

UCSF

UC San Francisco Previously Published Works

Title

Genome-wide association study of sporadic brain arteriovenous malformations

Permalink

<https://escholarship.org/uc/item/24d393tb>

Journal

Journal of Neurology Neurosurgery & Psychiatry, 87(9)

ISSN

0022-3050

Authors

Weinsheimer, Shantel

Bendjilali, Nasrine

Nelson, Jeffrey

et al.

Publication Date

2016-09-01

DOI

10.1136/jnnp-2015-312272

Peer reviewed



Published in final edited form as:

J Neurol Neurosurg Psychiatry. 2016 September ; 87(9): 916–923. doi:10.1136/jnnp-2015-312272.

Genome-wide association study of sporadic brain arteriovenous malformations

Shantel Weinsheimer^{1,*}, Nasrine Bendjilali^{2,*}, Jeffrey Nelson³, Diana E. Guo³, Jonathan G. Zaroff⁴, Stephen Sidney⁴, Charles E. McCulloch⁵, Rustam Al-Shahi Salman⁶, Jonathan N. Berg⁷, Bobby P. C. Koeleman⁸, Matthias Simon⁹, Azize Boström⁹, Marco Fontanella¹⁰, Carmelo L. Sturiale¹¹, Roberto Pola¹², Alfredo Puca¹¹, Michael T. Lawton¹³, William L. Young^{3,†}, Ludmila Pawlikowska^{3,14}, Catharina J. M. Klijn^{15,‡}, and Helen Kim^{3,5,14,‡} on behalf of the GEN-AVM Consortium

¹Institute of Biological Psychiatry, Mental Health Center, Sct. Hans MHS – Capital Region of Denmark, Roskilde, Denmark ²Department of Mathematics, Rowan University, Glassboro, New Jersey, USA ³Center for Cerebrovascular Research, Department of Anesthesia and Perioperative Care, University of California, San Francisco, San Francisco, California, USA ⁴Kaiser Permanente of Northern California, Division of Research, Oakland, California, USA ⁵Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, California, USA ⁶Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh, UK ⁷Department of Clinical Genetics, University of Dundee, Dundee, UK ⁸Department of Medical Genetics, University Medical Center, Utrecht, The Netherlands ⁹Department of Neurosurgery, University of Bonn Medical Center, Bonn, Germany ¹⁰Division of Neurosurgery, University of Torino, and Division of Neurosurgery, University of Brescia, Brescia, Italy ¹¹Institute of Neurosurgery, Catholic University of Rome, Rome, Italy ¹²Institute of Medicine, Catholic University of Rome, Rome, Italy ¹³Department of Neurological Surgery, University of California, San Francisco, San Francisco,

Correspondence to: Helen Kim, PhD, Department of Anesthesia and Perioperative Care, University of California, San Francisco, 1001 Potrero Avenue, Box 1363, San Francisco, CA 94143, Phone: (415) 206-8906, Helen.Kim2@ucsf.edu.

*These authors contributed equally to the manuscript.

†Deceased

‡Co-senior Authors

Collaborators

Gen-AVM Consortium Collaborating Investigators are listed in the online supplement.

Contributors

DEG, WLY, MTL, HK, JGZ, SS.; MS, AB, MF, CLS, RP, AP, CJMK, RA-SS and JNB collected the data.

WLY, LP, CJMK and HK participated in study design.

DEG, JB, BPCK and LP performed/participated in genetic analysis.

SW, NB, JN and CEM performed statistical analysis.

SW, NB, JN, CEM, BPCK, LP, CJMK and HK interpreted the results.

SW and NB drafted the manuscript.

JN, CEM, DEG, JGZ, SS, BPCK, MS, AB, MF, CLS, RP, AP, RA-SS, JNB, MTL, LP, CJMK and HK edited and reviewed the manuscript.

Competing interests

None declared.

Ethics approval

Approved by the Review Boards of all the institutions involved in this study. Written informed consent obtained from all participants.

Provenance and peer review

Not commissioned; externally peer reviewed.

California, USA ¹⁴Institute for Human Genetics, University of California, San Francisco, San Francisco, California, USA ¹⁵Department of Neurology, Radboud University Medical Center, Nijmegen, The Netherlands

Abstract

Background—The pathogenesis of sporadic brain arteriovenous malformations (BAVM) remains unknown, but studies suggest a genetic component. We estimated the heritability of sporadic BAVM and performed a genome-wide association study (GWAS) to investigate association of common single-nucleotide polymorphisms (SNPs) with risk of sporadic BAVM in the international, multicenter Genetics of Arteriovenous Malformation (GEN-AVM) consortium.

Methods—The Caucasian discovery cohort included 515 BAVM cases and 1,191 controls genotyped using Affymetrix genome-wide SNP arrays. Genotype data was imputed to 1000 Genomes Project data, and well-imputed SNPs (>0.01 minor allele frequency) were analyzed for association with BAVM. Fifty-seven top BAVM-associated SNPs (51 SNPs with $P < 10^{-05}$ or $P < 10^{-04}$ in candidate pathway genes, and 6 candidate BAVM SNPs) were tested in a replication cohort including 608 BAVM cases and 744 controls.

Results—The estimated heritability of BAVM was 17.6% (SE 8.9%, age and sex-adjusted $p = 0.015$). None of the SNPs were significantly associated with BAVM in the replication cohort after correction for multiple testing. Six SNPs had a nominal $P < 0.1$ in the replication cohort and map to introns in *EGFEM1P*, *SP4*, and *CDKAL1* or near *JAG1* and *BNC2*. Of the six candidate SNPs, two in *ACVRL1* and *MMP3* had a nominal $P < 0.05$ in the replication cohort.

Conclusion—We performed the first GWAS of sporadic BAVM in the largest BAVM cohort assembled to date. No GWAS SNPs were replicated, suggesting that common SNPs do not contribute strongly to BAVM susceptibility. However, heritability estimates suggest a modest but significant genetic contribution.

Keywords

brain arteriovenous malformation; genetics; genome-wide association study; risk factor; stroke

INTRODUCTION

Brain arteriovenous malformations (BAVMs) are vascular lesions in which blood shunts directly from the arterial to venous circulation through a nidus with no intervening capillary bed. BAVMs are rare, with a detection rate of 1.3 per 100,000 person-years[1] and a prevalence of 10–18 per 100,000.[2] Patients with BAVM are susceptible to intracranial hemorrhage (ICH), the presenting symptom in half of all patients.[1] The genetic basis for several Mendelian syndromes that display BAVM as part of the phenotype has been defined, including hereditary hemorrhagic telangiectasia (HHT, OMIM #187300) caused by mutations in *ENG*, *ACVRL1*, *SMAD4* and possibly *BMP9*,[3–6] and capillary malformation-arteriovenous malformation (CM-AVM, OMIM #608354) caused by mutations in *RASA1*. [7] Several studies also suggest familial cases of BAVM occur outside

the context of these genetic syndromes.[8, 9] However, the genes underlying these linkage signals have not been identified and the pathogenesis of sporadic BAVM remains elusive.

Candidate gene studies have been the primary approach for evaluating genetic risk of BAVM. Several common single nucleotide polymorphisms (SNPs) have been reported to be associated with BAVM.[10–15] Here we report an estimate of heritability of BAVM and results from the first genome-wide association study (GWAS) of common SNPs in Caucasian BAVM patients, initiated by the Genetics of Arteriovenous Malformation (GEN-AVM) Consortium, a major international effort to better understand the genetics of BAVM.

METHODS

Study population and design

Caucasian BAVM cases and healthy controls were recruited from institutions in the USA, The Netherlands, Germany, Italy, and Scotland. BAVM diagnosis, morphological, and clinical characteristics were recorded using standardized definitions.[16] The diagnosis of AVM was confirmed locally at each site by digital subtraction angiography, pathology, and/or MR or CT angiography. Patients with known diagnoses of HHT or other Mendelian vascular disorders were excluded.

Study cohorts are described below and in more detail in the online supplementary data. Written informed consent was obtained from all participants, and the study was approved by the respective Institutional Review Boards.

Discovery cohort—The discovery cohort started with 556 BAVM cases and 1,250 controls, and was analyzed in two phases. Phase 1 comprised 371 BAVM patients and 563 controls. BAVM cases were recruited from the University of California, San Francisco (UCSF) or Kaiser Permanente of Northern California (KPNC). Shared control data genotyped in the same laboratory as cases included 216 healthy controls participating in a narcolepsy study[17] and 347 transplant donors from a kidney transplantation study.[18] All cases and controls provided either blood or saliva specimens for genetic studies, and were of self-reported Caucasian race/ethnicity.

The Phase 2 discovery cohort started with 185 BAVM cases (149 recruited from the University Medical Center Utrecht, The Netherlands; 36 recruited from UCSF). In the absence of a second set of in-house genotyped controls, we used European Caucasian controls with GWAS data participating in the Wellcome Trust Case Control Consortium (WTCCC) British 1958 Birth Cohort (<http://www.wtccc.org.uk/>). Starting with a set of 2,706 WTCCC controls, we genetically matched controls to the phase 2 BAVM cases using GEMTools (version January 10, 2011).[19] Using 10,527 markers (minor allele frequency (MAF)>5%, $r^2<0.2$), we identified 687 well-matched controls (distance from case to closest control < 0.1).

Replication cohort—The replication cohort initially included 623 BAVM cases and 757 controls, all Caucasian, participating in the GEN-AVM Consortium. 608 cases and 744 controls remained after QC (online supplementary table S1).

Genotyping and quality control (QC) filtering

Discovery cohort—Phase 1 samples were genotyped at the UCSF Genomics Core Facility (GCF) using the Affymetrix® Genome-Wide Human SNP Array 6.0 (Affymetrix, Santa Clara, California), according to the manufacturer's protocols (<http://www.affymetrix.com>). Genotypes for 906,600 SNPs for cases and controls were called together using the Birdseed v2 algorithm implemented in Affymetrix® Genotyping Console™ Software. SNPs with $MAF < 0.01$ or deviating from Hardy-Weinberg (HWE) equilibrium ($P < 10^{-05}$) in controls were excluded. Samples with $> 5\%$ missing genotypes, cryptic duplicates or disagreement between computed and reported gender were discarded. Only autosomal SNPs were analyzed. After QC, the overall average genotyping call rate was 99% in 338 cases and 504 controls. To adjust for potential population stratification in Phase 1, 72,456 unlinked markers distributed uniformly along the genome with $MAF > 5\%$ and low pair-wise linkage disequilibrium (LD) ($r^2 < 0.2$) were used to calculate principal components using EIGENSTRAT v3.0.[20] Genomic inflation factor (λ) and quantile-quantile (Q-Q) plots were used to compare the genome-wide distribution of the test statistic against the expected null distribution.[21]

Phase 2 samples were obtained later and newer technology was in use at the UCSF GCF laboratory; these cases were genotyped using the custom Affymetrix Axiom® Genome-Wide World Array EUR1, containing 675,369 probes.[22] Genotypes were called using Affymetrix® Genotyping Console™ software. Indel polymorphisms (6,252 probes), SNPs with $> 3\%$ missing genotypes, $MAF < 0.01$ or HWE $P < 10^{-05}$ were removed. Samples with call rate $< 97\%$, cryptic duplicates and sex mismatches were dropped from the analysis. WTCCC controls were genotyped in an outside laboratory using the Affymetrix® Genome-Wide Human SNP Array 6.0, with similar QC as described for Phase 1. After QC, there were 177 BAVM cases (144 Utrecht and 33 UCSF) and 687 matched WTCCC controls in the Phase 2 discovery cohort.

Genotype imputation—To combine results across different genotyping platforms, we performed genotype imputation with the 1000 Genomes Project European haplotypes as a reference using MaCH[23] and Minimac,[24] and combining cases and controls together in Phase 1 samples. For Phase 2, cases and controls were imputed separately as they were genotyped on different platforms. Analysis of imputed data (i.e., SNP dosage) was restricted to well-imputed SNPs ($r^2 > 0.8$) in Phases 1 and 2 cohorts with the same direction of effect and with $MAF \geq 1\%$.

Replication cohort—SNPs selected for replication were genotyped at the UCSF GCF using the Illumina Golden Gate Veracode assay and the Assay Design Tool (ADT) (<https://my.illumina.com/custom/UploadVeraCodePrelim>). SNPs with designability scores < 0.6 were replaced with the next most significantly associated SNP in the same LD block if one with a higher designability score (≥ 0.6) was available.

A total of 61 SNPs were selected for genotyping in the replication cohort using the following criteria: (1) 22 SNPs with $P < 10^{-05}$ in meta-analysis of discovery cohorts; (2) 30 SNPs with $P < 10^{-04}$ in meta-analysis and in 752 genes from eight candidate pathways related

to BAVM biology (transforming growth factor (TGF)-beta signaling, notch, vascular endothelial growth factor (VEGF), inflammation, mitogen-activated protein kinase (MAPK), vascular endothelial growth, vascular development, hedgehog); (3) 3 SNPs with $P < 10^{-05}$ in Phase 1 but not passing imputation QC in Phase 2; and (d) 6 candidate SNPs reportedly associated with sporadic BAVM (rs522616, rs2071219, rs1333040, rs10486391, rs1800587, rs1143627).[10–15] Of these, 57 SNPs were successfully genotyped in the replication cohort (call rate $>96.9\%$). Samples with $>10\%$ missing genotypes were excluded, resulting in 608 cases and 744 controls.

Statistical analysis

Heritability analysis—We computed the narrow-sense heritability (both unadjusted and age and sex adjusted) of BAVM susceptibility using Phase 1 samples as these cases and controls were genotyped on the same platform and lab. The variance or liability explained by common SNPs on the array was estimated using an expectation maximization algorithm implemented within GCTA (Genome-wide Complex Trait Analysis, version 1.13).[25, 26] Prior to analysis, we removed: (1) SNPs with MAF <0.01 , (2) SNPs with missingness >0.02 , (3) subjects with >0.02 missing genotypes, (4) SNPs with significant differences ($P < 0.05$) in missingness between cases and controls, (5) individuals more than 5 standard deviations away from the mean of the first two principal components, (6) SNPs out of HWE ($P < 0.05$) within cases, within controls, and across all individuals, and (7) one subject from each pair with high genome-wide similarities (relatedness >0.025). This resulted in a final sample size for heritability analysis that differed slightly from the GWAS Phase 1 analysis (described below), and included 297 cases, 468 controls, and 484,737 SNPs. Additionally, we performed a permutation test by permuting case/control status (or case/control status and age and sex simultaneously for the adjusted estimates) 1,000 times to determine whether heritability was significantly greater than zero.

GWAS analysis—Logistic regression analysis of SNP dosages was performed using an additive model adjusting for age, sex, and the top three principal components in Phase 1 and adjusting for sex in Phase 2 (age was not available for WTCCC controls, and samples were genetically matched). To combine Phase 1 and Phase 2 results, we performed an effect-size based meta-analysis using METAL software (version 2011-03–25).[27]

In replication, we considered GWAS SNPs with one-sided $P < 0.05$ with the same direction of effect in the replication cohort and in the meta-analysis of discovery cohorts as nominally statistically significant. We computed the minimum detectable odds ratios (OR) at 80% power for minor allele frequencies ranging from 1% to 50% using the CaTS power calculator.[28] Our study was powered to detect ORs ranging from 1.3–3.1 for SNPs with MAF between 0.01–0.50 (online supplementary table S2). Permutation P-values were used to correct for multiple comparisons. Each SNP was regressed on age, sex, and cohort using linear regression and corresponding sets of residuals statistically uncorrelated with age, sex, and cohort were calculated and then permuted. Next, logistic regression analysis of case/control status was run with age, sex, cohort, and permuted residuals as predictors. P-values for the residual effect for the 51 SNPs selected from GWAS meta-analysis were computed based on the minimum permutation P-value across all SNPs within each of 1,000

permutations. The 6 candidate SNPs were analyzed separately, assuming the same direction of effect as published papers and applying the same permutation method as described above.

Sensitivity analysis was performed to evaluate the effect of including four individuals in the study who were subsequently found to have an HHT diagnosis (3 in discovery cohort and 1 in replication cohort). Excluding these individuals did not significantly alter results (data not shown).

Functional analysis—Bioinformatic evaluation of the functional impact of top SNPs from the replication analysis and SNPs in LD with them ($r^2 > 0.8$) was performed using RegulomeDB (<http://regulome.stanford.edu>) for evaluation of regulatory potential of noncoding SNPs and GTEx Database Portal (<http://www.gtexportal.org>) for gene expression levels and expression quantitative trait loci (eQTL) in relevant tissues (brain and artery).

RESULTS

Discovery and replication cohort demographics are summarized in table 1 and in online supplementary table S1. The mean age of BAVM cases was around 40 years and approximately 50% were male, similar for all cohorts. Hemorrhage at presentation was higher in the replication cohort compared to the discovery cohort (55% vs. 38%, $P < 0.001$).

Heritability analysis

The unadjusted estimate of heritability of BAVM was 19.3% (SE 8.9%, $P = 0.004$) and the age and sex-adjusted estimate was 17.6% (SE 8.9%, $P = 0.015$), suggesting that heritability due to additive genetic effects is significantly greater than zero.

GWAS in discovery cohort

A total of 4,300,568 SNPs with $MAF > 0.01$ were well-imputed ($MACH r^2 > 0.8$) in both Phase 1 and Phase 2. GWAS results are summarized in Manhattan plots for Phase 1 (figure 1A) and Phase 2 (figure 1B). In Phase 1, Q-Q analysis indicated potential population sub-structure ($\lambda = 1.11$, online supplementary figure S1), which was reduced by including the top three principal components as covariates ($\lambda = 1.04$, online supplementary figure S1). Six SNPs were marginally associated with BAVM at $P < 0.0001$ (figure 1A, table 2). In Phase 2, five SNPs were marginally associated with BAVM at $P < 0.0001$ (figure 1B, table 2), including one SNP that met genome-wide level of significance (rs2292155, intron *ZNF423*, $P = 3.49 \times 10^{-24}$). However, this SNP was not significantly associated with BAVM in Phase 1 ($P = 0.029$).

Results from meta-analysis of the combined discovery cohort of 515 BAVM cases and 1,191 controls are summarized in figure 1C and table 2. One intronic SNP, rs2292155 in *ZNF423*, was significantly associated with BAVM ($OR = 0.54$, $P = 2.88 \times 10^{-10}$) in the meta-analysis. An additional 20 SNPs had $P < 10^{-5}$ including 10 SNPs mapping within genes (*SLC22A20*, *POLA2*, *SP4*, *DNAH11*, *CDKAL1* and *SP110*).

Replication analysis

A total of 61 SNPs were selected for replication (22 SNPs with $P < 10^{-05}$ in meta-analysis of discovery cohorts; 30 SNPs with $P < 10^{-04}$ in meta-analysis and in candidate pathways related to BAVM biology; 3 SNPs with $P < 10^{-05}$ in Phase 1 but not passing imputation QC in Phase 2; and 6 candidate SNPs reported to be associated with sporadic BAVM). Fifty-seven SNPs were successfully genotyped in the replication cohort. None of these SNPs were significantly associated with BAVM after correction for multiple comparisons (table 2). Six SNPs with nominal one-sided $P < 0.10$ mapped to the following genes/regions: rs9863784 (intergenic *LOC389174/EGFEMIP*), rs6040426 (intergenic *JAG1/LOC728573*), rs1332479 (intergenic *BNC2/CCDC171*), rs10233357 (intron *SP4*), rs9460577 and rs9465934 (intron 9 *CDKALI*). The top BAVM-associated SNP from the discovery cohort (*ZNF423* rs2292155) was not associated with BAVM in the replication cohort ($P = 0.746$). Further inspection of the SNP cluster plot for rs2292155 revealed a genotype clustering error, resulting in a falsely low MAF in Phase 2 cases that was driving the association.

Two of the six previously reported candidate SNPs showed nominally significant associations with BAVM (*ACVRL1* rs2071219, OR=0.83, $P = 0.014$, and *MMP3* rs522616, OR=0.82, $P = 0.023$, table 2). However, these associations did not survive correction for multiple testing.

Functional analysis *in silico*

We queried the RegulomeDB for the top six SNPs from the replication analysis and found that four SNPs (rs17144483, rs7033995, rs62551099, and rs116744349) in LD with the top SNPs ($r^2 > 0.8$ in the 1000 Genomes Caucasians (Utah residents with ancestry from northern and western Europe, CEU)) are likely to affect regulatory protein binding (online supplementary table S3A).

All top six loci contain genes that are expressed in brain and/or artery tissue according to GTEx database (online supplementary table S3B). For example, *SP4* (SNP rs10233357) is highly expressed in brain tissue and moderately expressed in artery tissue, but was not differentially expressed in the blood of BAVM patients compared to controls.[29] Known eQTLs for each locus (defined by gene/nearest gene) include: *EGFEMIP* (n=2), *HRHI* (n=1), *SP4* (n=1), *BNC2* (n=3) and *CDKALI* (n=14) (online supplementary table S3C), but none are in high LD with the top BAVM-associated SNPs. Only one of these eQTL (rs9358372, *CDKALI*) regulates expression in vascular tissue (artery) (online supplementary table S3C).

DISCUSSION

In order to evaluate the role of common genetic variation in sporadic BAVM, we assembled the largest multi-national sporadic BAVM cohort studied to date, comprising 515 Caucasian cases and 1,191 controls in the discovery phase and 608 Caucasian cases and 744 controls in the replication phase, and performed the first GWAS of sporadic BAVM. We used the genome-wide common variation data to estimate sporadic BAVM heritability, and found a significant genetic influence to BAVM (18%, SE 8.9%). This heritability estimate is in the

range of what has been reported for other vascular diseases using similar methods. For example, heritability estimates of primary ICH was estimated at 29% (SE 11%) for non-apolipoprotein E (APOE) loci and 15% (SE 10%) for APOE.[30] Thus, our heritability analysis suggests a modest but significant genetic influence to BAVM.

Our GWAS meta-analysis identified 51 SNPs associated at a $P < 10^{-4}$ threshold. However, upon replication none of these findings met the corrected threshold for significance. Six SNPs showed a trend ($P < 0.1$) toward association with BAVM in the replication cohort. Two of these are intronic SNPs in *CDKAL1* encoding CDK5 regulatory subunit associated protein 1-like 1 with unknown function. GWAS studies have linked *CDKAL1* with susceptibility to type 2 diabetes.[31] Although *CDKAL1* is highly expressed in artery tissue (online supplementary table S3B), we did not detect differential expression in the blood of BAVM cases compared to controls.[29] Interestingly, *CDKN2B-AS1* is a non-coding RNA located within the *CDKN2A-CDKN2B* gene cluster on 9p21, the strongest genetic locus for cardiovascular diseases and linked to intracranial aneurysm.[32] More recently, the 9p21 SNP rs1333040 has been associated with BAVM,[15, 33] and replicated in our discovery cohort (Phase 1). However, this genetic association appears to be explained by the presence of BAVM-associated aneurysms.[33, 34]

Genes near the other top loci included: *EGFEM1P*, *HRH1*, *JAG1*, *SP4*, *CCDC171* and *BNC2*, several of which are implicated in vascular biology. *EGFEM1P* is a pseudogene. *HRH1* encodes an integral membrane protein, which is a G-protein coupled receptor that mediates capillary permeability.[35] *JAG1* is the ligand for the NOTCH1 receptor and plays a role in hematopoiesis; *JAG1* mutations cause Alagille syndrome, an autosomal dominant disorder with diverse clinical features including vascular anomalies with significant morbidity and mortality.[36] *SP4* encodes a transcription factor. *BNC2* encodes basonuclin 2, which is essential during embryogenesis.[37]

The five genes at the top BAVM-associated loci are all moderately expressed in brain or artery tissue, and four SNPs in high LD with top-associated SNPs in *CDKAL1*, *SP4*, and *BNC2* are predicted to have a likely role in regulatory protein binding. In addition, while several eQTL have been reported within the top loci, only one is associated with gene expression in a relevant tissue (artery, *CDKAL1*).

Two of the six previously reported BAVM-associated candidate SNPs showed nominal association in our replication cohort: *ACVRL1* IVS3–35 A>G and *MMP3* rs522616. The *ACVRL1* IVS3–35 A>G association was originally reported in two of the cohorts included in our study (UCSF and Germany),[10, 38] but was not associated in the Dutch cohort.[39] However, a prior meta-analysis combining the three cohorts revealed that the *ACVRL1* variant remained associated with BAVM.[39] Our results confirm that the *ACVRL1* IVS3–35 A>G association with BAVM appears to be present in multiple, but not all, Caucasian populations. Recently, the same SNP has been reported to be associated with AVMs in HHT.[40] The *MMP3* rs522616 association was previously reported in a Chinese BAVM case-control study.[13] Our study is the first replication of that finding, and represents only the second common polymorphism, after *ACVRL1* IVS3–35 A>G, identified in candidate gene studies that is associated with BAVM in more than one cohort; and the first that may be

associated with sporadic BAVM in two ethnic groups: Caucasians and Chinese. Association of promoter activity with *MMP3* rs522616 genotype has been previously reported.[13] The lack of association with BAVM for the other 4 of 6 previously reported candidate SNPs in our cohort could represent non-replication, or could be due to population differences, low power to replicate small effects, and, in the case of the 9p21 SNP (rs1333040), differences in the prevalence of associated feeding artery aneurysms.[33, 34]

Conclusion

In summary, heritability estimates based on genome-wide common variation suggest a significant genetic influence in sporadic BAVM. However, our study did not identify any common SNPs representing strong genetic risk factors for BAVM in Caucasians. Several of the top GWAS hits implicated genes involved in vascular biology and may represent smaller effects that we were not powered to reliably detect. Taken together with our previous copy number variation analysis, these results suggest that common genetic variation is not a major risk factor for BAVM in Caucasians. Other potential genetic mechanisms may nonetheless contribute to sporadic BAVM, including modest effect common variants, such as the six suggestive but non-significant loci revealed by our replication analysis, rare genetic variants, or somatic or epigenetic variation. Larger cohorts and different study designs will be required to evaluate these other hypotheses.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors would like to acknowledge the important contributions of the late William L. Young, MD, who pioneered and supported BAVM genetic studies, and the research staff at respective sites for their assistance with database management and manuscript preparation. They also thank all patients with BAVM for their participation in research studies.

Funding

USA studies: NIH grants P01 NS044155, R01 NS034949, K23 NS058357 for AVM cases and controls; P50 NS2372 and U19 AI063603 for shared GWAS control data.

The Netherlands AVM cohort: Van Leersum Fund, Royal Netherlands Academy of Arts and Sciences; a grant from *Running-for-Nona*. CJMK is supported by a clinical established investigator grant from the Dutch Heart Foundation (2012T077) and ASPASIA grant from The Netherlands Organisation for Health Research and Development, ZonMw (015008048).

German AVM study: Funded in part by grants to MS from the University of Bonn (BONFOR).

Italian AVM cohorts: D1 intramural grant from the Catholic University School of Medicine, Rome, Italy; grants from the *Ministero dell'Università e della Ricerca Scientifica (MURST)* of Italy; intramural grant from the University of Brescia, Italy.

SIVMS study: Supported by the Chief Scientist Office of the Scottish Government Health Department; project grant from The Stroke Association; MRC clinical training, clinician scientist and senior clinical fellowships to RA-SS.

References

1. Al-Shahi R, Bhattacharya JJ, Currie DG, et al. Prospective, population-based detection of intracranial vascular malformations in adults: the Scottish Intracranial Vascular Malformation Study (SIVMS). *Stroke*. 2003; 34:1163–1169. [PubMed: 12702837]
2. Al-Shahi R, Fang JS, Lewis SC, et al. Prevalence of adults with brain arteriovenous malformations: a community based study in Scotland using capture-recapture analysis. *J Neurol Neurosurg Psychiatry*. 2002; 73:547–551. [PubMed: 12397149]
3. McAllister KA, Grogg KM, Johnson DW, et al. Endoglin, a TGF-beta binding protein of endothelial cells, is the gene for hereditary haemorrhagic telangiectasia type 1. *Nat Genet*. 1994; 8:345–351. [PubMed: 7894484]
4. Johnson DW, Berg JN, Baldwin MA, et al. Mutations in the activin receptor-like kinase 1 gene in hereditary haemorrhagic telangiectasia type 2. *Nat Genet*. 1996; 13:189–195. [PubMed: 8640225]
5. Gallione CJ, Repetto GM, Legius E, et al. A combined syndrome of juvenile polyposis and hereditary haemorrhagic telangiectasia associated with mutations in MADH4 (SMAD4). *Lancet*. 2004; 363:852–859. [PubMed: 15031030]
6. Wooderchak-Donahue WL, McDonald J, O'Fallon B, et al. BMP9 mutations cause a vascular-anomaly syndrome with phenotypic overlap with hereditary hemorrhagic telangiectasia. *Am J Hum Genet*. 2013; 93:530–537. [PubMed: 23972370]
7. Eerola I, Boon LM, Mulliken JB, et al. Capillary malformation-arteriovenous malformation, a new clinical and genetic disorder caused by RASA1 mutations. *Am J Hum Genet*. 2003; 73:1240–1249. [PubMed: 14639529]
8. van Beijnum J, van der Worp HB, Schippers HM, et al. Familial occurrence of brain arteriovenous malformations: a systematic review. *J Neurol Neurosurg Psychiatry*. 2007; 78:1213–1217. [PubMed: 17259353]
9. van Beijnum J, van der Worp HB, Algra A, et al. Prevalence of brain arteriovenous malformations in first-degree relatives of patients with a brain arteriovenous malformation. *Stroke*. 2014; 45:3231–3235. [PubMed: 25236872]
10. Pawlikowska L, Tran MN, Achrol AS, et al. Polymorphisms in transforming growth factor-beta-related genes ALK1 and ENG are associated with sporadic brain arteriovenous malformations. *Stroke*. 2005; 36:2278–2280. [PubMed: 16179574]
11. Kim H, Hysi PG, Pawlikowska L, et al. Common variants in interleukin-1-beta gene are associated with intracranial hemorrhage and susceptibility to brain arteriovenous malformation. *Cerebrovasc Dis*. 2009; 27:176–182. [PubMed: 19092239]
12. Su H, Kim H, Pawlikowska L, et al. Reduced expression of integrin alphavbeta8 is associated with brain arteriovenous malformation pathogenesis. *Am J Pathol*. 2010; 176:1018–1027. [PubMed: 20019187]
13. Zhao Y, Li P, Fan W, et al. The rs522616 polymorphism in the matrix metalloproteinase-3 (MMP-3) gene is associated with sporadic brain arteriovenous malformation in a Chinese population. *J Clin Neurosci*. 2010; 17:1568–1572. [PubMed: 20822909]
14. Fontanella M, Rubino E, Crobbedu E, et al. Brain arteriovenous malformations are associated with interleukin-1 cluster gene polymorphisms. *Neurosurgery*. 2012; 70:12–17. [PubMed: 21768915]
15. Sturiale CL, Fontanella MM, Gatto I, et al. Association between polymorphisms rs1333040 and rs7865618 of chromosome 9p21 and sporadic brain arteriovenous malformations. *Cerebrovasc Dis*. 2014; 37:290–295. [PubMed: 24820060]
16. Atkinson RP, Awad IA, Batjer HH, et al. Reporting terminology for brain arteriovenous malformation clinical and radiographic features for use in clinical trials. *Stroke*. 2001; 32:1430–1442. [PubMed: 11387510]
17. Hallmayer J, Faraco J, Lin L, et al. Narcolepsy is strongly associated with the T-cell receptor alpha locus. *Nat Genet*. 2009; 41:708–711. [PubMed: 19412176]
18. Flechner SM, Goldfarb D, Solez K, et al. Kidney transplantation with sirolimus and mycophenolate mofetil-based immunosuppression: 5-year results of a randomized prospective trial compared to calcineurin inhibitor drugs. *Transplantation*. 2007; 83:883–892. [PubMed: 17460558]

19. Luca D, Ringquist S, Klei L, et al. On the use of general control samples for genome-wide association studies: genetic matching highlights causal variants. *Am J Hum Genet.* 2008; 82:453–463. [PubMed: 18252225]
20. Price AL, Patterson NJ, Plenge RM, et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet.* 2006; 38:904–909. [PubMed: 16862161]
21. Devlin B, Roeder K. Genomic control for association studies. *Biometrics.* 1999; 55:997–1004. [PubMed: 11315092]
22. Hoffmann TJ, Kvale MN, Hesselton SE, et al. Next generation genome-wide association tool: Design and coverage of a high-throughput European-optimized SNP array. *Genomics.* 2011; 98:79–89. [PubMed: 21565264]
23. Li Y, Willer CJ, Ding J, et al. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet Epidemiol.* 2010; 34:816–834. [PubMed: 21058334]
24. Howie B, Fuchsberger C, Stephens M, et al. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet.* 2012; 44:955–959. [PubMed: 22820512]
25. Yang J, Lee SH, Goddard ME, et al. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet.* 2011; 88:76–82. [PubMed: 21167468]
26. Lee SH, Wray NR, Goddard ME, et al. Estimating missing heritability for disease from genome-wide association studies. *Am J Hum Genet.* 2011; 88:294–305. [PubMed: 21376301]
27. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics.* 2010; 26:2190–2191. [PubMed: 20616382]
28. Skol AD, Scott LJ, Abecasis GR, et al. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat Genet.* 2006; 38:209–213. [PubMed: 16415888]
29. Weinsheimer S, Xu H, Achrol AS, et al. Gene expression profiling of blood in brain arteriovenous malformation patients. *Transl Stroke Res.* 2011; 2:575–587. [PubMed: 22184505]
30. Devan WJ, Falcone GJ, Anderson CD, et al. Heritability estimates identify a substantial genetic contribution to risk and outcome of intracerebral hemorrhage. *Stroke.* 2013; 44:1578–1583. [PubMed: 23559261]
31. Peng F, Hu D, Gu C, et al. The relationship between five widely-evaluated variants in CDKN2A/B and CDKAL1 genes and the risk of type 2 diabetes: a meta-analysis. *Gene.* 2013; 531:435–443. [PubMed: 24012816]
32. Tromp G, Weinsheimer S, Ronkainen A, et al. Molecular basis and genetic predisposition to intracranial aneurysm. *Ann Med.* 2014; 46:597–606. [PubMed: 25117779]
33. Bendjilali N, Nelson J, Weinsheimer S, et al. Common variants on 9p21.3 are associated with brain arteriovenous malformations with accompanying arterial aneurysms. *J Neurol Neurosurg Psychiatry.* 2014; 85:1280–1283. [PubMed: 24777168]
34. Kremer PH, Koeleman BP, Pawlikowska L, et al. Evaluation of genetic risk loci for intracranial aneurysms in sporadic arteriovenous malformations of the brain. *J Neurol Neurosurg Psychiatry.* 2015; 86:524–529. [PubMed: 25053769]
35. Lu C, Diehl SA, Noubade R, et al. Endothelial histamine H1 receptor signaling reduces blood-brain barrier permeability and susceptibility to autoimmune encephalomyelitis. *Proc Natl Acad Sci U S A.* 2010; 107:18967–18972. [PubMed: 20956310]
36. Turnpenny PD, Ellard S. Alagille syndrome: pathogenesis, diagnosis and management. *Eur J Hum Genet.* 2012; 20:251–257. [PubMed: 21934706]
37. Vanhoutteghem A, Maciejewski-Duval A, Bouche C, et al. Basonuclin 2 has a function in the multiplication of embryonic craniofacial mesenchymal cells and is orthologous to disco proteins. *Proc Natl Acad Sci U S A.* 2009; 106:14432–14437. [PubMed: 19706529]
38. Simon M, Franke D, Ludwig M, et al. Association of a polymorphism of the ACVRL1 gene with sporadic arteriovenous malformations of the central nervous system. *J Neurosurg.* 2006; 104:945–949. [PubMed: 16776339]
39. Boshuisen K, Brundel M, de Kovel CG, et al. Polymorphisms in ACVRL1 and endoglin genes are not associated with sporadic and HHT-related brain AVMs in Dutch patients. *Transl Stroke Res.* 2013; 4:375–378. [PubMed: 24323303]

40. Pawlikowska L, Nelson J, Guo DE, et al. The ACVRL1 c.314-35A>G polymorphism is associated with organ vascular malformations in hereditary hemorrhagic telangiectasia patients with ENG mutations, but not in patients with ACVRL1 mutations. *Am J Med Genet A*. 2015; 167:1262–1267. [PubMed: 25847705]

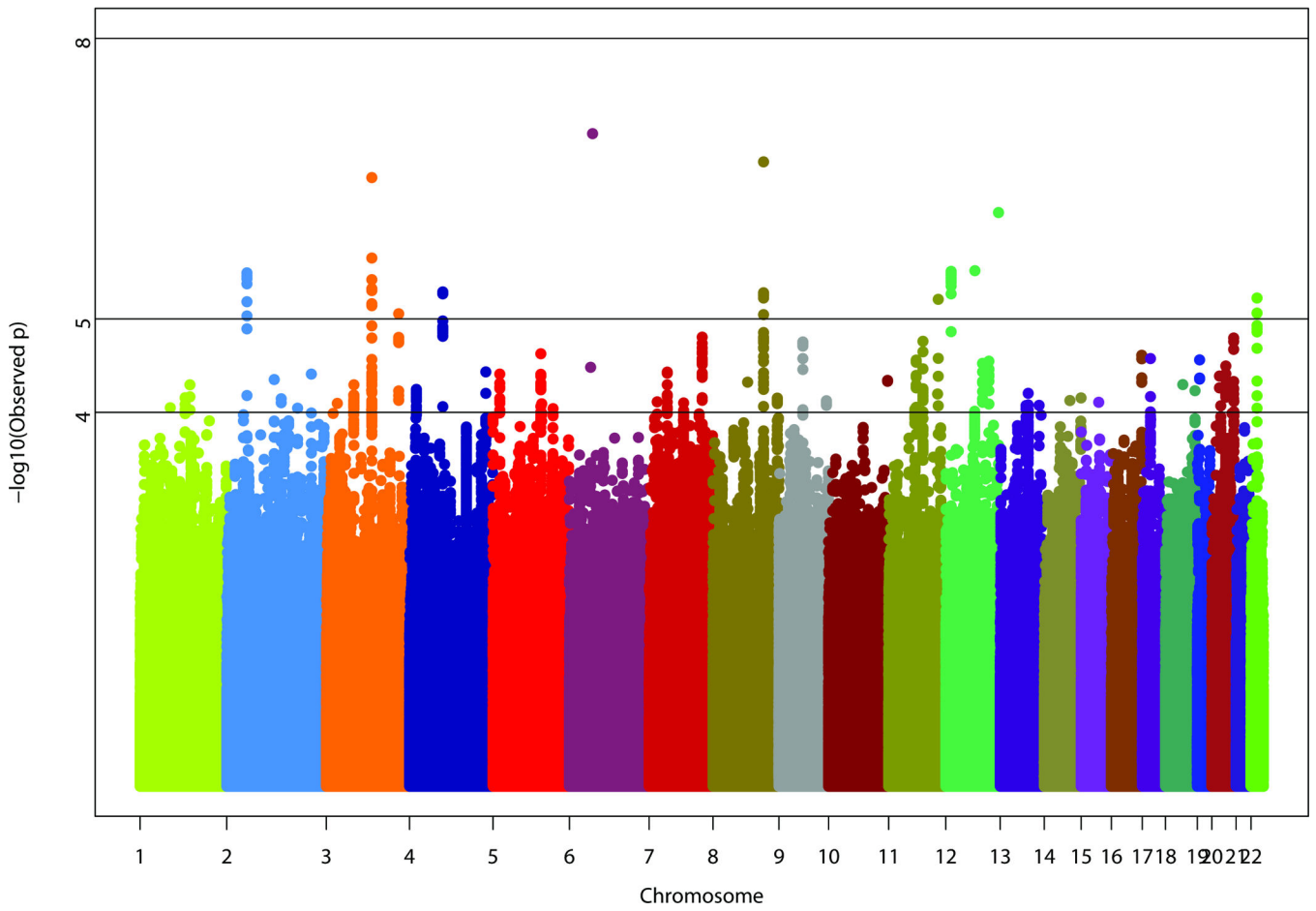
Author Manuscript

Author Manuscript

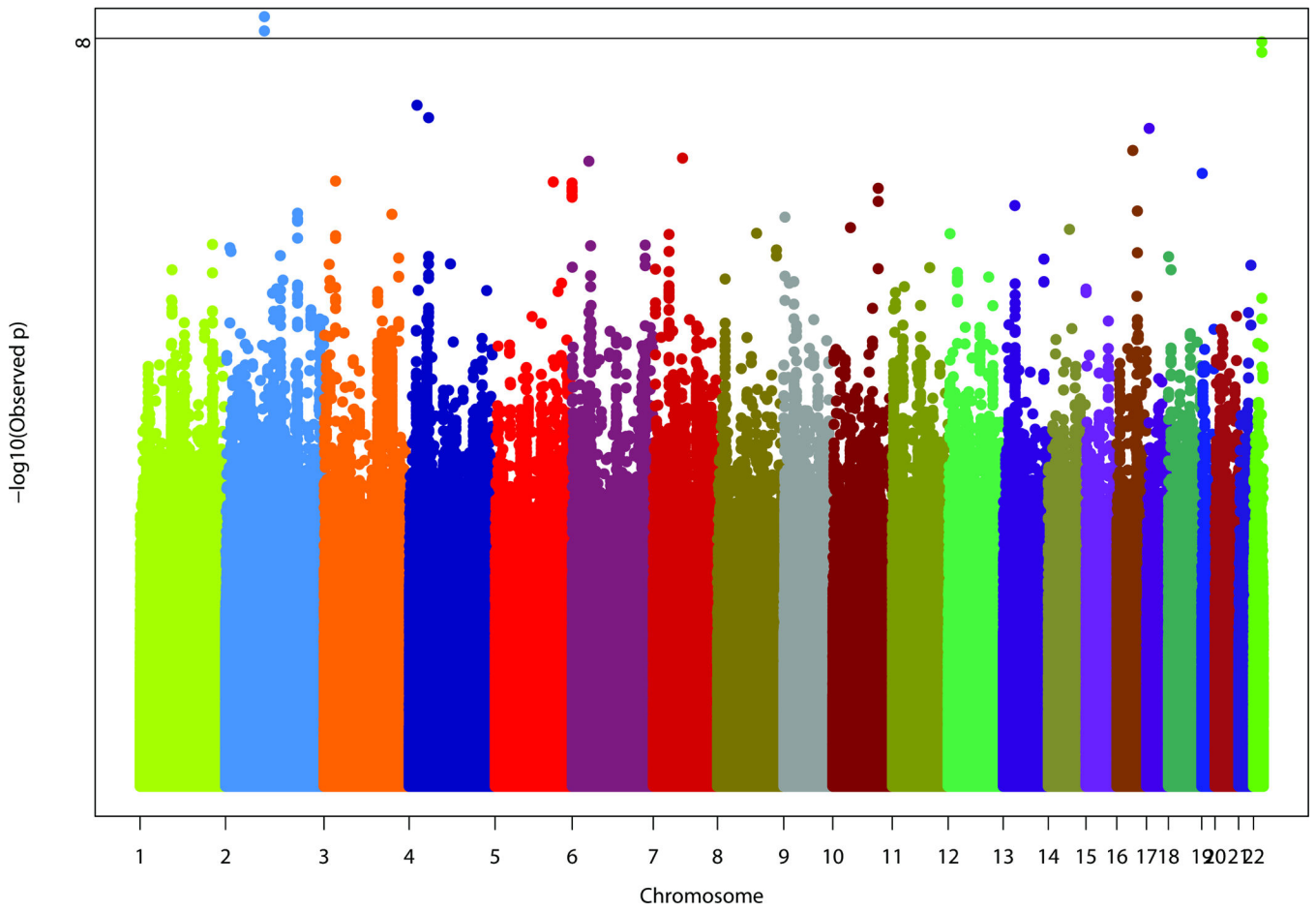
Author Manuscript

Author Manuscript

A



B



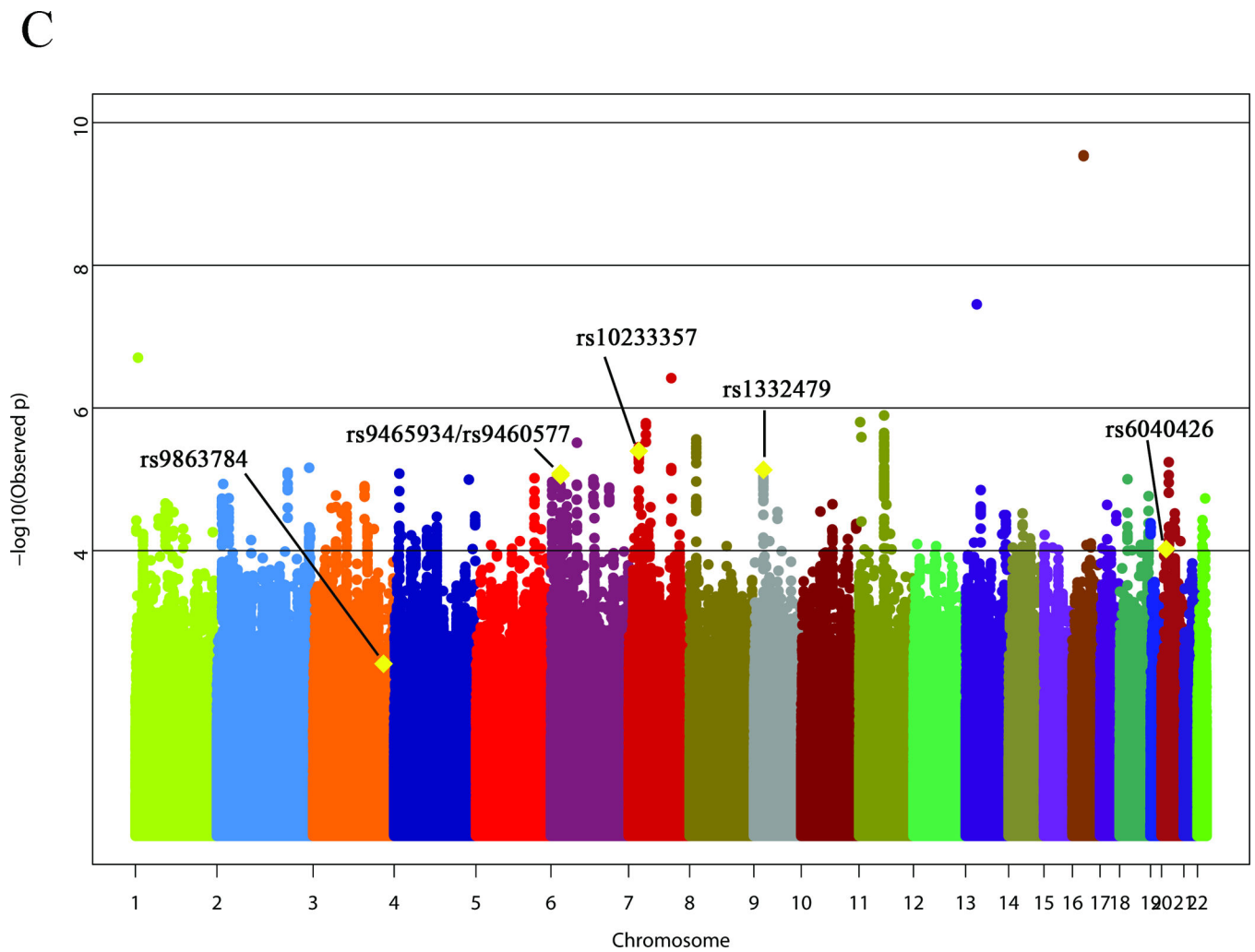


Figure 1. Results of the genome-wide association analysis for brain arteriovenous malformation susceptibility

(A) Discovery Phase 1; (B) Discovery Phase 2; (C) Meta-analysis. Data plotted includes imputed P-values (i.e., the association P-value corresponds to the imputed dosages). The six top single nucleotide polymorphisms from the replication analysis are shown on the meta-analysis plot (C).

Table 1

BAVM cases and controls in the GEN-AVM discovery and replication cohorts.

	Discovery Phase 1		Discovery Phase 2		Replication	
	Case	Control	Case	Control	Case	Control
Sample size	338	504	177	687	608	744
Male (%)	45.6	50.2	54.2	50.7	54.0	56.2
Age (years), mean (SD)	38.7 (17.6)	49.0 (14.0)	38.3 (16.1)	40.0 (0)	42.9 (17.3)	40.7 (16.8)
ICH* (%)	37.6	NA	46.9	NA	54.7	NA
AVM size (cm) ^{**} , mean (SD)	3.0 (1.6)	NA	2.9 (1.7)	NA	2.7 (1.3)	NA

* ICH (%) available for all cases in discovery cohort, and 594/608 cases in the replication cohort

** Maximum AVM nidus diameter in centimeters; measurement available for 260/338 cases in phase 1, 31/177 cases in phase 2, and 560/851 cases in the replication cohort

AVM, arteriovenous malformation; BAVM, brain arteriovenous malformation; ICH, intracerebral hemorrhage; NA, not available

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

SNP	AI	chr	gene [†]	*Reason for inclusion	Discovery Cohort										Replication Cohort (668 cases / 744 controls)														
					Phase I (538 cases / 504 controls)					Phase 2 (177 cases / 687 controls)					Meta Analysis														
					MAFcase	MAFcontrol	OR	95% CI	P	MAFcase	MAFcontrol	OR	95% CI	P	OR	95% CI	P	MAFcase	MAFcontrol	OR	95% CI	P _{2-sided}	P _{1-sided}	P _{permutation}					
rs11968071	G	6	SLC22A23, PXDC1	b	0.08	0.06	1.66	1.06-2.58	2.65E-02	0.13	0.07	2.22	1.49-3.31	1.52E-04	1.95	1.45-2.62	1.09E-05	0.09	0.97	0.74-1.26	0.811	0.594	1.000						
rs17144860	A	7	DNAH11	a	0.06	0.03	2.11	1.25-3.58	4.92E-03	0.07	0.03	2.68	1.58-4.52	4.15E-04	2.38	1.64-3.45	5.03E-06	0.04	0.95	0.64-1.39	0.788	0.606	1.000						
rs7165109	A	15	ROSA, VPS13C	b	0.09	0.06	1.63	1.07-2.50	2.35E-02	0.10	0.05	1.93	1.29-2.88	1.78E-03	1.78	1.33-2.39	9.77E-05	0.07	0.96	0.71-1.29	0.786	0.607	1.000						
rs314076	G	11	SLC22A20	a	0.16	0.23	0.61	0.46-0.81	4.01E-04	0.16	0.24	0.59	0.43-0.82	8.77E-04	0.60	0.49-0.74	2.62E-06	0.20	1.03	0.85-1.25	0.756	0.622	1.000						
rs1037475	A	4	METAP1	b	0.50	0.45	1.29	1.04-1.59	2.10E-02	0.50	0.40	1.47	1.17-1.85	8.85E-04	1.37	1.17-1.60	7.97E-05	0.44	0.97	0.83-1.14	0.756	0.632	1.000						
rs11731878	C	4	METAP1	b	0.49	0.42	1.29	1.04-1.60	2.05E-02	0.50	0.39	1.51	1.20-1.90	4.50E-04	1.39	1.19-1.63	4.59E-05	0.44	0.97	0.83-1.14	0.726	0.637	1.000						
rs10485548	C	20	MACROD2, RIF16B	a	0.04	0.03	1.60	0.92-2.80	9.74E-02	0.10	0.04	2.81	1.79-4.41	1.46E-05	2.25	1.58-3.19	5.76E-06	0.05	0.94	0.67-1.31	0.712	0.644	1.000						
rs7651994	A	3	XIRP1, CX3CR1	b	0.10	0.07	1.71	1.12-2.61	1.28E-02	0.12	0.07	2.07	1.37-3.12	7.83E-04	1.89	1.40-2.53	2.53E-05	0.08	0.95	0.72-1.25	0.711	0.644	1.000						
rs12456228	A	18	CHST9, CDH2	b	0.07	0.04	1.84	1.09-3.11	2.15E-02	0.10	0.05	2.24	1.43-3.50	6.17E-04	2.06	1.47-2.90	2.95E-05	0.07	0.90	0.67-1.22	0.514	0.743	1.000						
rs2292155	T	16	ZNF423	a	0.34	0.40	0.78	0.62-0.98	2.91E-02	0.11	0.37	0.21	0.14-0.30	3.49E-24	0.54	0.45-0.66	2.88E-10	0.37	1.06	0.90-1.24	0.508	0.746	1.000						
rs7004769	A	8	LOC157273	a	0.17	0.22	0.63	0.48-0.83	6.96E-04	0.13	0.22	0.57	0.41-0.79	4.83E-04	0.60	0.49-0.75	2.76E-06	0.21	1.07	0.88-1.29	0.501	0.749	1.000						
rs10111721	C	8	PKHDLL1	c	0.00	0.02	0.01	0.00-0.27	9.02E-06	0.01	0.01	1.11	0.36-3.40	8.59E-01	0.71	0.24-2.08	5.34E-01	0.01	1.32	0.62-2.81	0.475	0.762	1.000						
rs10274813	A	7	CCM2	b	0.03	0.01	3.04	1.33-6.95	6.50E-03	0.04	0.02	2.68	1.38-5.20	5.27E-03	2.82	1.68-4.72	8.50E-05	0.02	0.84	0.52-1.35	0.471	0.764	1.000						
rs77960920	T	22	PHF21B	b	0.07	0.11	0.56	0.38-0.83	2.68E-03	0.06	0.11	0.44	0.27-0.73	4.43E-04	0.51	0.38-0.70	1.85E-05	0.11	1.14	0.89-1.45	0.314	0.843	1.000						
rs1028492	T	22	PHF21B	b	0.07	0.12	0.57	0.39-0.84	2.98E-03	0.06	0.12	0.52	0.33-0.82	2.50E-03	0.55	0.41-0.73	5.83E-05	0.12	1.11	0.89-1.44	0.307	0.847	1.000						
rs62410786	C	4	LOC152742, LOC441009	a	0.06	0.12	0.48	0.33-0.71	1.10E-04	0.07	0.11	0.54	0.33-0.86	6.06E-03	0.50	0.37-0.68	8.31E-06	0.11	0.99	0.94-1.55	0.143	0.928	1.000						
rs10136495	G	14	NID2, PTGDR	b	0.29	0.33	0.77	0.61-0.98	3.08E-02	0.24	0.34	0.62	0.47-0.81	2.80E-04	0.70	0.59-0.84	8.77E-05	0.33	1.27	1.07-1.49	0.006	0.997	1.000						
rs2071219	C	6	ACVRL1	d	0.42	0.42	0.98	0.75-1.27	0.88	0.42	0.45	0.84	0.65-1.08	0.18	0.91	0.75-1.09	0.29	0.41	0.83	0.71-0.98	0.028	0.014	0.077						
rs522616	G	1	MME3	d	0.24	0.21	1.15	0.90-1.47	0.28	0.22	0.23	0.93	0.70-1.23	0.61	1.05	0.87-1.26	0.63	0.17	0.82	0.67-1.00	0.046	0.023	0.125						
rs1333040	C	8	CDKN2B-AS1	d	0.39	0.44	0.79	0.63-0.99	0.04	0.44	0.44	1.03	0.81-1.32	0.80	0.89	0.76-1.05	0.18	0.40	0.93	0.80-1.08	0.356	0.168	0.669						
rs10486391	T	6	ITGB8	d	0.38	0.45	0.84	0.68-1.04	0.11	0.41	0.41	0.99	0.78-1.26	0.95	0.91	0.77-1.06	0.22	0.40	0.97	0.83-1.14	0.702	0.351	0.918						
rs1143627	A	7	IL1B	d	0.38	0.35	1.20	0.95-1.51	0.13	0.32	0.32	0.99	0.77-1.28	0.95	1.10	0.92-1.30	0.29	0.34	1.00	0.85-1.19	0.955	0.477	0.983						
rs1800587	A	7	IL1A	d	0.28	0.30	0.96	0.76-1.21	0.72	0.29	0.30	0.98	0.76-1.26	0.86	0.97	0.82-1.15	0.70	0.29	0.97	0.82-1.15	0.711	0.644	0.997						

*Reasons for inclusion are: (a) $P < 10^{-5}$ in meta-analysis, (b) $P < 10^{-4}$ in meta-analysis and in candidate pathway, (c) $P < 10^{-5}$ in Phase 1 but not passing imputation quality control in Phase 2, and (d) candidate SNP

[†]Gene annotation based on HG-19. SNPs with two genes listed are intergenic

AI, minor allele; BAVM, brain arteriovenous malformation; chr, chromosome; MAF, minor allele frequency; P_{1-sided}, one-sided P-value based on directionality in discovery cohort; P_{permutation}, permutation P-value adjusting for multiple comparisons; SNP, single nucleotide polymorphism