fibrillation, dementia, and heart failure, and a selection of controls sampled based on the age-, and sex-distribution of MI cases. Sampling weights were used in the analysis to account for the study design and to weight back to the underlying cohort selected for genotyping (N=3,980). Because the other cohorts were predominantly white, the African American participants were excluded from the initial meta-analysis to limit the potential for false positive associations due to population stratification. At the time of the meta-analysis, genotyping had been attempted in 2,101 white participants, and was successful in 2,022

persons; the latter constitute the CHS sample for this study. QC criteria used to define successful genotyping are listed in Section 3, 35 persons were excluded for a subject-specific call rate $\leq 95\%$.

FHS

The 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;³ their Offspring and spouses of the Offspring, followed since 1971;⁴ and children from the largest Offspring families enrolled in 2000 (Gen 3).⁸ Gen 3 participants were not included in this analysis since they are young (mean age 40 ± 9 years) and few have suffered strokes. The Original cohort enrolled 5209 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 5124 persons (including 3514 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 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 also asked to identify their ethnicity as either 'Hispanic or Latino' or neither. Almost all

the FHS Original and Offspring participants are white/Caucasian and none were excluded from the discovery cohort. FHS participants had DNA extracted and provided consent for genotyping in the 1990s. All available eligible participants were genotyped at Affymetrix (Santa Clara, CA) through an NHLBI funded SNP-Health Association Resource (SHARe) project using the Affymetrix GeneChip[®] Human Mapping 500K Array Set and 50K Human Gene Focused Panel.[®] In 272 persons (31 with stroke), small amounts of DNA were extracted from stored whole blood and required whole genome amplification prior to genotyping. Cell lines were available for most of the remaining participants. Genotyping was attempted in 5293 participants, and 4,519 persons met QC criteria. Failures (call rate <97%, extreme heterozygosity or high Mendelian error rate) were largely restricted to persons with whole-genome amplified DNA and DNA extracted from stored serum samples. We also excluded 156 participants who were less than 45 years old at the time of the DNA draw, 135 persons with prevalent stroke and 97 persons who did not have stroke surveillance on follow-up after genotyping; the remaining 4,131 subjects constitute the FHS sample for this study.

Rotterdam Study

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.⁵ All inhabitants aged ≥ 55 years ($n = 10,275$) were invited and the participation rate was 78%, yielding a total of 7983 subjects. All participants gave written informed consent to retrieve information from treating physicians. Baseline

measurements were obtained from 1990 to 1993 and consisted of an interview at home and two visits to the research center for physical examination. At this baseline examination ancestry was determined by self-report. Participants were asked to identify with one of the following categories that best described their ancestry: Dutch, Caucasian, Asian, Indian, Indonesian, Mediterranean, Negroid. Less than 1% of participants chose an ancestry other than Dutch or Caucasian. All Rotterdam study participants were included in the discovery cohort. Survivors have been re-examined three times: in 1993-1995, 1997-1999, and 2002-2004. All persons attending the baseline examination in 1990-93 consented to genotyping and had DNA extracted. This DNA was genotyped using the Illumina Infinium II HumanHap550chip v3.0[®] array in 2007-2008 according to the manufacturer's protocols. Genotyping was attempted in persons with high-quality extracted DNA (n=6449). From these 6449, samples with low call rate (<97.5%, n=209), with excess autosomal heterozygosity (> 0.336, n=21), with sex-mismatch (n=36), or if there were outliers identified by the IBS clustering analysis (>3 standard deviations from population mean, n=102 or IBS probabilities > 97%, n=129) were excluded from the study population with some persons meeting more than one exclusion criterion; in total, 5974 samples were available with good quality genotyping data. Of these, 170 persons were excluded for prevalent stroke and 41 persons for lack of adequate follow-up after genotyping. The remaining 5763 persons constitute the Rotterdam Study sample for the present analyses.

3. Stroke surveillance and classification

Stroke Surveillance

ARIC:

Persons were screened for possible stroke events through annual phone interviews, follow-up examinations, community hospital surveillance, and death certificates. Any reported hospitalization led to screening and, if suitable, to medical record abstraction. Potential stroke events were selected for abstraction of records if the discharge diagnosis included a cerebrovascular disease code (International Classification of Diseases, 9th Revision, codes 430 to 438), if a cerebrovascular procedure was mentioned in the discharge summary, or if the CT or MRI report showed evidence of acute cerebrovascular disease.⁹ All suspected events were classified by computer algorithm and also by an expert physician reviewer, blinded to the automated results. A second physician reviewer adjudicated disagreements between the computer and the initial reviewer.

CHS

Participants were examined annually from enrollment to 1999, and since then continue to be under surveillance for stroke.^{10, 11} 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.

FHS

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;¹²⁻¹⁴ they include 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 a large number of stroke events in the FHS (although ascertained prospectively) were prevalent at the time of DNA collection and

were excluded from these analyses. While this reduced the sample size from FHS, the meta-analyses presented here focused on incident events.

Rotterdam

All participants have been continuously monitored for major events (including stroke) through automated linkage of the study database with files from general practitioners and the municipality. In addition physician files from nursing homes and general practitioner records of participants who moved out of the Ommoord district were reviewed twice a year. For suspected stroke and TIA events, additional information (including neuroimaging) was obtained from hospital records and research physicians discussed available information with an experienced stroke neurologist to verify all diagnoses.^{15, 16}

Stroke Classification

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. In the Rotterdam study which did not subtype ischemic strokes as ABI or cardioembolic, we excluded those ischemic strokes that occurred in the presence of concomitant atrial fibrillation. For the analysis of ABI, participants were censored in the ARIC, CHS and FHS when they developed an alternative type of stroke and in the Rotterdam study if they developed atrial fibrillation.

Neuroimaging was available for 98% of all events in ARIC, 87% of all events in CHS, 90% of all events in FHS and 65% of all events in the Rotterdam study. In the current analyses we included events through December 31st 2004 for ARIC and the Rotterdam Study, and through December 31st 2005 for the CHS and FHS studies.

4. Genotyping quality control filters and imputation:

Subject specific quality control filters included filters for call rate, heterozygosity, and number of Mendelian errors per individual. SNP specific quality control filters included filters for call rate, minor allele frequency, Hardy-Weinberg equilibrium, and differential missingness by outcome or genotype (mishap test in PLINK);

<http://pngu.mgh.harvard.edu/purcell/plink/>).

Imputation

The set of genotyped input SNPs used for imputation in each study was selected based on their highest quality GWA data. We used a callrate >95% in CHS and ARIC; >97% in FHS and >98% in Rotterdam; a minor allele frequency >0.01 in each study; a Hardy-Weinberg $p > 1 \times 10^{-5}$ in CHS and $p > 1 \times 10^{-6}$ in ARIC, FHS, and Rotterdam; and a test of differential missingness by the “mishap” test in PLINK $p > 1 \times 10^{-9}$ in each study. We used either the Markov Chain Haplotyping (MaCH) package

(<http://www.sph.umich.edu/csg/abecasis/MACH>) version 1.0.15 software (ARIC, FHS and Rotterdam; imputed to plus strand of NCBI build 36, HapMap release #22) or BIM-BAM¹⁷ (CHS, imputed to plus strand of NCBI build 35) programs. For each imputed SNP a reliability of imputation was estimated (as the ratio of the empirically observed dosage variance to the expected binomial dosage variance: O/E ratio). For the primary meta-analysis using inverse-variance weighting less weight is given to imputed SNPs with low observed dosage variance (resulting in higher variance of the estimate). For the secondary meta- analysis using the inverse square root (N) weighting the ratio was used to compute an effective sample size.

Direct Genotyping

We used the Taqman system to directly genotype SNPs that reached genome-wide significance in the initial GWAS. Rs 11833579 had been imputed in all 4 cohorts and although the estimated reliability of the imputation had been high in three of the four cohorts (O/E ratio: 0.94 to 0.97) we directly genotyped this SNP in all 4 discovery cohorts and in the ARIC African-American replication sample. Rs1245791 had been genotyped by the platforms used in the ARIC, FHS and Rotterdam studies and imputed in the CHS alone (O/E ratio of 0.47), hence it was directly genotyped in CHS. Both SNPs were also genotyped in a small sample of 574 CHS African-Americans and in the white clinical case-control series. Primers and probe sequences may be obtained from the authors on request.

5. Screening for latent population substructure

All four discovery cohorts were screened for latent population substructure (including cryptic relatedness) using suitable programs (i.e, EIGENSTRAT in ARIC and FHS,^{18, 19} an IBD matrix in Rotterdam,²⁰ and using principal components analysis in CHS) and either found no evidence of occult population admixture, or found that the principal components identified were not related to the stroke phenotype.

We studied quantile-quantile (Q-Q) plots to ensure that the p-value distributions in each of the four cohorts conformed to a null distribution at all but the extreme tail. We also calculated the genomic inflation factor (λ), which measures over-dispersion of test-statistics from association tests indicating population stratification and can be used to apply genomic control.²¹ The λ for total stroke was 1.010 for ARIC, 1.053 for CHS, 1.029 for FHS, and 0.999 for the Rotterdam Study. For ischemic stroke it was 1.009 for ARIC, 1.064 for CHS, 1.037 for FHS, and 1.003 for the Rotterdam Study.

6. Meta-analysis techniques

After quality control and filtering within each study, the ARIC had either genotyped or imputed data for 2,448,797 SNPs, CHS for 2,531,169 SNPs, the FHS for 2,540,223 SNPs, and the Rotterdam study for 2,543,887 SNPs. SNPs with a minor allele frequency ≤ 0.005 were excluded. We restricted the present meta-analysis to the 2,194,468 autosomal SNPs common to all 4 studies. Our primary meta-analysis technique was inverse-variance weighting (also known as fixed-effects meta-analysis) after applying genomic control within each individual study. Beta estimates were weighted by their inverse variance and a combined estimate was obtained by summing the weighted betas and dividing by the summed weights. Hence results for SNPs imputed with low certainty were down-weighted because the low reliability-of-imputation (see Section 4) ensures a large variance. In contrast, studies with large sample sizes and with directly genotyped or well-imputed SNPs had a greater effect on the meta-analysis p-value because of small variances. For verification purposes we also undertook meta-analyses using inverse square root (N) weighting without genomic control and these results are included in the Supplementary tables 1 and 2. We undertook the meta-analyses at the Rotterdam study using the R-library, MetABEL (<http://mga.bionet.nsc.ru/~yurii/ABEL/>).²²

We estimated the genomic inflation factor (λ) after meta-analysis. The estimate of λ was 1.012 for total stroke and 1.006 for ischemic stroke indicating no significant inflation of p-values. The quantile-quantile (Q-Q) plot of our inverse variance meta-analysis results for total stroke (Figure 1A) and ischemic stroke (Figure 1B) show the distribution of the observed test statistic (negative log of p-values, on the y-axis) plotted

against the distribution of test statistic expected under the null-hypothesis (on the x -axis). The observed p -values conform to the null distribution across the entire range until $p < 1.0 \times 10^{-5}$; we expect that true positive associations reside beyond this region and that population substructure is negligible.

7. Replication Samples:

Our original African-American replication sample consisted of a subset of the cohort of 4,266 African-American participants from the ARIC study. The overall design of the ARIC study is described in Section 2, and the methods used for stroke surveillance and classification are described in Section 3 of the Supplementary Appendix.^{1,7} As of June 2008, genotyping had been completed on a sample of 2,572 African-American participants drawn randomly from the total ARIC African-American cohort. Only individuals free of stroke or TIA at baseline were included in the analyses (n=2,483). An additional 8 individuals were excluded for incident SAH. Finally, 45 individuals were excluded for failure to meet QC criteria; these individuals had a sex mismatch or were discordant for more than 5% of 47 previously genotyped SNPs. The remaining 2,430 individuals were used in the replication analyses.

Our second replication sample comprised 652 persons with prevalent stroke (313 women, mean age 76.5 ± 7.1 , 501 of these persons had an ischemic type of stroke and of these 400 had an atherothrombotic stroke) and age-matched controls. The stroke cases included persons with prevalent stroke in the Rotterdam Study at baseline, and a clinical series of ischemic strokes obtained from three hospitals. The first was the Erasmus MC University Medical Center, Rotterdam (n=271, 246 atherothrombotic), the second was the St Elisabeth Hospital, Tilburg (n=99, of which 76 were atherothrombotic), and the third was the TweeSteden Hospital, Tilburg (n=131, of which 31 were atherothrombotic). These clinical databases aim to collect clinical information, blood samples and DNA in all ischemic stroke patients evaluated at these hospitals. Controls were drawn from the group of Rotterdam Study participants who were stroke-free at baseline and were within

the same age range as the cases. These persons contributed one set of person-years to the prevalent analysis and a second, non-overlapping set of person-years to the incident analyses. Therefore the two analyses are independent of each other, this was confirmed with simulation experiments. A third, small and underpowered sample that was genotyped consisted of 574 African-Americans in the CHS who were free of prevalent cardiovascular disease at baseline and had been followed prospectively for incident stroke. Over a 10 year follow-up period a total of 85 incident strokes (68 ischemic, 57 atherothrombotic) were observed. Baseline characteristics of the replication cohorts are shown in Supplementary Table 4.

8. Candidate genes: selection and results

SNP selection

We searched Pubmed for original articles published before July 1st 2008, which reported statistically significant associations between specific genotypes and stroke. Since over 200 putative candidate genes for stroke have been described in the literature, we focused on genes identified as potential candidates for total stroke and ischemic stroke in whites by meta-analyses of the published literature. We used the search terms *meta-analysis*, *stroke* and *gene* and identified a total of 41 publications; we retrieved and reviewed all of them. We identified 12 meta-analyses,²³⁻³⁴ each of which studied white participants, and reported a significant association of at least one SNP within a gene with the risk of either total stroke or ischemic stroke in the pooled analyses. A total of 9 candidate genes were identified by this process. We selected SNPs that had been specifically examined in these meta-analyses, ignoring data based on linkage studies. Based on these criteria we studied selected SNPs in the following genes: 5,10-methylene tetrahydrofolate reductase (*MTHFR*),²⁴ Factor 5 Leiden (*F5*),²⁴ phosphodiesterase D (*PDE4D*)^{31, 33, 34}, plasminogen activator inhibitor type 1 or serpin peptidase inhibitor (*SERPINE1*),^{23, 29} prothrombin (*F2*),²⁴ glycoprotein 1b alpha (*GP1BA*),²⁷ arachidonate 5-lipoxygenase-activating protein (*ALOX5AP*), apolipoprotein E (*APOE*).^{28, 32, 34} We also selected a proxy SNP for the ACE insertion/deletion that has been associated with the risk of ischemic stroke.³⁵ Finally, we selected two intergenic SNPs at chromosome 4q25 and 9p21 that have been related previously to atrial fibrillation and coronary artery disease, respectively, and more recently to cardioembolic (rs2200733)³⁶ and thrombotic (rs1537378) strokes.

Results are presented in Supplementary Table 6 below. Rs16954257 in the *GPIBA* gene, a proxy SNP for rs2243093 previously associated with stroke,²⁷ was significantly associated with total stroke (p=0.02) and ischemic stroke (p=0.02) in our meta-analysis. For the prothrombin (*F2*) gene neither the reported SNP nor a proxy SNP with LD>0.8 was genotyped or imputed in our dataset.²⁴ However, six other SNPs within that gene (rs3136447, rs2070852, rs5896, rs3136457, rs3136480 and rs2282687) reached p<10⁻⁴ for total stroke (see Supplementary Table 1 rows 206 to 211).

9. Population attributable risk

Population attributable risk is an index incorporating both frequency of a risk factor and the risk ratio of outcome with and without the risk factor. It represents the proportion of a specific disease in the population attributable to this risk factor. Estimation of PAR is usually more accurate in cohort studies than in case/control studies where the odds-ratio may overestimate risk. The PAR for individual SNPs such as rs11833579 and rs12425791 was calculated as follows:

$$PAR = 1 - \frac{1}{q^2 + 2pq(HR) + p^2(HR^2)}$$

where q = major allele frequency; p = minor allele frequency and HR =the hazard ratio from the meta-analysis based on an additive model for increasing dose of the risk allele.³⁷

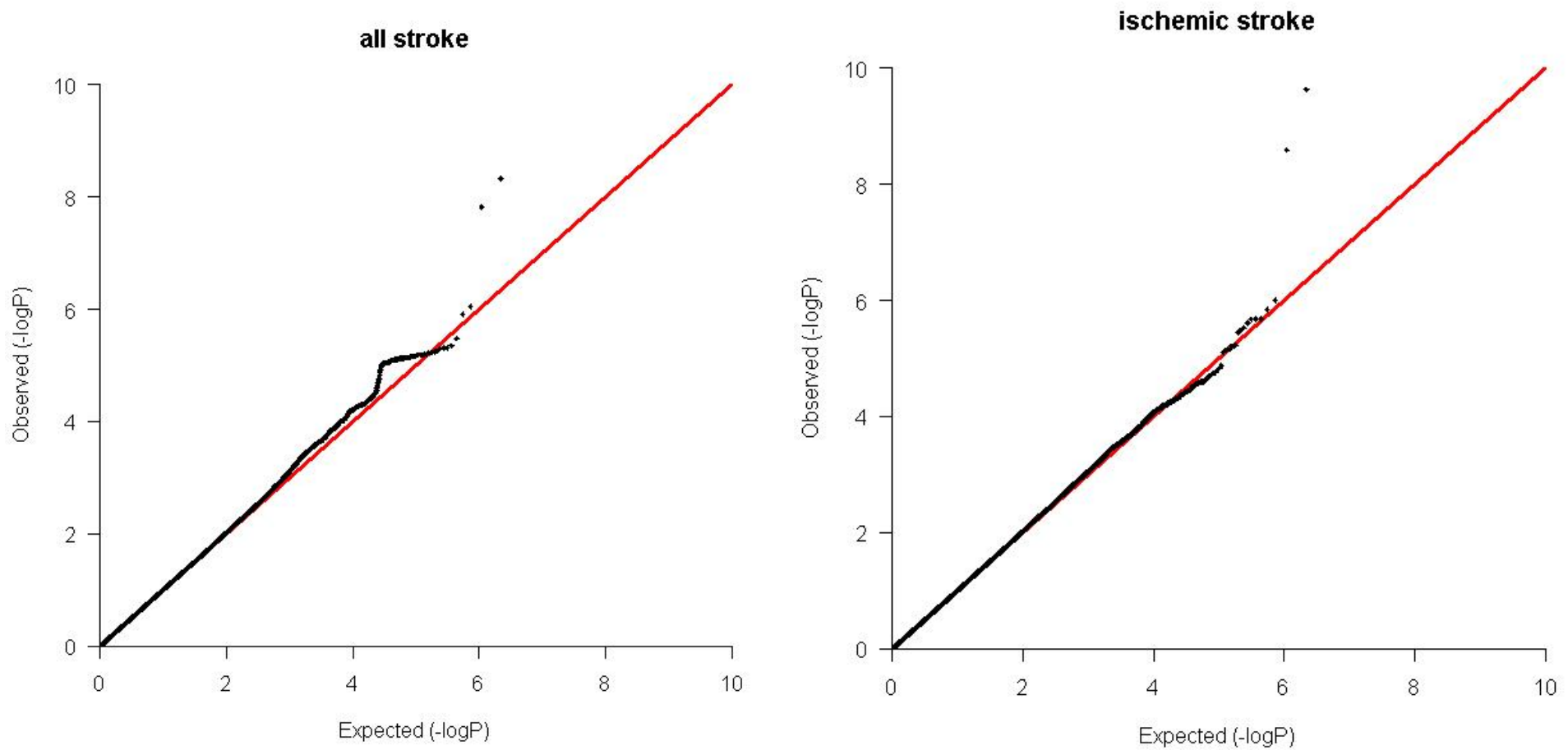
11. Reference List

1. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. *Am J Epidemiol* 1989;129:687-702.
2. Fried LP, Borhani NO, Enright P et al. The Cardiovascular Health Study: design and rationale. *Ann Epidemiol* 1991;1:263-276.
3. Dawber TR, Kannel WB. The Framingham study. An epidemiological approach to coronary heart disease. *Circulation* 1966;34:553-555.
4. Feinleib M, Kannel WB, Garrison RJ, McNamara PM, Castelli WP. The Framingham Offspring Study. Design and preliminary data. *Prev Med* 1975;4:518-525.
5. Hofman A, Breteler MM, van Duijn CM et al. The Rotterdam Study: objectives and design update. *Eur J Epidemiol* 2007;22:819-829.
6. Harris TB, Launer LJ, Eiriksdottir G et al. Age, Gene/Environment Susceptibility-Reykjavik Study: multidisciplinary applied phenomics. *Am J Epidemiol* 2007;165:1076-1087.
7. Psaty BM, O'Donnell CJ, Gudnason V et al. Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: Design of prospective meta-analyses of genome-wide association studies from five cohorts. *Circ Cardiovasc Genet* 2009; 2:73-80.
8. Splansky GL, Corey D, Yang Q et al. The Third Generation Cohort of the National Heart, Lung, and Blood Institute's Framingham Heart Study: design, recruitment, and initial examination. *Am J Epidemiol* 2007;165:1328-1335.
9. Rosamond WD, Folsom AR, Chambless LE et al. Stroke incidence and survival among middle-aged adults: 9-year follow-up of the Atherosclerosis Risk in Communities (ARIC) cohort. *Stroke* 1999;30:736-743.
10. Longstreth WT, Jr., Bernick C, Fitzpatrick A et al. Frequency and predictors of stroke death in 5,888 participants in the Cardiovascular Health Study. *Neurology* 2001;56:368-375.
11. 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.
12. Carandang R, Seshadri S, Beiser A et al. Trends in incidence, lifetime risk, severity, and 30-day mortality of stroke over the past 50 years. *JAMA* 2006;296:2939-2946.
13. Seshadri S, Beiser A, Kelly-Hayes M et al. The lifetime risk of stroke: estimates from the Framingham Study. *Stroke* 2006;37:345-350.

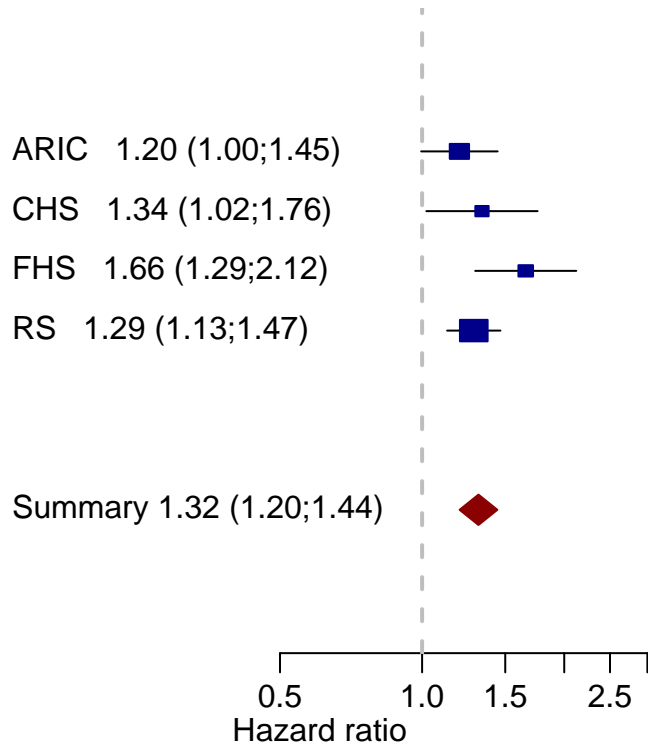
14. **Wolf PA, Kannel WB, Dawber TR. Prospective investigations: the Framingham study and the epidemiology of stroke. *Adv Neurol* 1978;19:107-120.**
15. **Bots ML, Looman SJ, Koudstaal PJ, Hofman A, Hoes AW, Grobbee DE. Prevalence of stroke in the general population. The Rotterdam Study. *Stroke* 1996;27:1499-1501.**
16. **Hollander M, Koudstaal PJ, Bots ML, Grobbee DE, Hofman A, Breteler MM. Incidence, risk, and case fatality of first ever stroke in the elderly population. The Rotterdam Study. *J Neurol Neurosurg Psychiatry* 2003;74:317-321.**
17. **Servin B, Stephens M. Imputation-based analysis of association studies: candidate regions and quantitative traits. *PLoS Genet* 2007;3:e114.**
18. **Patterson N, Price AL, Reich D. Population structure and eigenanalysis. *PLoS Genet* 2006;2:e190.**
19. **Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006;38:904-909.**
20. **Richards JB, Rivadeneira F, Inouye M et al. Bone mineral density, osteoporosis, and osteoporotic fractures: a genome-wide association study. *Lancet* 2008;371:1505-1512.**
21. **Bacanu SA, Devlin B, Roeder K. The power of genomic control. *Am J Hum Genet* 2000;66:1933-1944.**
22. **Aulchenko YS, Ripke S, Isaacs A, van Duijn CM. GenABEL: an R library for genome-wide association analysis. *Bioinformatics* 2007;23:1294-1296.**
23. **Attia J, Thakkestian A, Wang Y et al. The PAI-1 4G/5G gene polymorphism and ischemic stroke: an association study and meta-analysis. *J Stroke Cerebrovasc Dis* 2007;16:173-179.**
24. **Casas JP, Hingorani AD, Bautista LE, Sharma P. Meta-analysis of genetic studies in ischemic stroke: thirty-two genes involving approximately 18,000 cases and 58,000 controls. *Arch Neurol* 2004;61:1652-1661.**
25. **Cronin S, Furie KL, Kelly PJ. Dose-related association of MTHFR 677T allele with risk of ischemic stroke: evidence from a cumulative meta-analysis. *Stroke* 2005;36:1581-1587.**
26. **Kelly PJ, Rosand J, Kistler JP et al. Homocysteine, MTHFR 677C-->T polymorphism, and risk of ischemic stroke: results of a meta-analysis. *Neurology* 2002;59:529-536.**

27. Maguire JM, Thakkinstian A, Sturm J et al. Polymorphisms in platelet glycoprotein 1b α and factor VII and risk of ischemic stroke: a meta-analysis. *Stroke* 2008;39:1710-1716.
28. McCarron MO, Delong D, Alberts MJ. APOE genotype as a risk factor for ischemic cerebrovascular disease: a meta-analysis. *Neurology* 1999;53:1308-1311.
29. Tsantes AE, Nikolopoulos GK, Bagos PG et al. Plasminogen activator inhibitor-1 4G/5G polymorphism and risk of ischemic stroke: a meta-analysis. *Blood Coagul Fibrinolysis* 2007;18:497-504.
30. Wu AH, Tsongalis GJ. Correlation of polymorphisms to coagulation and biochemical risk factors for cardiovascular diseases. *Am J Cardiol* 2001;87:1361-1366.
31. Staton JM, Sayer MS, Hankey GJ et al. Association between phosphodiesterase 4D gene and ischaemic stroke. *J Neurol Neurosurg Psychiatry* 2006;77:1067-1069.
32. Baum L, Ng HK, Wong KS et al. Associations of apolipoprotein E exon 4 and lipoprotein lipase S447X polymorphisms with acute ischemic stroke and myocardial infarction. *Clin Chem Lab Med* 2006;44:274-281.
33. Bevan S, Dichgans M, Gschwendtner A, Kuhlenbaumer G, Ringelstein EB, Markus HS. Variation in the PDE4D gene and ischemic stroke risk: a systematic review and meta-analysis on 5200 cases and 6600 controls. *Stroke* 2008;39:1966-1971.
34. Lovkvist H, Smith JG, Luthman H et al. Ischaemic stroke in hypertensive patients is associated with variations in the PDE4D genome region. *Eur J Hum Genet* 2008;16:1117-1125.
35. Abdollahi MR, Huang S, Rodriguez S et al. Homogeneous assay of rs4343, an ACE I/D proxy, and an analysis in the British Women's Heart and Health Study (BWHHS). *Dis Markers* 2008;24:11-17.
36. Gretarsdottir S, Thorleifsson G, Manolescu A et al. Risk variants for atrial fibrillation on chromosome 4q25 associate with ischemic stroke. *Ann Neurol* 2008;64:402-409.
37. Rockhill B, Newman B, Weinberg C. Use and misuse of population attributable fractions. *Am J Public Health* 1998;88:15-19.

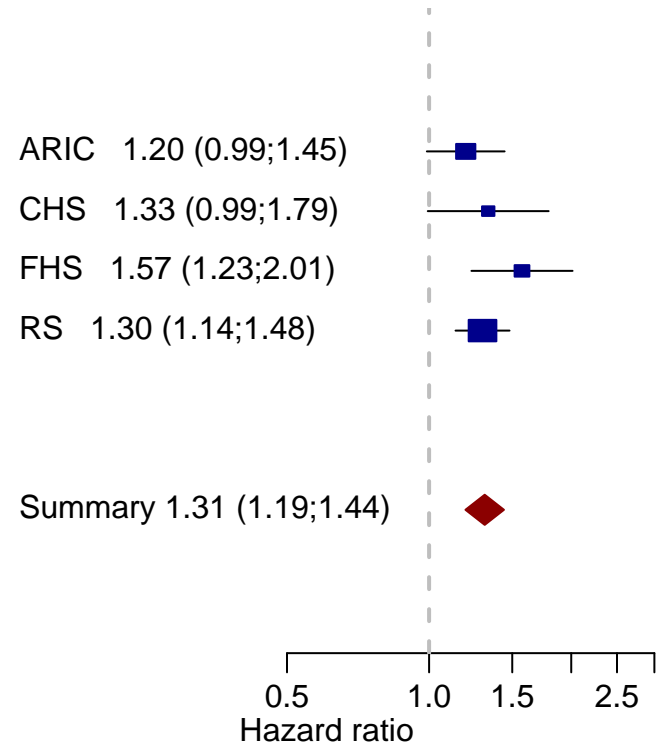
Figures 1A and 1B. Quantile-quantile (QQ)-plot showing the observed versus the expected p-values after meta-analysis for total stroke and ischemic stroke respectively. The red line shows the distribution under the null-hypothesis (also uploaded as separate figures).



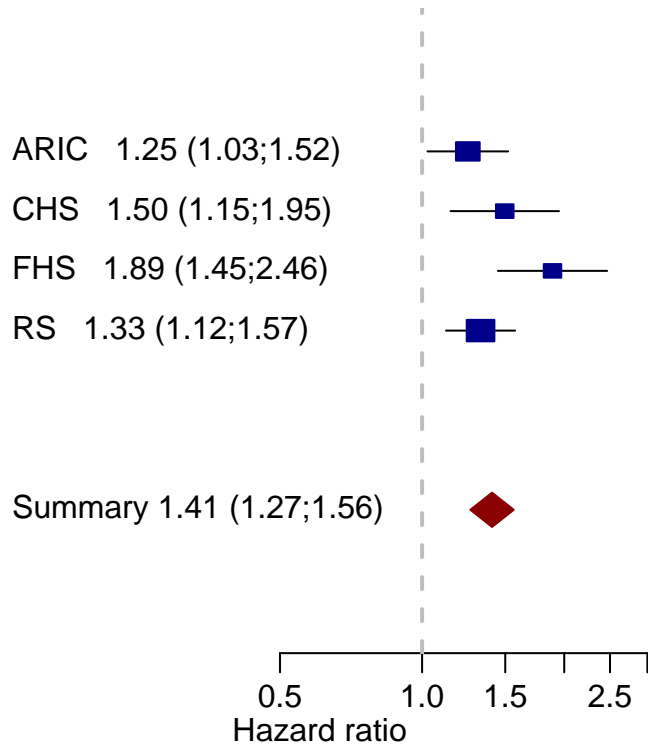
A. all stroke (rs11833579)



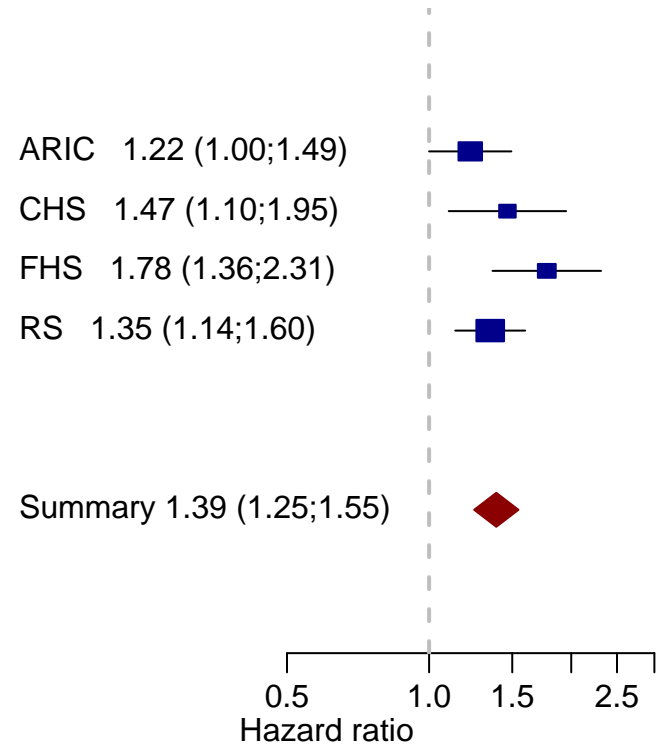
B. all stroke (rs12425791)



C. ischemic stroke (rs11833579)

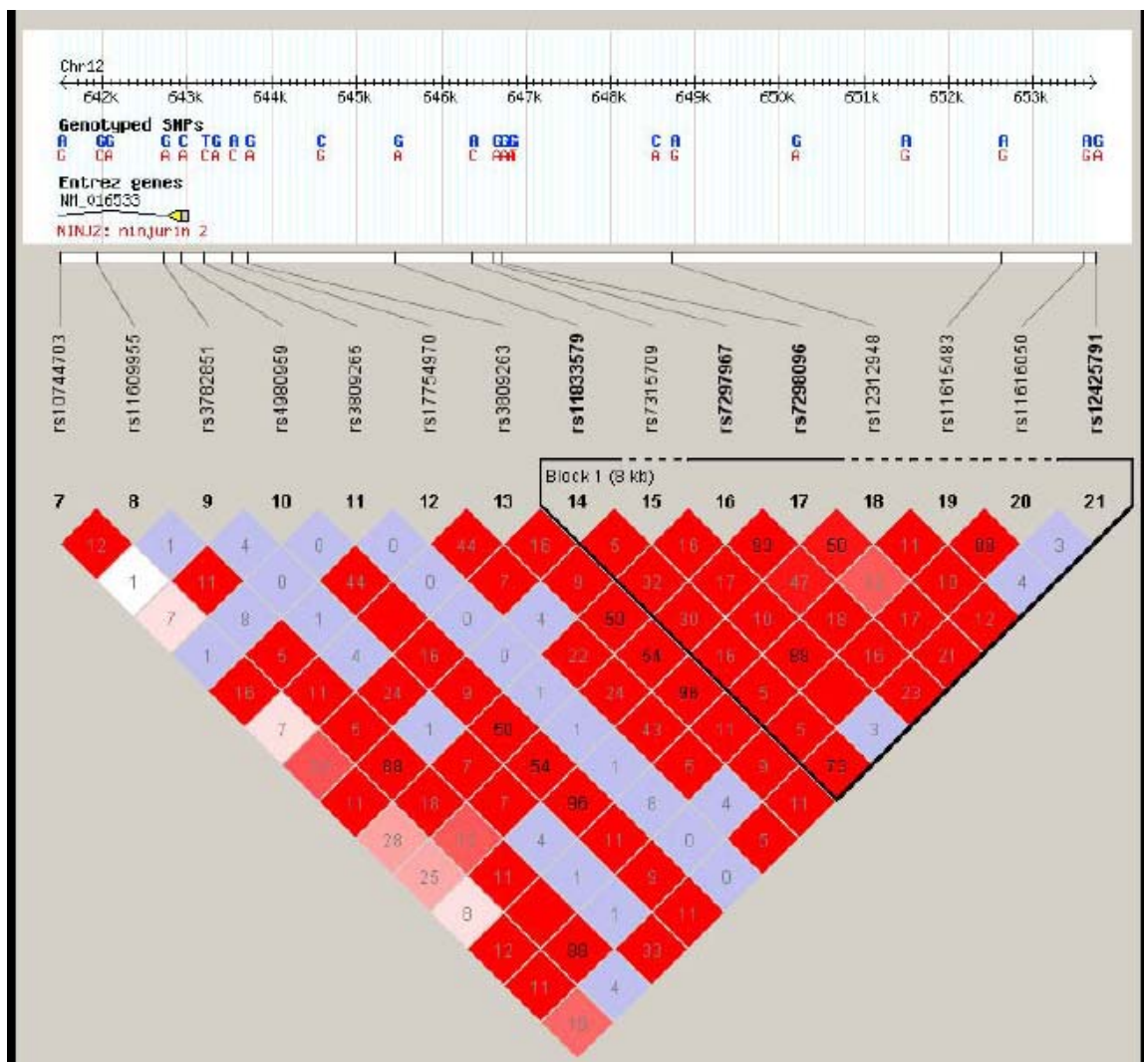


D. ischemic stroke (rs12425791)



Supplementary Appendix: Figure 2: Forest plots based on initial GWAS for total stroke: rs11833579 (A) and rs12425791 (B) and for ischemic stroke: rs11833579 (C) and rs12425791 (D).

Figure 3: Linkage-disequilibrium plot of SNPs adjacent to the *NINJ2* gene that were significantly associated with total stroke and ischemic stroke and SNPs at the 5' end of the *NINJ2* gene (also uploaded as separate figure). SNPs in bold are those that were significantly associated with total stroke and ischemic stroke



Tables 1 (total stroke) and 2 (ischemic stroke): results for all associations with a meta-analysis p-value lower than 10^{-4}

Results are ordered by chromosomal number and position. The table provides details specific to the individual participating cohorts including whether SNP was genotyped with sufficient reliability to be used in imputation [1] or was imputed [0], and the hazard ratios (with 95% confidence intervals) and p-values within each participating cohort. It also provides hazard ratios (95% confidence intervals) and p-values based on fixed-effects (inverse variance-weighted) meta-analysis and Z-test statistics and p-values based on inverse square root (N) weighting meta-analysis. The Human Gene Organization (HUGO) Gene Nomenclature System symbols for the two genes located closest to each SNP and the distance of the associated SNP from the 5' end (start) of the gene are also shown. Standardized gene annotations for all SNP results were derived programatically from the UCSC Genome Browser RefSeq gene track (hg18). Distances to genes are given in kilo-base (kb) pairs, based on NCBI build #36.

name	chromosome	position	strand	allele1	allele2	new allele	ref allele	r2	r3	r4	r5	r6	r7	r8	r9	r10	r11	r12	r13	r14	r15	r16	r17	r18	r19	r20	r21	r22	r23	r24	r25	r26	r27	r28	r29	r30	r31	r32	r33	r34	r35	r36	r37	r38	r39	r40	r41	r42	r43	r44	r45	r46	r47	r48	r49	r50	r51	r52	r53	r54	r55	r56	r57	r58	r59	r60	r61	r62	r63	r64	r65	r66	r67	r68	r69	r70	r71	r72	r73	r74	r75	r76	r77	r78	r79	r80	r81	r82	r83	r84	r85	r86	r87	r88	r89	r90	r91	r92	r93	r94	r95	r96	r97	r98	r99	r100	r101	r102	r103	r104	r105	r106	r107	r108	r109	r110	r111	r112	r113	r114	r115	r116	r117	r118	r119	r120	r121	r122	r123	r124	r125	r126	r127	r128	r129	r130	r131	r132	r133	r134	r135	r136	r137	r138	r139	r140	r141	r142	r143	r144	r145	r146	r147	r148	r149	r150	r151	r152	r153	r154	r155	r156	r157	r158	r159	r160	r161	r162	r163	r164	r165	r166	r167	r168	r169	r170	r171	r172	r173	r174	r175	r176	r177	r178	r179	r180	r181	r182	r183	r184	r185	r186	r187	r188	r189	r190	r191	r192	r193	r194	r195	r196	r197	r198	r199	r200	r201	r202	r203	r204	r205	r206	r207	r208	r209	r210	r211	r212	r213	r214	r215	r216	r217	r218	r219	r220	r221	r222	r223	r224	r225	r226	r227	r228	r229	r230	r231	r232	r233	r234	r235	r236	r237	r238	r239	r240	r241	r242	r243	r244	r245	r246	r247	r248	r249	r250	r251	r252	r253	r254	r255	r256	r257	r258	r259	r260	r261	r262	r263	r264	r265	r266	r267	r268	r269	r270	r271	r272	r273	r274	r275	r276	r277	r278	r279	r280	r281	r282	r283	r284	r285	r286	r287	r288	r289	r290	r291	r292	r293	r294	r295	r296	r297	r298	r299	r300	r301	r302	r303	r304	r305	r306	r307	r308	r309	r310	r311	r312	r313	r314	r315	r316	r317	r318	r319	r320	r321	r322	r323	r324	r325	r326	r327	r328	r329	r330	r331	r332	r333	r334	r335	r336	r337	r338	r339	r340	r341	r342	r343	r344	r345	r346	r347	r348	r349	r350	r351	r352	r353	r354	r355	r356	r357	r358	r359	r360	r361	r362	r363	r364	r365	r366	r367	r368	r369	r370	r371	r372	r373	r374	r375	r376	r377	r378	r379	r380	r381	r382	r383	r384	r385	r386	r387	r388	r389	r390	r391	r392	r393	r394	r395	r396	r397	r398	r399	r400	r401	r402	r403	r404	r405	r406	r407	r408	r409	r410	r411	r412	r413	r414	r415	r416	r417	r418	r419	r420	r421	r422	r423	r424	r425	r426	r427	r428	r429	r430	r431	r432	r433	r434	r435	r436	r437	r438	r439	r440	r441	r442	r443	r444	r445	r446	r447	r448	r449	r450	r451	r452	r453	r454	r455	r456	r457	r458	r459	r460	r461	r462	r463	r464	r465	r466	r467	r468	r469	r470	r471	r472	r473	r474	r475	r476	r477	r478	r479	r480	r481	r482	r483	r484	r485	r486	r487	r488	r489	r490	r491	r492	r493	r494	r495	r496	r497	r498	r499	r500	r501	r502	r503	r504	r505	r506	r507	r508	r509	r510	r511	r512	r513	r514	r515	r516	r517	r518	r519	r520	r521	r522	r523	r524	r525	r526	r527	r528	r529	r530	r531	r532	r533	r534	r535	r536	r537	r538	r539	r540	r541	r542	r543	r544	r545	r546	r547	r548	r549	r550	r551	r552	r553	r554	r555	r556	r557	r558	r559	r560	r561	r562	r563	r564	r565	r566	r567	r568	r569	r570	r571	r572	r573	r574	r575	r576	r577	r578	r579	r580	r581	r582	r583	r584	r585	r586	r587	r588	r589	r590	r591	r592	r593	r594	r595	r596	r597	r598	r599	r600	r601	r602	r603	r604	r605	r606	r607	r608	r609	r610	r611	r612	r613	r614	r615	r616	r617	r618	r619	r620	r621	r622	r623	r624	r625	r626	r627	r628	r629	r630	r631	r632	r633	r634	r635	r636	r637	r638	r639	r640	r641	r642	r643	r644	r645	r646	r647	r648	r649	r650	r651	r652	r653	r654	r655	r656	r657	r658	r659	r660	r661	r662	r663	r664	r665	r666	r667	r668	r669	r670	r671	r672	r673	r674	r675	r676	r677	r678	r679	r680	r681	r682	r683	r684	r685	r686	r687	r688	r689	r690	r691	r692	r693	r694	r695	r696	r697	r698	r699	r700	r701	r702	r703	r704	r705	r706	r707	r708	r709	r710	r711	r712	r713	r714	r715	r716	r717	r718	r719	r720	r721	r722	r723	r724	r725	r726	r727	r728	r729	r730	r731	r732	r733	r734	r735	r736	r737	r738	r739	r740	r741	r742	r743	r744	r745	r746	r747	r748	r749	r750	r751	r752	r753	r754	r755	r756	r757	r758	r759	r760	r761	r762	r763	r764	r765	r766	r767	r768	r769	r770	r771	r772	r773	r774	r775	r776	r777	r778	r779	r780	r781	r782	r783	r784	r785	r786	r787	r788	r789	r790	r791	r792	r793	r794	r795	r796	r797	r798	r799	r800	r801	r802	r803	r804	r805	r806	r807	r808	r809	r810	r811	r812	r813	r814	r815	r816	r817	r818	r819	r820	r821	r822	r823	r824	r825	r826	r827	r828	r829	r830	r831	r832	r833	r834	r835	r836	r837	r838	r839	r840	r841	r842	r843	r844	r845	r846	r847	r848	r849	r850	r851	r852	r853	r854	r855	r856	r857	r858	r859	r860	r861	r862	r863	r864	r865	r866	r867	r868	r869	r870	r871	r872	r873	r874	r875	r876	r877	r878	r879	r880	r881	r882	r883	r884	r885	r886	r887	r888	r889	r890	r891	r892	r893	r894	r895	r896	r897	r898	r899	r900	r901	r902	r903	r904	r905	r906	r907	r908	r909	r910	r911	r912	r913	r914	r915	r916	r917	r918	r919	r920	r921	r922	r923	r924	r925	r926	r927	r928	r929	r930	r931	r932	r933	r934	r935	r936	r937	r938	r939	r940	r941	r942	r943	r944	r945	r946	r947	r948	r949	r950	r951	r952	r953	r954	r955	r956	r957	r958	r959	r960	r961	r962	r963	r964	r965	r966	r967	r968	r969	r970	r971	r972	r973	r974	r975	r976	r977	r978	r979	r980	r981	r982	r983	r984	r985	r986	r987	r988	r989	r990	r991	r992	r993	r994	r995	r996	r997	r998	r999	r1000	r1001	r1002	r1003	r1004	r1005	r1006	r1007	r1008	r1009	r1010	r1011	r1012	r1013	r1014	r1015	r1016	r1017	r1018	r1019	r1020	r1021	r1022	r1023	r1024	r1025	r1026	r1027	r1028	r1029	r10
------	------------	----------	--------	---------	---------	---------------	---------------	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-----

Table 3: Results of NINJ2 associations with total and ischemic stroke adjusted for age-, sex- and various stroke risk factors

	Rs11833579	Rs12425791
Total stroke		
Adjusted for age-,sex	1.26 (1.16-1.37)	1.30 (1.19-1.42)
Adjusted for age-,sex- and systolic blood pressure	1.24 (1.14-1.34)	1.30 (1.20-1.41)
Adjusted for age-,sex- and presence of hypertension (JNC-7 criteria)	1.25 (1.15-1.36)	1.30 (1.19-1.42)
Adjusted for age-,sex- and presence of diabetes	1.25 (1.15-1.36)	1.28 (1.18-1.40)
Adjusted for age-,sex- and current smoking status	1.26 (1.16-1.37)	1.30 (1.19-1.42)
Adjusted for age-,sex- and presence of atrial fibrillation	1.28 (1.18-1.39)	1.33 (1.21-1.45)
Adjusted for age-,sex- and all of the above	1.27 (1.17-1.39)	1.31 (1.20-1.43)
Ischemic Stroke		
Adjusted for age-,sex	1.33 (1.21-1.46)	1.33 (1.21-1.47)
Adjusted for age-,sex- and systolic blood pressure	1.30 (1.19-1.43)	1.36 (1.24-1.50)
Adjusted for age-,sex- and presence of hypertension (JNC-7 criteria)	1.33 (1.21-1.46)	1.36 (1.24-1.51)
Adjusted for age-,sex- and presence of diabetes	1.32 (1.20-1.45)	1.34 (1.21-1.48)
Adjusted for age-,sex- and current smoking status	1.32 (1.20-1.45)	1.35 (1.22-1.49)
Adjusted for age-,sex- and presence of atrial fibrillation	1.32 (1.20-1.45)	1.38 (1.25-1.52)
Adjusted for age-,sex- and all of the above	1.34 (1.21-1.47)	1.35 (1.22-1.50)

*All results are presented as HR (hazard ratio) and 95% CI (confidence interval) per copy of minor allele, these results are based on directly genotyped data for rs11833579 and rs12425791.

Table 4. Characteristics of Participants in Replication Cohorts

STUDY	ARIC African-American	CHS African-American	The Netherlands (White)	
	cohort	cohort	Cases	Controls
Number in sample*	2430	574	652	3613
Women %	63%	65%	48%	62%
Mean follow-up (yrs)	14.7 (3.6)	9.9 (4.03)	-	-
<i>at DNA draw</i>	53.6 (5.7)	72.5 (5.5)	76.5 (7.0)	74.6 (7.0)
<i>at stroke</i>	63.8 (6.9)	79.97 (5.8)	76.5 (7.0)	74.6 (7.0)
Number of total strokes	214	85	652	-
Number of ischemic strokes	191	68	501	-
Number of atherothrombotic strokes	153	57	400	-

Table 5: Association of rs12425791 and rs11833579 with various stroke related phenotypes in discovery and replication cohorts

	Discovery						Replication								
	Using genotyped and imputed GWAS data			Using only directly genotyped data			ARIC African - American cohort			CHS African-American cohort			Netherlands, white case-control sample cohort		
	n	HR	p	n	HR	p	n	RR	p	n	HR	p	n	HR	p
Rs12425791															
Total stroke	1544	1.31	1.5x10 ⁻⁸	1544	1.30	5.0x10 ⁻⁹	215	1.35	3.6x10 ⁻²	85	0.58	ns	652	1.17	3.5x10 ⁻²
Ischemic stroke	1164	1.39	2.6x10 ⁻⁹	1164	1.33	1.2x10 ⁻⁸	191	1.42	1.9x10 ⁻²	68	0.49	ns	501	1.19	3.7x10 ⁻²
Atherothrombotic	-	-	-	876	1.37	3.3x10 ⁻⁸	153	1.41	4.1x10 ⁻²	57	0.51	ns	400	1.29	5.2x10 ⁻³
Rs11833579															
Total stroke	1544	1.32	4.8x10 ⁻⁹	1544	1.26	6.1x10 ⁻⁸	215	1.15	ns	85	0.72	ns	652	1.10	ns
Ischemic stroke	1164	1.41	2.3x10 ⁻¹⁰	1164	1.33	2.3x10 ⁻⁹	191	1.20	ns	68	0.65	ns	501	1.11	ns
Atherothrombotic	-	-	-	876	1.35	4.0x10 ⁻⁸	153	1.28	ns [†]	57	0.65	ns	400	1.19	4.8x10 ⁻²
Meta-meta analysis															
	All data in whites			All data											
	n	HR	p	n	HR	p									
Rs12425791															
Total stroke	2196	1.27	1.0x10 ⁻⁹	2496	1.26	7.9x10 ⁻¹⁰									
Ischemic stroke	1665	1.29	2.4x10 ⁻⁹	1924	1.29	1.1x10 ⁻⁹									
Atherothrombotic	1029	1.35	7.2x10 ⁻¹⁰	1486	1.33	5.2x10 ⁻¹⁰									
Rs11833579															
Total stroke	2196	1.21	1.2x10 ⁻⁷	2496	1.19	4.3x10 ⁻⁷									
Ischemic stroke	1665	1.27	6.4x10 ⁻⁹	1924	1.24	2.3x10 ⁻⁸									
Atherothrombotic	1029	1.30	1.3x10 ⁻⁸	1486	1.28	2.2x10 ⁻⁸									

n= number of strokes in sample

[†] Two-sided p-values are shown, one-sided p-value for this replication analysis is 0.04 falling below a threshold of p<0.05.

Table 6: Associations with Putative Candidate Stroke SNPs in the CHARGE meta-analysis

Chr	Gene Name	Gene Symbol	Candidate SNP	proxy SNP	coded allele	Total stroke		Ischemic stroke		
						Hazard ratio	pvalue	Hazard ratio	pvalue	
1	5,10-methylenetetrahydrofolate reductase	MTHFR	C677T = rs1801133		C	0.94 (0.86-1.02)	0.12	0.93 (0.84-1.02)	0.11	
1	Factor V leiden	F5	1691G>A = rs6025	NA						
4	intergenic		Rs2200733		C	0.97 (0.86-1.08)	0.55	1.00 (0.88-1.15)	0.95	
5	Phosphodiesterase 4D	PDE4D	SNP 39 = rs3887175		T	1.01 (0.91-1.13)	0.81	1.00 (0.88-1.13)	0.96	
			SNP 41 = rs12188950		C	0.97 (0.87-1.08)	0.57	0.97 (0.86-1.10)	0.66	
			SNP 45 = rs12153798		C	1.02 (0.91-1.13)	0.78	1.03 (0.91-1.16)	0.67	
			SNP 56 = rs702553	NA						
			SNP 83 = rs1396476		G	1.03 (0.93-1.14)	0.56	0.99 (0.88-1.12)	0.87	
			SNP 87 = rs966221		C	0.98 (0.91-1.06)	0.68	0.96 (0.88-1.06)	0.43	
			SNP 89 = rs2910829		G	1.01 (0.94-1.09)	0.80	1.01 (0.93-1.11)	0.76	
7	Serpin peptidase inhibitor	SERPINE1	11053G/T = rs7242		G	0.98 (0.91-1.06)	0.63	0.99 (0.91-1.08)	0.86	
			-675 G4/G5 = rs1799768	NA						
9	intergenic		Rs1537378		C	1.05 (0.97-1.14)	0.21	1.06 (0.97-1.16)	0.22	
11	Prothrombin	F2	20210G>A = rs1799963	NA	See Supplementary Table A rows 206-211 for other SNPs (not in LD with candidate SNP) in this gene at $p < 10^{-4}$					
11	Glycoprotein 1 b alpha	GP1BA	HPA-2 (Thr/Met) = rs6065		C	0.92 (0.80-1.05)	0.219	0.93 (0.79-1.09)	0.35	
			Kozak sequence -5 C/T = rs2243093	rs16954257	T	0.86 (0.76-0.98)	0.020	0.83 (0.72-0.97)	0.02	
13	Arachidonate 5-lipoxygenase-activating	ALOX5AP	SG13S25 = rs17222814	NA						

	protein									
17	Angiotensin converting enzyme	ACE	I/D = rs1799752	rs4343	G	1.04 (0.96-1.12)	0.32	1.03 (0.94-1.12)	0.54	
19	Apolipoprotein E	APOE	rs7412, rs429358	NA						

If the reported SNP was not present in our dataset, a proxy SNP was selected using the SNAP program

<http://www.broad.mit.edu/mpg/snap/>

A proxy SNP was required to have an $r^2 > 0.8$ with the reported SNP. If no such proxy SNP was present in our dataset this is indicated as NA (not available).