GENOMEWIDE ASSOCIATION STUDY OF INTRACRANIAL ANEURYSMS CONFIRMS ROLE OF ANRIL AND SOX17 IN DISEASE RISK

SUPPLEMENTAL MATERIAL

SUPPLEMENTAL METHODS

Discovery Sample 1

The initial discovery sample consisted of individuals recruited through the Familial Intracranial Aneurysm (FIA) I study,[1] in which familial cases that met one of several criteria involving multiple family members with IA, which would make the family appropriate for linkage analysis,[2] were recruited through 26 clinical centers (41 sites) in North America, New Zealand, and Australia. Exclusion criteria included: (i) a fusiform-shaped unruptured IA of a major intracranial trunk artery; (ii) an IA which is part of an arteriovenous malformation; (iii) a family or personal history of polycystic kidney disease, Ehlers Danlos syndrome, Marfan's syndrome, fibromuscular dysplasia, or Moya-Moya disease; or (iv) failure to obtain informed consent from the patient or family members. All medical records and relevant accompanying data were reviewed by a Verification Committee. For the present analysis, only individuals having an IA based on an intra-arterial angiogram, operative report, autopsy, or size ≥7 mm on non-invasive imaging (MRA, CTA) were considered "definite" cases. A set of independent unrelated cases was obtained by selecting one individual with definite IA from each FIA I family self-reported as Caucasian (n=389). The FIA study was approved by the Institutional Review Boards/Ethics Committees at all clinical and analytical centers and recruitment sites.

Controls for the Discovery Sample 1 were obtained from two population-based studies. The first was the NINDS-funded case-control Genetic and Environmental Risk Factors for Hemorrhage Stroke (GERFHS) study, which was designed to identify the important environmental and genetic risk factors for IA-related SAH as well as for spontaneous intracerebral hemorrhage. Controls identified by random-digit telephone dialing from the Greater Cincinnati/Northern Kentucky community and matched to enrolled cases by age (±5 years), gender, and race, had the same interview questions regarding environmental risk factors as FIA study participants. A set of 113 unrelated, Caucasian controls were selected for genotyping. In addition, 290 Caucasian controls free of stroke and IA were selected from the Cincinnati Control Cohort. The subjects in this cohort were identified by random-digit dialing from the Greater Cincinnati region during 2006. These subjects had blood drawn for DNA extraction as well as extensive interviews including detailed environmental exposures as well as detailed medical history of every major disease. Both studies were approved by the Institutional Review Boards of the University of Cincinnati and all participating hospitals.

Results from a subset of the cases (n=343) and controls (n=374) from Discovery Sample 1 were previously reported in Deka et al, 2010.[3]

Discovery Sample 2

During FIA II study recruitment, the requirement for family history of IA was removed and both familial and sporadic IA cases were enrolled. The same exclusion

criteria were in place and all cases underwent the same rigorous review from the Verification Committee. A set of 829 Caucasian IA cases was selected for genotyping from this sample, and an additional 61 Caucasian sporadic aneurysmal SAH cases from the Greater Cincinnati/Northern Kentucky region were obtained from GERFHS.

This sample was augmented by Caucasian cases and controls identified from other studies, including those from the Australasian Cooperative Research on Subarachnoid hemorrhage Study (ACROSS), which was a prospective, population-based, case-control study of SAH undertaken in three cities in Australia and one city in New Zealand during the mid-1990s.[4] ACROSS included incidence cases of SAH secondary to documented or presumed ruptured IA who were frequency-matched (by sex, 10-year age strata, and city of residence) to controls selected from electoral rolls in each city. Detailed information about key exposures, such as smoking, hypertension, family history of stroke/IA, was obtained by standardized interviews with subjects (or proxies) and where possible, blood samples were obtained for storage and future DNA extraction. Samples from a total of 160 cases and 168 controls were available for genotyping. This study was approved by the institutional review committees at 10 sites.

In addition, IA cases were recruited from a prospective cohort study of adult patients with spontaneous SAH due to IA confirmed by non-contrast CT and cerebral angiogram who were admitted to a tertiary-care referral center in San Francisco during 2003 to 2008. Additional FIA exclusion criteria were also applied to yield 184 samples from Caucasian subjects with detailed medical histories and blood banked for DNA. This study was approved by the institutional review committee at University of California, San Francisco.

Genotypic data from a further set of 1148 white controls was obtained through a collaborative agreement with the Atherosclerosis Risk in Communities (ARIC) study. In the ARIC sample, a subset of subjects who never had a stroke or TIA was matched to the Discovery Sample 2 cases by sex and, where possible, by age (±5 years). However, because the age of the ARIC controls was limited to 44–66, cases younger than 39 or older than 71 at onset were matched to controls outside of the 5-year criterion. Genotyping had been previously performed using the Affymetrix 6.0 array.[5]

Family history

Positive family history for cases in Discovery Sample 1 was validated (supplemental text). For most Discovery Sample 2 cases and for controls in Discovery Sample 1, positive family history was based only on self-report by the subject. Family history was considered positive if any relative (not necessarily first-degree relative) was reported to have had a ruptured or unruptured IA. Family history of IA was not collected in controls for the ARIC or ACROSS studies.

Genotyping and Quality Review

Genotyping was performed using the Axiom array at the Affymetrix core labs for all samples except for the ARIC controls. Twenty-five internal samples were genotyped twice for quality control. This yielded a total of 2,219 samples sent for genotyping. However, only 2,140 samples with a QC (dQC) value ≥0.82 and an initial call rate of 97% were released. All released genotypes underwent a common quality review pipeline which included identification of sample duplicates, related individuals, and

gender discrepancies, which resulted in the removal of 59 samples. Prior to performing imputation, SNPs were excluded if there was: (i) improper mapping to Genome Reference Consortium GRCh37; (ii) a minor allele frequency (MAF) <0.03; (ii) a SNP call <95%; (iv) a Hardy Weinberg Equilibrium (HWE) p-value in controls of $p<10^{-2}$ and $p<10^{-4}$ in cases. From the 567,096 SNPs on the Axiom array, 473,238 were retained following this quality review.

A principal component analysis (PCA) was performed using Eigenstrat[6] and data from 11 HapMap phase III populations to identify clusters using the first two eigenvectors computed using the SNPs typed on both platforms. Samples clustering with the European American (CEU) reference set (PC1: 0.0024–0.0078; PC2: 0.0007–0.0049) were retained, and those outside this cluster which were likely to contain African, Asian, or Hispanic admixture were removed from further analysis (n=47 of the Axiom-genotyped samples); 16 non-European American samples from the ARIC set were also removed. Coordinates of study subjects for PCs 1 and 2 are illustrated in Supplemental Figure 2.

Genotypic data for the ARIC samples was obtained from the Affy 6.0 array.[5] These data also underwent quality review and SNPs were removed based on the same criteria listed above. From the 793,799 SNPs on the Affy 6.0 array that were provided by ARIC following their initial data review, a total of 619,514 were retained for analysis in this study.

Power for the Samples

Discovery Sample 1 had low power to detect an allelic association at the genome-wide significance threshold (5×10^{-8}), but had 50% power to detect an association with an odds ratio of 1.5 at the 10^{-4} screening threshold for common SNPs (MAF≥0.4). At the threshold for genome-wide significant evidence of association, Discovery Sample 2 had 80% or greater power to detect an odds ratios of 1.4 or greater for common SNPs (MAF≥0.4). Similarly, the meta-analysis sample had 90% or greater power to detect odds ratios of 1.4 or greater with common SNPs (MAF≥0.4) at the genome-wide significance level.

Imputation

Imputation was performed for all autosomes using IMPUTE2 (https://mathgen.stats.ox.ac.uk/impute/impute_v2.html). All distinct samples genotyped on the Axiom array (n=2,115) were imputed together using the 1000Genomes haplotypes (n=1094; data freeze from Nov. 2010, Jun. 2011 phased haplotype release, mapped to GRCh37) as the phased reference panel. IMPUTE2 implements a method that can utilize an additional reference panel data as well (https://mathgen.stats.ox.ac.uk/impute/using_multi_population_reference_panels.html); we included the 1,148 ARIC samples as another unphased reference panel to maximize available information at each imputed SNP.

Because Discovery Sample 2 was genotyped on two platforms, with nearly complete confounding by type of sample (i.e., all cases on Axiom; the majority of controls on Affy 6.0), extensive and detailed quality review was performed to ensure that spurious association was not detected based on platform effects. As suggested by

Sinnott and Kraft,[7] we reviewed several SNP metrics, including imputation quality (information) and differences in SNP minor allele frequency in controls genotyped on the Axiom platform, and the ARIC controls genotyped on the Affy 6.0. We removed all SNPs with low to moderate imputation quality (information score <0.50) as well as those SNPs with a significant difference in minor allele frequency between the two sources of control samples ($p<0.1$). To further reduce the influence of rare SNPs, which would typically have less accurate imputation, we removed all SNPs with a minor allele frequency less than 5%. Using this aggressive filtering approach, we retained 453,699 SNPs for analysis of Discovery Sample 2. Remaining uncertainty in the imputed genotypes after application of the aggressive information score and minor allele frequency filters was modeled using the "-method score" option in IMPUTE2. We would expect a slight loss of power in the association tests due to the uncertainty in genotypes; however, previous studies indicate this power loss is minimal, on the order of 7% of the effective sample size on average[8]

A few critical SNPs were not genotyped in the ARIC sample. Therefore, we performed imputation of specific SNPs using IMPUTE2. The same 1000G haplotypes are used as the phased reference samples. The 2115 samples genotyped on the Axiom array were used as the unphased reference panel.

Gene x Smoking Relationship

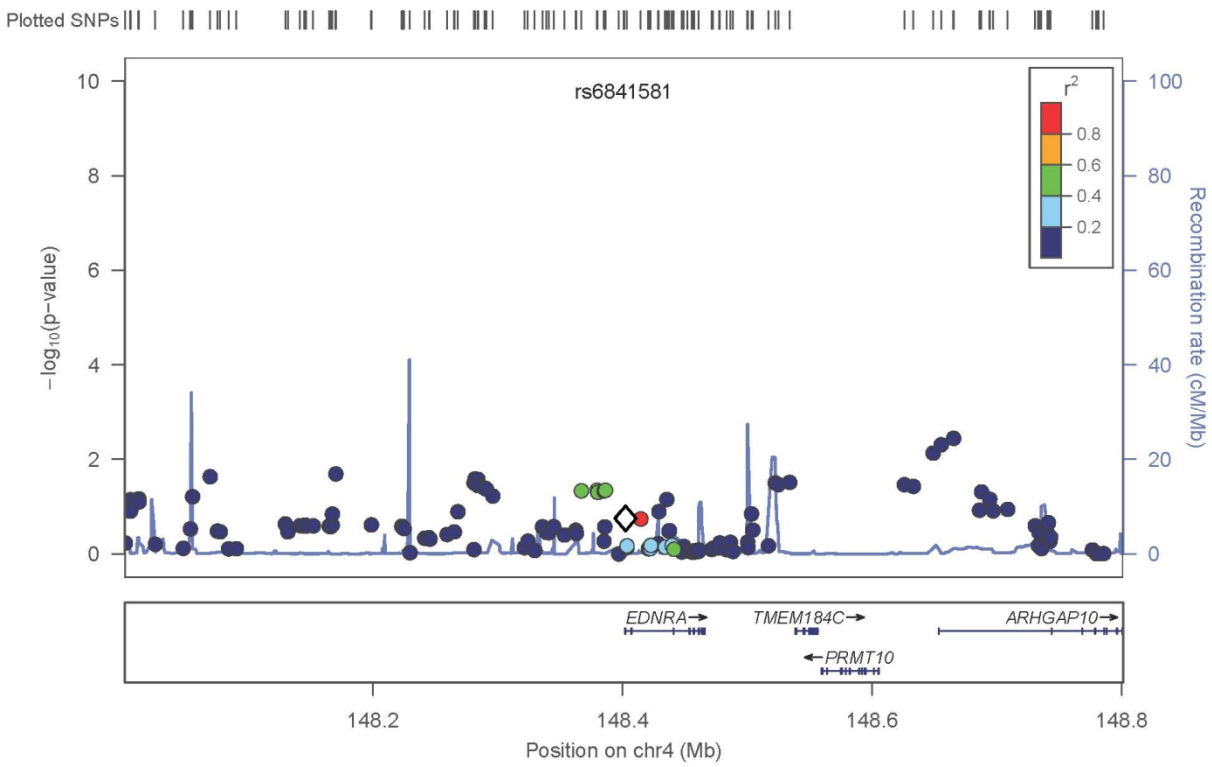
Since cigarette smoking is a very strong risk factor in IA, we examined the possible interaction of the most highly associated SNPs on chromosomes 8 and 9 with cigarette smoking. On chromosome 8, we used a SNP (rs1072737) that was imputed in the samples genotyped on the Axiom array. The imputation procedure generates each individual's probability of each genotype for this SNP. To avoid ambiguity, in cases where the probability of one SNP genotype was greater than 80%, we assigned the individual that genotype. However, if all genotypic probabilities were less than 80%, then that individual was omitted from the analysis (9% of samples removed; n=285). On chromosome 9 we used SNP (rs6475606), which was genotyped on both the Axiom and Affy 6.0 arrays. Using the same cases and controls from discovery samples 1 and 2, a logistic regression model was employed to test for departures from a multiplicative relationship between the risk allele scores (no risk allele=0, 1 risk allele=1, and 2 risk alleles=2) and cumulative exposure to smoking as measured by pack-years. Since the distribution of pack-years is highly skewed, we used the log of pack-years in the logistic model, with 0.05 pack-years assigned to the never-smokers. The logistic models were fitted for discovery samples 1 and 2, and each model was adjusted for age and sex. We combined estimates using meta-analysis with individual sample results weighted by the inverse variance of the sample estimates.

ADDITIONAL INTERPRETATION FOR TABLE 3

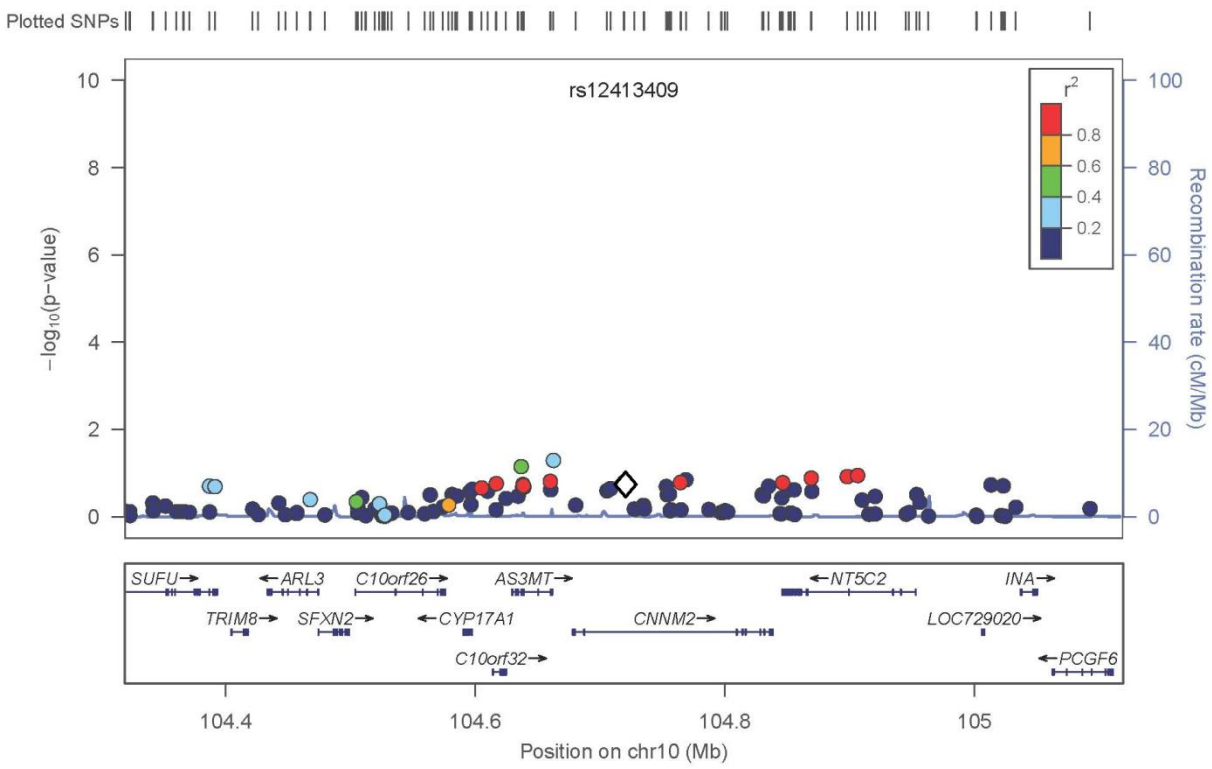
The OR for any given number of K pack-years can be calculated using the following equation: $OR = \exp(\beta(\ln(K) + 2.99))$ where β = the regression coefficient for log(pkyrs). $\beta=0.155$ for rs6475606 and $\beta=0.154$ for rs1072737. For example, the odds ratio for 40 pack years of smoking for subjects in the model of the rs6475606 risk allele = $\exp(0.155(\ln(40) + 2.99)) = 2.82$. To determine the odds ratio for presence of two risk

alleles of rs6475606 (homozygous state) and 40 pack years of smoking, one would multiply $(1.36)^2$ =1.85 (two risk alleles) x 2.82 which equals an odds ratio of 5.22.

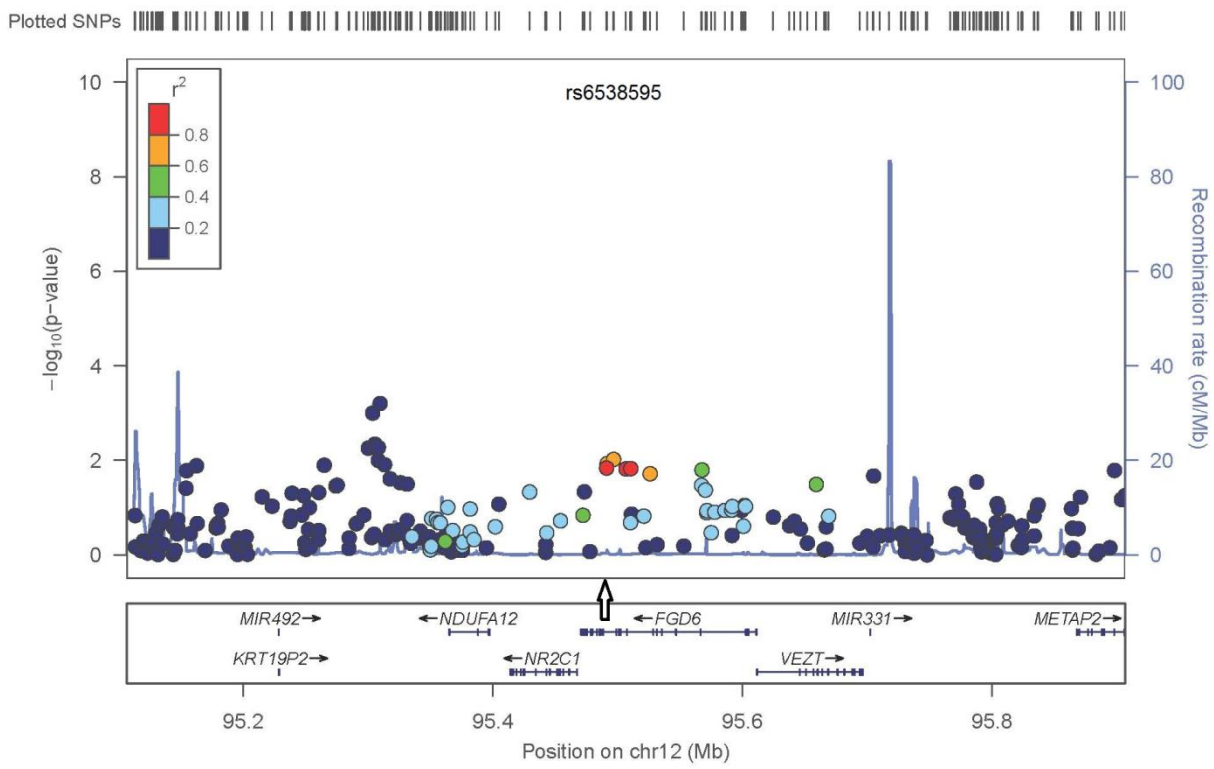
SUPPLEMENTAL FIGURES AND FIGURE LEGENDS



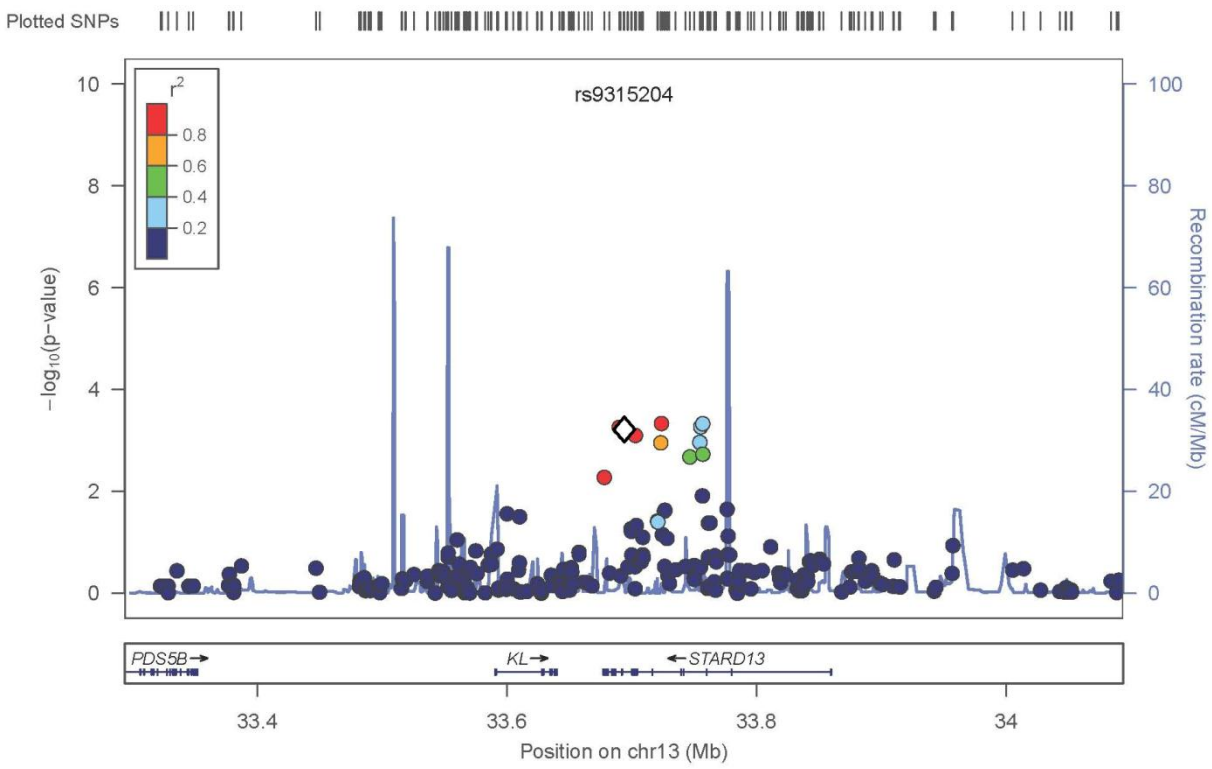
Supplemental Figure 1A



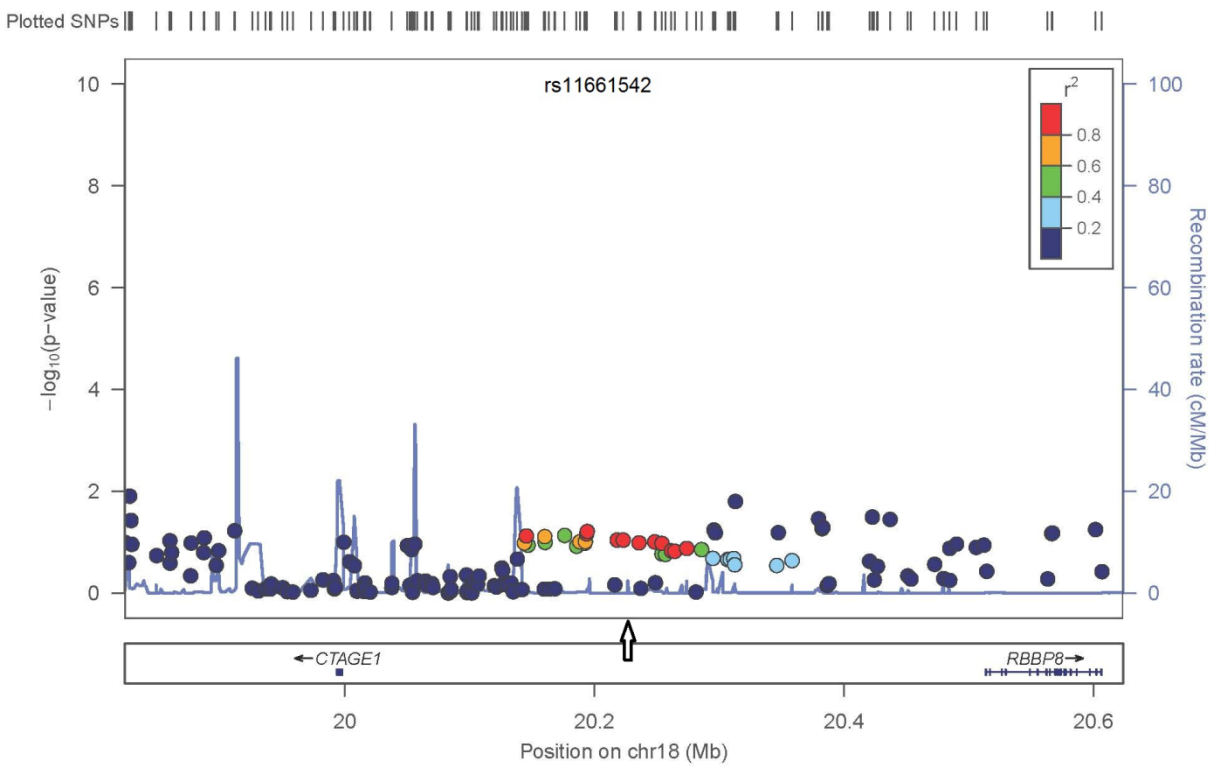
Supplemental Figure 1B



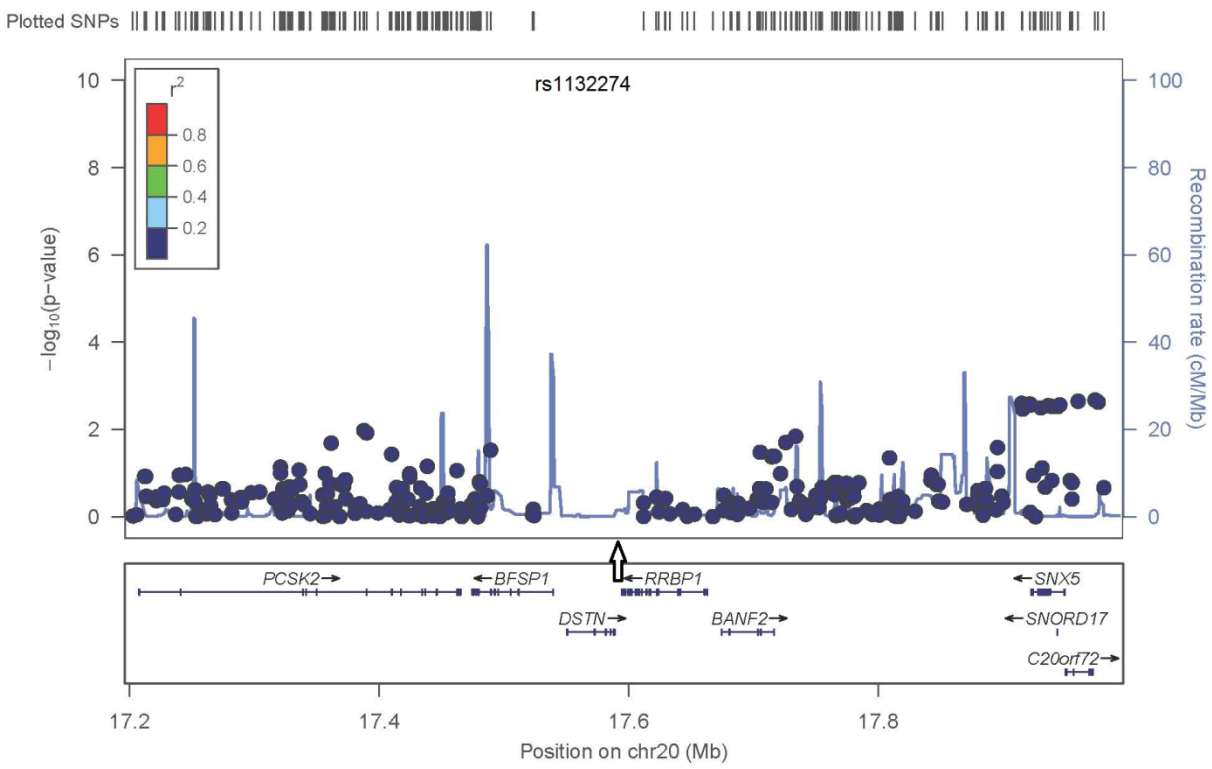
Supplemental Figure 1C



Supplemental Figure 1D



Supplemental Figure 1E



Supplemental Figure 1F

Figure 1: Comparison with previously reported results from genomewide association studies.[9, 10] (A) Chromosome 4q31.23 (*EDNRA*); (B) Chromosome 10q24.32 (*CNNM2*); (C) Chromosome 12q22; (D) Chromosome 13q13.1 (*KL/STARD13*); (E) Chromosome 18q11.2 (*RBBP8*); (F) Chromosome 20p12.1. When the most significant SNP from the initial report was available in the meta analysis, a white diamond indicates the SNP, and the SNP is listed at the top of the figure. If the SNP was not available in the meta analysis sample, an arrow is used to denote the position of that SNP within this map of markers. Each circle symbol within the graph indicates the p-value for a SNP at that position in the meta analysis. The color of the square symbol denotes the extent of linkage disequilibrium (as computed by r^2) with the SNP reported in the initial report.

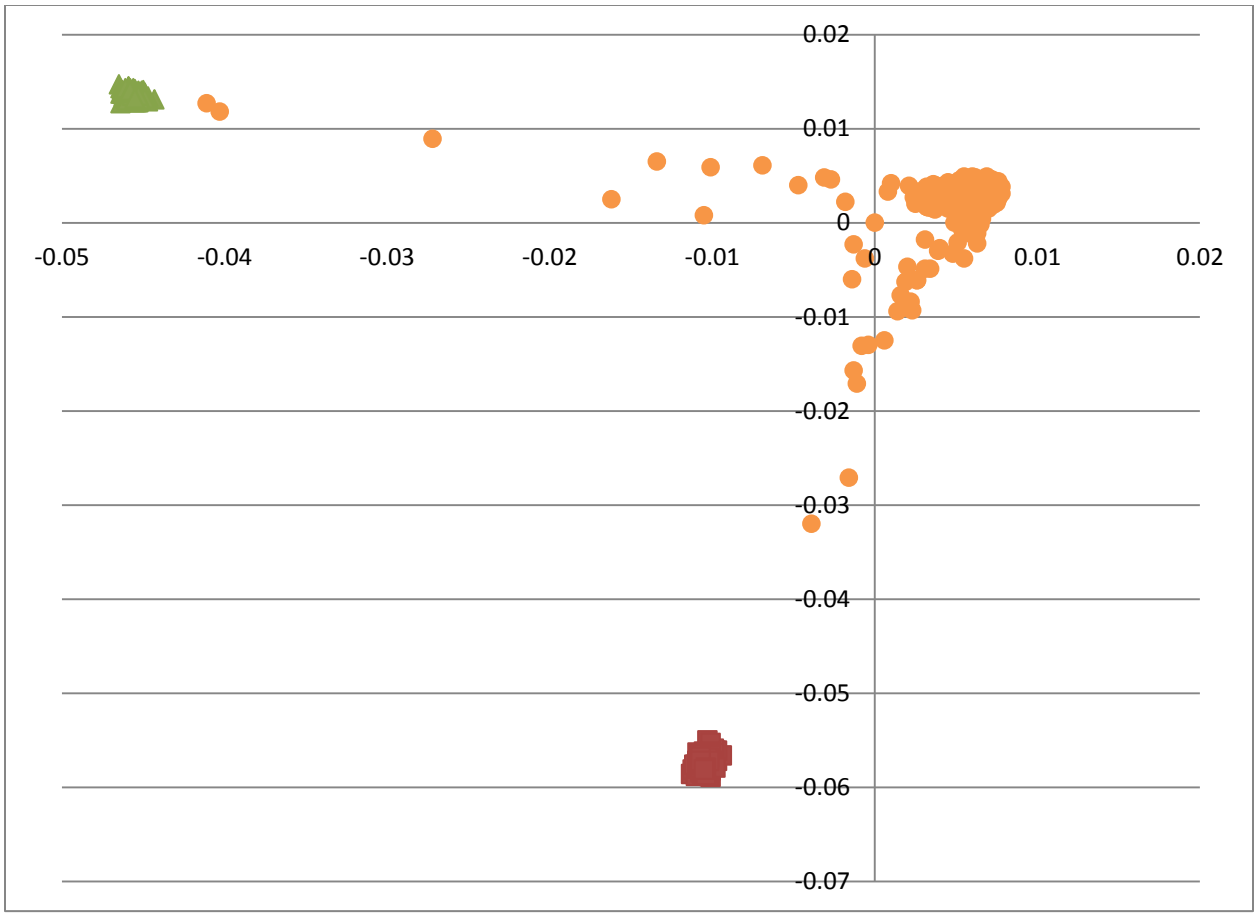


Figure 2: Principal component clustering plot for genotyped study subjects. Genotyped individuals are shown for PC1 (x-axis) and PC2 (y-axis). Reference populations YRI (African) and CHB/JPT (Asian) are shown as green and red symbols, respectively, with study subjects as orange circles. As described in the text, genotyped subjects clustering outside the area defined by the CEU (European-American) reference sample (PC1: 0.0024–0.0078; PC2: 0.0007–0.0049) were identified and excluded from association analyses.

SUPPLEMENTAL REFERENCES

1. Broderick JP, Sauerbeck LR, Foroud T, Huston J, 3rd, Pankratz N, Meissner I, et al. The familial intracranial aneurysm (fia) study protocol. *BMC Med Genet*. 2005;6:17
2. Foroud T, Sauerbeck L, Brown R, Anderson C, Woo D, Kleindorfer D, F,et al. Genome screen to detect linkage to intracranial aneurysm susceptibility genes: The familial intracranial aneurysm (fia) study. *Stroke*. 2008;39:1434-1440
3. Deka R, Koller DL, Lai D, Indugula SR, Sun G, Woo D, et al. The relationship between smoking and replicated sequence variants on chromosomes 8 and 9 with familial intracranial aneurysm. *Stroke*. 2010;41:1132-1137
4. Anderson C, Ni Mhurchu C, Scott D, Bennett D, Jamrozik K, Hankey G. Triggers of subarachnoid hemorrhage: Role of physical exertion, smoking, and alcohol in the australasian cooperative research on subarachnoid hemorrhage study (across). *Stroke*. 2003;34:1771-1776
5. Ikram MA, Seshadri S, Bis JC, Fornage M, DeStefano AL, Aulchenko YS, et al. Genomewide association studies of stroke. *N Engl J Med*. 2009;360:1718-1728
6. 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
7. Sinnott JA, Kraft P. Artifact due to differential error when cases and controls are imputed from different platforms. *Hum Genet*. 2012;131:111-119
8. Nair RP, Duffin KC, Helms C, Ding J, Stuart PE, Goldgar D, et al. Genome-wide scan reveals association of psoriasis with il-23 and nf-kappab pathways. *Nat Genet*. 2009;41:199-204
9. Yasuno K, Bakircioglu M, Low SK, Bilguvar K, Gaal E, Ruigrok YM, et al. Common variant near the endothelin receptor type a (ednra) gene is associated with intracranial aneurysm risk. *Proc Natl Acad Sci U S A*. 2011;108:19707-19712
10. Yasuno K, Bilguvar K, Bijlenga P, Low SK, Krischek B, Auburger G, et al. Genome-wide association study of intracranial aneurysm identifies three new risk loci. *Nat Genet*. 2010;42:420-425

Clinical Centers – University of Alabama at Birmingham: W. Fisher (PI), H. Forson (coordinator); Clinical Trials Research Unit, University of Auckland and Auckland City Hospital, New Zealand: C. Anderson (PI), E. Mee (PI), C. Howe (coordinator), S. Vos (coordinator); Royal Perth Hospital, Sir Charles Gairdner Hospital, Royal Adelaide Hospital, Royal Melbourne Hospital, Alfred Hospital, Westmead Hospital, Royal North Shore Hospital, Royal Prince Alfred Hospital, Australia: C. Anderson (PI), G. Hankey (PI), N. Knuckey (PI), J. Laidlaw (PI), P. Reilly (PI), N. Dorsch (PI), M. Morgan (PI), M. Besser (PI), J. Rosenfeld (PI), K. Athanasiadis (coordinator), A Claxton (coordinator), V. Dunne (coordinator), J. Griffith (coordinator), J Davidson (coordinator), S. Pope (coordinator), Amanda Froelich (coordinator); Brigham & Women's Hospital: A. Day (PI), R. Brach (coordinator); University of Cincinnati: D. Woo (co-PI), M. Zuccarello (co-PI), A. Ringer (co-PI), H. Yeh (co-PI), K. Franklin (coordinator); Cleveland Clinic Foundation: P. Ramussen (PI), D. Andrews-Hinders (coordinator), T. Wheeler (coordinator); Columbia University: E. S. Connolly (PI), R. Sacco (co-PI), D. LaMonica (coordinator); University of Florida: S. B. Lewis (PI), A. Royster (coordinator); Indianapolis Neurosurgical Group: T. Payner (PI), N. Miracle (coordinator); Johns Hopkins: K. Murphy (PI), B. Kohler (coordinator); Massachusetts General Hospital: C. Ogilvy (PI), D. Buckley (coordinator), J. Manansala (coordinator); London Health Science Center Research Inc.: G. Ferguson (PI), C. Mayer (coordinator), J. Peacock (coordinator); Notre Dame Hospital: G. Rouleau (PI), A. Desjarlais (coordinator); University of Maryland: E. F. Aldrich (PI), C. Aldrich (coordinator), C. Byard (coordinator); Mayo Clinic: R. D. Brown (PI), L. Jaeger (coordinator); University of Michigan: L. Morgenstern (PI), M. Concannon (coordinator); New Jersey Medical School: A. I. Qureshi (PI), P. Harris-Lane (coordinator); Northwestern University: H. Batjer (PI), G. Joven (coordinator), S. Thompson (coordinator); University of Ottawa: M. T. Richard (PI), A. Hopper (PI); University of Pittsburgh: A. B. Kassam (PI), K. Lee (coordinator); University of California, San Francisco: C. Johnston (PI), K. Katsura (coordinator); University of Southern California: S. Giannotta (PI), D. Fishback (coordinator); Stanford University Medical Center: G. Steinberg (PI), D. Luu (coordinator), M. Coburn (coordinator); University of Texas at Houston: M. Malkoff (PI), A. Wojner (coordinator); University of Virginia: N. Kassel (PI), B. Worrall (co-PI), G.

Radakovic (coordinator); University of Washington: D. Tirschwell (PI), P. Tanzi (coordinator); Washington University: C. Derdeyn (PI), M. Catanzaro (coordinator); University of Manitoba: A. Kaufmann (PI), D. Gladish (coordinator).

Executive Committee: Craig Anderson, MD; Joan E. Bailey-Wilson, PhD; Joseph P. Broderick, MD; Robert Brown, Jr. MD; E. Sander Connelly, MD; Ranjan Deka, PhD; Tatiana Foroud, PhD; John Huston, MD; Nerissa Ko, MD; Carl Langefeld, PhD; Irene Meissner, MD; Guy Rouleau MD, PhD; Laura Sauerbeck, RN, MS; Bradford Worrall, MD; Daniel Woo, MD, MS.

Imaging Phenotyping Committee: John Huston, MD; David Kallmes, MD; Harry Cloft, MD.

Medical Record Phenotyping: Joseph P. Broderick, MD; Robert Brown, Jr. MD; Daniel Woo, MD, MS; Irene Meissner, MD; David Wiebers, MD; Dawn Kleindorfer, MD; Matthew L. Flaherty, MD; Jason Mackey, MD, MS; Shannon Kohake, MD; Sharyl Martini, MD, PhD; Felipe De Los Rios La Rosa, MD.