Genome-Wide Association Study of Brain Arteriolosclerosis

Lincoln M.P. Shade1

Yuriko Katsumata1,2

Timothy J. Hohman3

Julie A. Schneider4

Kwangsik Nho

Andrew Saykin

Shubhabrata Mukherjee

Alzheimer’s Disease Genetics Consortium

Peter T. Nelson2,5

David W. Fardo1,2

2021-03-03

# Author Affiliations

1Department of Biostatistics, College of Public Health, University of Kentucky, Lexington, KY

2Sanders-Brown Center on Aging and Alzheimer’s Disease Research Center, University of Kentucky, Lexington, KY

3Vanderbilt Memory & Alzheimer’s Center, Department of Neurology, Vanderbilt University Medical Center, Nashville, Tennessee

4Departments of Neurology and Pathology, Rush University Medical Center, Chicago, IL, 60612, USA.

5Pathology and Laboratory Medicine, University of Kentucky, Lexington, KY

## Corresponding Author

Lincoln Shade

Email: [lincoln.shade@uky.edu](mailto:lincoln.shade@uky.edu)

Phone: 859-893-5798

# Abstract

Brain arteriolosclerosis (B-ASC) is a small-vessel cerebrovascular disease involving sclerotic thickening of arterioles in the brain. B-ASC pathology is common in aged autopsy cohorts, is associated with worse cognitive functioning, and is associated with other neuropathologies. The genetics of other contributors of dementia have been studied extensively, but to date no in-depth study has been conducted on genetic risk of autopsy-proven B-ASC. We perform the first genome-wide association study (GWAS) on B-ASC using multiple cohorts. We then follow-up on identified risk variants with functional analysis to investigate potential biological mechanisms. Individual GWAS and mega-analyses were conducted using data from participants in the National Alzheimer’s Coordinating Center (NACC) and the Religious Orders Study and Memory and Aging Project (ROSMAP). Variants identified in GWAS were then checked for quantitative trait loci (QTL) associations in the Genotype Tissue Expression (GTEx) project. Colocalization analysis was then performed on identified QTL in brain and vascular tissues. Top variants were also examined for association with neuroimaging correlates of B-ASC in the Alzheimer’s Disease Neuroimaging Initiative. One locus on chromosome six with lead variant rs2603462 is significantly associated with B-ASC in NACC and colocalizes with *ELOVL4* gene expression in the brain. Other variants in both cohorts achieved less stringent thresholds of association and colocalize with multiple QTL in GTEx. One locus on chromosome ten near *SORCS3* with lead variant rs7902929 suggestively associated with B-ASC in NACC was validated in ROSMAP. Genetic loci associated with B-ASC pathology were identified using multiple cohorts. Most loci associated with B-ASC in one cohort were not associated in the other. Multiple identified risk loci colocalize with gene expression or splicing QTL, providing evidence for potential biological mechanisms through which variants may influence disease risk. Due to the relatively small sample size and heterogeneity between the cohorts used, future studies should be done to better characterize genetic risk of B-ASC.

# Keywords

Please provide 4 to 6 keywords which can be used for indexing purposes.

# Introduction

Brain arteriolosclerosis (B-ASC) is a subtype of cerebral small vessel neuropathologic change characterized by thickening of arteriole walls in the brain. These changes can include hypertrophy or atrophy of vascular smooth muscle and luminal extracellular deposition of elastin and collagen [[1](#ref-ighodaro2017)–[3](#ref-buchman2013)]. B-ASC is commonly found in autopsied elderly individuals, with 39-80% of participants showing some B-ASC pathology in large autopsy studies [[1](#ref-ighodaro2017), [3](#ref-buchman2013), [4](#ref-chou2013) ]. B-ASC is associated with multiple neuropathologies including Alzheimer’s disease (AD), limbic-predominant age-related TDP-43 encephalopathy (LATE), micro-infarcts, and large-vessel infarcts [[1](#ref-ighodaro2017), [5](#ref-arvanitakis2016), [6](#ref-neltner2014) ]. B-ASC is also associated with cognitive decline after adjusting for the presence of other neuropathologies [[1](#ref-ighodaro2017), [5](#ref-arvanitakis2016)]. Despite the clinical importance of B-ASC, its risk factors, other than age and sex, remain largely uncharacterized. Hypertension and diabetes are established clinical risk factors for arteriolosclerosis in the kidneys, but their association with B-ASC is inconsistent; in an autopsy based cohort stratified by age of death (<80 years vs. ≥80 years), hypertension was significantly associated with B-ASC in the younger group only, and diabetes was not associated with B-ASC in either group [[1](#ref-ighodaro2017), [7](#ref-wu2005)–[9](#ref-cameron2006)].

Genome-wide association studies (GWAS) have provided a powerful resource for investigating genomic risk of complex diseases through analysis of millions of common genetic variants with diseases of interest, and have to date identified tens-of-thousands of variants associated with disease [[10](#ref-buniello2019)]. GWAS have successfully been used to identify genetic risk loci for stroke and for imaging-based phenotypes of cerebrovascular disease such as white matter hyperintensities (WMH) and brain infarcts; however, we found no published GWAS for autopsy-proven B-ASC to date [[11](#ref-traylor2016)–[13](#ref-chauhan2019)]. Given the unique and complex structures of brain arterioles and their associated structures, including astrocytes and other components of the blood-brain barrier, there is reason to suspect that the genomic risk for B-ASC may not be wholly shared with other cerebrovascular phenotypes. In one study of Religious Orders Study/Memory and Aging Project (ROSMAP) participants, 167 independent genetic variants previously meeting genome-wide significance threshold (P < 5×10-8) in GWAS of stroke or stroke risk factorswere tested for association with B-ASC pathology [[4](#ref-chou2013)]. The authors found 6 variants nominally associated with B-ASC at the 0.01 < P-value < 0.05 significance level, none of which remained significant after correction for multiple tests [[4](#ref-chou2013)]. Investigating genomic risk factors of B-ASC at the genome-wide level may provide important insights into its pathophysiological development as well as its relationship to neuroimaging and other neuropathological phenotypes.

In the present study, we conduct a GWAS using B-ASC pathology as the endophenotype analyzing neuropathology and genotype data across four aged autopsy cohorts. In the first stage of our study, we analyzed data from the National Alzheimer’s Coordinating Center (NACC) Neuropathology Dataset linked to genotype data from the Alzheimer’s Disease Genetics Consortium (ADGC; NACC/ADGC when referring to combined datasets). In the second stage, we performed a GWAS using data from ROSMAP participants and then mega-analyze both datasets together. In the third stage, we attempted to replicate our findings froms stages one and two in two smaller autopsy cohorts consisting of Alzheimer’s Disease Neuroimaging Inititative (ADNI) and Adult Changes in Thought (ACT) participants. To investigate potential biologically functional correlates to disease risk, we then perform colocalization analysis on B-ASC associated variants identified as quantitative trait loci (QTL) using data from the Genetic Tissue Expression (GTEx) Project and gene-based association analyses [[14](#ref-thegeno2013)].

# Methods

## Study Participants

In our first study, we linked neuropathology data from participants in thirty-one National Institute on Aging-funded Alzheimer’s Disease Research Centers (ADRCs) studies from the NACC Neuropathology Data Set to ADGC genotype data [[15](#ref-welcome), [16](#ref-besser2018)]. Each ADRC has its own recruitment strategies, populations, and study design, and data are collected and aggregated by NACC. A total of **XXXX** participants had both neuropathology and genotype data available. We then excluded all participants diagnosed with any of nineteen rare neurological conditions (see **Supplementary Table S1** for full exclusion criteria).

The ROSMAP study has been previously described in detail and consists of harmonized data from two longitudinal cohorts: The Religious Orders Study (ROS) and the Memory and Aging Project (MAP). ROS began in 1994 and recruited older Catholic Priests and Sisters from around the United States. MAP began in 1997 and recruited older adults who at the time had no diagnosis of dementia. A total of **XXXX** ROSMAP participants had both autopsy and genotyping data available.

The ADNI study…

The ACT study…

## **Definitions of B-ASC variables used**

In the NACC Neuropathology Data Set, B-ASC is graded as an ordinal variable with possible values of 0 (none), 1 (mild), 2 (moderate), or 3 (severe). Grading is performed by trained Neuropathologists at each ADRC and is a global rating, meaning no instruction is given to examine specific brain regions for B-ASC. Both the ADNI and ACT studies follow the NACC Neuropathology Form when grading neuropathological phenotypes. The B-ASC variable in ROSMAP is graded on B-ASC histological changes exclusively in the basal ganglia. Vessel intimal pathology is first graded on a scale from 0 (none) to 6 (severe), and then collapsed to a four-level ordinal variable with the same labels as in NACC: 0 (none), 1 (mild), 2 (moderate), and 3 (severe) [[17](#ref-buchman2011)].

## **Quality control (QC) of genotype data**

The data in our study were all subset from datasets whose QC measures have been previously described . We performed additional standard QC procedures on all genotyping data using PLINK v1.9 and KING [[18](#ref-chang2015)–[22](#ref-marees2018)]. Variants were excluded if they had (1) a minor allele frequency (MAF) less than 5%; (2) a call rate of less than 95%; or (3) a Hardy-Weinberg equilibrium exact test P-value < 10-6. All participants were checked for duplicate genotype information across studies, and all participants with duplicated genotyping data were removed. Of participants with a high degree of relatedness estimated using identical by descent (IBD) indicated by 2nd-degree relation (proportion IBD > 0.18) or closer in PLINK, all but the participant with the highest genotyping rate in each related cluster were removed, with ties broken randomly. In NACC/ADGC analyses, 4799488 variants and 3382 participants passed QC protocols. In the ROSMAP dataset, 4889494 variants and 1187 participants passed QC protocols. Thirteen ADNI participants were identified as NACC/ADGC duplicates and removed, leaving 47 participants for analysis. A total of **XXX (~326)** ACT participants passed QC measures.

## **Identifying ethnic outliers**

In all cohorts, we performed principal component analysis (PCA) in PLINK v1.9 using a pruned subset of independent (linkage disequilibrium (LD) r2 < 0.05) variants from each dataset merged to data from the 1000 Genomes Project Phase 3 (1000 Genomes, n = 2504) [[23](#ref-Abecasis2012)]. All variants were checked to ensure that major and minor alleles matched in the study datasets and 1000 Genomes. We then plotted the standardized first and second principal components (PCs) for each participant using the ggplot2 R package in R version 4.0.3 [[24](#ref-rcoreteam2020)]. All participants whose plot positions were located within a Euclidean distance of 0.3 from the mean plot positions of 1000 Genomes participants in the EUR superpopulation were considered to be of European descent and were included in analyses. We then re-ran PCA for the included participants and used included the first five PCs to use as covariates in regression models.

## **Statistical analyses**

### **Single-variant analyses**

We first performed single-variant association analyses in the NACC/ADGC participants using logistic regression in PLINK v1.9 and ordinal regression in R using the MASS package [[19](#ref-chang), [24](#ref-rcoreteam2020)]. To create a dichotomous outcome variable for logistic regression, we counted participants with none or mild B-ASC as controls and those with moderate or severe B-ASC as cases. This decision was made based on previous studies that used these cut points and studies that found that moderate-to-severe B-ASC was associated with worse cognitive functioning [[1](#ref-ighodaro2017), [5](#ref-arvanitakis2016)]. Covariates in the regression models included age at death, sex, ADGC cohort indicators, and the first five PCs of the genetic relatedness matrix. An additive mode of inheritance was assumed in all analyses. We used a Bonferroni-corrected threshold of P-value < 5×10-8 for genome-wide significance and a predetermined “suggestive” threshold of P-value < 1×10-5. Then, using PLINK v1.9, we clumped variants meeting this threshold to create a set of independent variants (defined as LD r2 < 0.05). For suggestive variants in ordinal regression analyses, we tested the proportional odds assumption using Brant tests in the brant R package [[25](#ref-schlegel2020)]. Finally, we examined single variants previously found to be putatively associated with B-ASC in Chou et al. (2013) to determine if any were validated in NACC/ADGC at the P-value < 0.05 significance level [[4](#ref-chou2013)].

In our ROSMAP analyses, we first performed single-variant regression analyses with variants identified as suggestively significant in the NACC/ADGC analyses, using a significance threshold of P-value < 0.05 for these variants. We then performed ordinal and dichotomous genome-wide single-variant the procedures used in the NACC/ADGC analyses. Covariates included age at death, sex, cohort (ROS vs. MAP) and the first five PCs. We then performed a mega-analysis on both cohorts, including an indicator variable for ROSMAP participants. We also performed both fixed- and random-effects meta-analyses in PLINK v1.9 for individual ADGC and ROSMAP logistic GWAS results.

To seek replication of our results, we analyzed the top three variants from the NACC/ADGC and ROSMAP mega-analysis separately in the ADNI and ACT datasets. We then added the ADNI and ACT participants sequentially to our combined NACC/ADGC and ROSMAP dataset to perform additional mega-analyses.

### **Gene-based analyses**

# Placeholder

Following single-variant analyses, we performed gene-based and gene-set analyses using MAGMA [[26](#ref-leeuw2015)]. First, we annoted variants to genes using Hg37 positions, and annotated all variants within a 1000-kilobase window of a gene as associated with it, producing a total of **XXXXX** genes with at least one annotated variant. We then performed competitive gene-based analyses in MAGMA. For the gene-set analysis, we used gene sets curated for Enrichment Map ([https://enrichmentmap.readthedocs.io](https://enrichmentmap.readthedocs.io/en/latest/index.html#)) that included gene sets based on multiple resources, including Reactome, Gene Ontology, the Pathway Interaction Database, NetPath, and the PANTHER database [[[27](#ref-merico2010)]; ].

### **Colocalization analyses**

To investigate potential functional mechanisms of variants driving GWAS signals for B-ASC, we performed colocalization analyses for suggestive variants using the coloc R package and QTL summary statistic data from The Genotype-Tissue Expression (GTEx) project V8 publicly available data [[14](#ref-thegeno2013), [28](#ref-aguet2017)–[30](#ref-gtexpor)]. QTL are calculated in GTEx by performing single-variant analysis of gene expression for all variants within 1000 Kb of the transcription start or end site for each gene. First, we systematically checked to determine if suggestive variants from B-ASC analyses (P-value < 1×10-5) are significant expression QTL (eQTL) or splicing QTL (sQTL) using GTEx summary statistics of participants of European descent. We then analyzed variants that were eQTL or sQTL in brain, nerve, blood, cultured cell lines, or vascular tissues (see Supplementary Table S2 for a full description of tissues used) and had nominal P-values less than ten times greater than the variant with the smallest nominal P-value for the associated QTL phenotype. We used default prior probabilities in the coloc package of P1 = P2 = 1e-4 and P12 = 1e-5, meaning that each variant has a 1/10,000 prior probability of being associated with either trait, and a variant associated with either trait in turn has a 1/10 prior probability of being associated with both traits [[29](#ref-giambartolomei2014)]. Because the coloc package can currently only analyze dichotomous or continuous variables, for QTL variants from ordinal B-ASC GWAS, we used variant P-values in the corresponding logistic analyses. A posterior probability of colocalization (PPH4) of 50% or greater was chosen to indicate evidence for colocalization.

### **Sensitivity analyses**

Given previously identified potential differences in clinical risk factors for B-ASC in participants stratified by age of death, we re-did our analyses in NACC/ADGC and ROSMAP using only participants with an age of death of eighty or above. To further assess the robustness of our results under different model assumptions, we performed several sensitivity analyses in our NACC/ADGC dataset using the binary B-ASC outcome variable. We first performed single-variant analyses on each ADGC cohort and then meta-analyzed in PLINK v1.9 rather than using fixed-effect cohort indicators in our regression models. For variants meeting our suggestive threshold, we included related participants and performed mixed-effects analyses with a random effect incorporating the kinship matrix estimated from KING in R using the GMMAT, GENESIS, and SNPRelate packages [[31](#ref-chen2020)–[34](#ref-zheng2012)]. To overcome issues with computing PCs with samples with related participants, we used the PC-AiR method [[33](#ref-gogarten)]. Finally we performed mediation analyses on the subset of our NACC/ADGC with clinical variables available to test whether diabetes mellitus or hypertension status mediate the association between suggestive genetic variants and B-ASC pathology [[1](#ref-ighodaro2017), [35](#ref-baron)].

### **Testing Variants of Interest with WMH Volume**

We tested the top two variants from the single-variant meta-analysis for association with WMH volume in Alzheimer’s Disease Neuroimaging Initiative (ADNI). ***Insert section here. Either Kwangsik or someone else at IU should write this section or Lincoln should discuss with Kwangsik on what methods were used for analysis.***

# Results

Of the 3382 NACC participants that met inclusion criteria for Sanalysis, 935 (27.6%) had no B-ASC, 1023 (30.2%) had mild B-ASC, 1043 (30.8%) had moderate B-ASC, and 381 (11.3%) had severe B-ASC (see **Table 1**). ROSMAP participants that met inclusion for analysis had comparatively less B-ASC pathology (P-value = 0.002): 414 (34.9%) had no B-ASC, 405 (34.1%) had mild B-ASC, 284 (23.9%) had moderate B-ASC, and 84 (7.1%) had severe B-ASC. NACC participants were also significantly more likely to be male (nacc\_adgc\_unrelated[NPSEX == 1, round(50% vs.33%, P-value < 0.001) and had younger ages at death on average (mean age of death 82 vs. 89.6, P-value < 0.001) compared to ROSMAP participants. These comparative differences in the cohort demographics largely held in the secondary analysis that included only participants whose ages of death were eighty years or above, with the caveat that the difference in mean age of death was smaller between NACC and ROSMAP (87.7 vs. 90.7, P-value < 0.001).

| Variable | Labels | Overall | ADNI | NACC/ADGC | ROSMAP |
| --- | --- | --- | --- | --- | --- |
| B-ASC |  |  |  |  |  |
|  | None | 1350 (29.2%) | 1 (2.1%) | 935 (27.6%) | 414 (34.9%) |
|  | Mild | 1459 (31.6%) | 31 (66%) | 1023 (30.2%) | 405 (34.1%) |
|  | Moderate | 1340 (29%) | 13 (27.7%) | 1043 (30.8%) | 284 (23.9%) |
|  | Severe | 467 (10.1%) | 2 (4.3%) | 381 (11.3%) | 84 (7.1%) |
| Sex |  |  |  |  |  |
|  | Female | 2488 (53.9%) | 8 (17%) | 1680 (49.7%) | 800 (67.4%) |
|  | Male | 2128 (46.1%) | 39 (83%) | 1702 (50.3%) | 387 (32.6%) |
| Age of Death |  |  |  |  |  |
|  | Mean (SD) | 84 (9.3) | 83.3 (7.1) | 82 (9.4) | 89.6 (6.5) |
|  | Median [Min, Max] | 85 [47, 111] | 84 [59, 97] | 83 [47, 111] | 90 [66, 108.3] |

## Single-variant analyses

In the logistic regression analysis, 1424 (42.1%) and 368 (31%) of participants had either moderate or severe B-ASC and were counted as cases in NACC and ROSMAP, respectively. In the NACC/ADGC analysis, one locus, rs2603462 on chromosome 6q14.1, is genome-wide significantly associated with B-ASC (odds ratio [OR] = 1.45, P-value = 2.5^{-8}). We identified 13 other loci that meet our suggestive association threshold of P-value < 1×10-5, described in **Table 2**. No variants achieve genome-wide significance in any of the ROSMAP analyses, though **X** loci met our suggestive threshold in the primary logistic regression analysis. Of the 14 NACC/ADGC loci that met our suggestive threshold that were also tested in the ROSMAP analysis, rs7902929 on chromosome 10q25.1 was validated at P-value < 0.05 level (NACC/ADGC OR = 1.57, P-value = 7.8^{-6}; ROSMAP OR = 1.61, P-value = 0.0069). For the other 13 loci, two ROSMAP ORs are <1.00 and 11 have effect sizes in the same direction in effect as in NACC/ADGC, though of smaller magnitude. Results in the ordinal regressions were broadly similar to the logistic analyses in both cohorts. Of the top 14 loci in the NACC primary logistic regression, **X** also meet our suggestive threshold in the ordinal regression, along with **X** other loci. In the primary mega-analysis using logistic regression, **X** loci meet the suggestive threshold, with rs7902929 (OR = 1.58, P-value = 2.0 x 10-7) and rs2603462 (OR = 1.35, P-value = 3.0 x 10-7) producing the smallest P-values (see **Table 3**). Of the six variants found to be nominally associated with B-ASC by Chou et al. (2013), no variants were validated at the P-value < 0.05 significance level in the NACC/ADGC cohort.

| SNP | Gene | CHR | BP | A1/A2 | NACC OR [95% CI] | NACC P | ROSMAP OR [95% CI] | ROSMAP P |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| rs2603462 | TBD | 6 | 81,418,667 | C/A | 1.45 [1.27-1.66] | 2.5e-08 | 1.05 [0.84-1.33] | 0.65 |
| rs3774902 | TBD | 4 | 23,890,782 | A/G | 1.77 [1.42-2.21] | 4.5e-07 |  | NA |
| rs2418491 | TBD | 2 | 59,535,434 | G/A | 1.33 [1.19-1.49] | 4.8e-07 | 1.02 [0.84-1.23] | 0.84 |
| rs115980554 | TBD | 6 | 34,456,632 | C/T | 1.63 [1.33-2] | 3.4e-06 | 0.95 [0.65-1.37] | 0.77 |
| rs61944465 | TBD | 13 | 27,272,704 | G/A | 1.33 [1.18-1.5] | 3.6e-06 | 1.16 [0.93-1.46] | 0.19 |
| rs4470773 | TBD | 5 | 177,507,080 | G/C | 1.28 [1.15-1.42] | 3.9e-06 | 1.07 [0.89-1.29] | 0.47 |
| rs6574718 | TBD | 14 | 26,395,832 | C/T | 1.28 [1.15-1.42] | 5.3e-06 |  | NA |
| rs387083 | TBD | 19 | 35,847,115 | G/A | 1.28 [1.15-1.42] | 5.4e-06 | 1.02 [0.85-1.22] | 0.86 |
| rs28362345 | TBD | 6 | 31,165,836 | T/C | 1.27 [1.15-1.41] | 6.3e-06 | 0.94 [0.78-1.14] | 0.53 |
| rs2069126 | TBD | 20 | 19,749,957 | A/G | 1.27 [1.14-1.4] | 6.7e-06 | 1.11 [0.92-1.34] | 0.27 |
| rs7902929 | TBD | 10 | 107,237,532 | C/T | 1.57 [1.29-1.92] | 7.8e-06 | 1.61 [1.14-2.26] | 0.0069 |
| rs12700439 | TBD | 7 | 23,472,061 | T/C | 1.26 [1.14-1.4] | 8.0e-06 | 1.07 [0.89-1.29] | 0.47 |
| rs10790707 | TBD | 11 | 124,460,485 | A/G | 1.26 [1.14-1.4] | 8.1e-06 | 1.03 [0.85-1.24] | 0.76 |
| rs9895518 | TBD | 17 | 49,645,803 | A/G | 1.28 [1.15-1.42] | 8.3e-06 | 1.09 [0.9-1.32] | 0.39 |

In the ADNI neuropathology cohort, both rs2603462 (OR = **X**, P = **X**) and rs7902929 (OR = **X**, P =**X** ) were valiedated at the *P* < 0.05 significance level (see **Table 4**), while in the ACT neuropathology cohort, neither variant was validated.

In the secondary analysis using only participants with an age of death of 80 years or older, no variants achieved genome-wide significance (P-value < 5x10-8). In the NACC/ADGC analyses, we identified 10 and 9 independent loci in the logistic and ordinal regression analyses, respectively, that met the suggestive threshold. Three loci were suggestively associated with B-ASC in both the logistic and ordinal analyses: rs4491854, an intergenic SNP near EOMES on chromosome 3 (logistic P-value = 8.0x10-7, ordinal P-value = 6.2x10-6); rs11928305, an intronic SNP in FAM19A1/TAFA1 on chromosome 3 (logistic P-value = 6.1x10-6, ordinal P-value = 4.0x10-7); and rs1956605, and intergenic SNP 500 Kb from NOVA1 on chromosome 14 (logistic P-value = 4.9x10-6, ordinal P-value = 6.3x10-6). No suggestive variants from the NACC/ADGC analyses were validated in the ROSMAP analyses at the P-value < 0.05 significance threshold (see supplementary tables S3 and S4).

## Gene-based analyses

In the gene-based analyses, no genes achieved Bonferroni-adjusted significance (P-value < 2.5x10-6).

## Colocalization analyses

In the mega-analyses, five variants met our criteria for colocalization analysis: rs2603462 colocalizes with B-ASC and ELOVL4 expression in the cerebellar hemisphere with PPH4 of 92.6%; rs1343705 colocalizes with expression of the non-coding gene RP11-408A13.3 in the cerebellum with a PPH4 of 87.2%; rs35010424 colocalizes with TAS2R5 expression in the hypothalamus with PPH4 of 84.4% and OR9N1P expression in the nucleus accumbens with PPH4 of 57.7%; rs34349961 colocalizes with SPRED2 expression in cultured fibroblasts with PPH4 of 74.6%; and rs6936285 colocalizes with CD83 in whole blood with PPH4 of 73.7% (see Table 4). In the NACC/ADGC analyses, rs2603462 and 3 suggestive loci are significant QTLs in GTEx and met our criteria for colocalization analysis across eleven phenotypes in nine tissues. Three of these loci colocalized with a PPH4 > 50% across ten phenotypes. rs2603462 colocalizes with B-ASC and ELOVL4 expression in the cerebellar hemisphere with a PPH4 of 93.3%. rs2352974 is located in a gene-dense region of chromosome 3p21.31 and is a significant eQTL for DALRD3, FAM212A, MST1R, and TCTA in one tissue each, and a significant sQTL for RNF123 in four tissues; it colocalizes with B-ASC and seven of these phenotypes with PPH4 between 50.0% and 79.1%. No variants in the ROSMAP primary analyses met our criteria for colocalization analysis.

In the secondary NACC/ADGC analyses using only the oldest participants, eQTLs in GTEx were found for two genes, HLA-A and NMRAL2P. The HLA-A variant, rs9260090, is a significant eQTL in nine brain tissues, spanning the frontal cortex, basal ganglia, cingulate cortex, and cerebellum. Colocalization analysis in seven of these nine tissues showed evidence for colocalization with a posterior probability of colocalization (PPH4) greater than 50% and the brain cortex having the highest probability of colocalization (PPH4 = 82.1%). Evidence for colocalization is lowest in the hippocampus (PPH4 = 9.4%). The NMRAL2P variant, rs11718099, is an eQTL in the aortic artery, and colocalization analysis revealed a PPH4 of 86.2%. However, because rs11718099 was identified as being suggestively associated with B-ASC in the NACC/ADGC ordinal regression, its P-value from the logistic regression used for colocalization analysis is much larger (1.4 x 10-3 vs 8.7 x 10-6).

## Sensitivity analyses

Performing fixed-effects meta-analysis on the NACC/ADGC ADGC genotyping cohorts produced nearly identical effect sizes and P-values compared to including indicator variables for the non-reference cohorts in our primary analysis. Similarly, including the participants originally excluded due to relatedness did not produce notably different effect sizes or P-values for any of the suggestive variants. Finally, in the subset of our NACC/ADGC participants who had clinical data for diabetes and hypertension status (n = 1679), none of the suggestive variants were associated with diabetes or hypertension status at the P-value < 0.05 level, and all variant effect sizes on B-ASC in regressions including diabetes or hypertension were within the 95% confidence intervals in the base model excluding them.

## WMH Volume Analyses

Insert section here.

# Conclusions

In this study, we performed the first GWAS of autopsy-proven B-ASC using neuropathology and genotype data from four neuropathology cohorts derived from multiple research centers. We found a significant association between one locus on Chromosome 6 and B-ASC in the NACC/ADGC cohort (*P* = 2.5^{-8}). This locus colocalizes with *ELOVL4* gene expression in GTEx (PPH4 = 93.3%), providing evidence that this locus may affect B-ASC risk through mediating *ELOVL4* expression (see **Figure X**). Another locus on Chromosome 10 near *SORCS3* suggestively associated with B-ASC was validated in the ROSMAP cohort. These variants’ effects were successfully replicated in the ADNI cohort, but failed to replicate in the ACT cohort. We also found suggestive evidence for association between other loci and B-ASC risk in both cohorts and that some of these loci colocalize with gene eQTL and sQTL in GTEx.

Most genetic loci identified as being associated with B-ASC in NACC were not validated in ROSMAP. This could be potentially explained via the significantly different demographic attributes and study designs between the two cohorts, as ROSMAP participants were simultaneously older at death and had lower risk of B-ASC pathology, indicating that there may be unaccounted-for genetic confounding and differential selection bias between cohorts. For instance, ROS recruits from Catholic sisters and brothers, who are on average more highly educated than the general population. Each ADRC also has its own recruitment population. Differences in neuropathological grading of B-ASC in each cohort may also contribute to the heterogeneity of results, as B-ASC is graded globally in NACC but is graded in the basal ganglia in ROSMAP. Despite these limitations, one suggestive locus on chromosome ten identified in NACC/ADGC was validated in ROSMAP and had nearly the same affect size in each cohort (OR 1.58 vs. 1.60). Furthermore, 11 of 13 B-ASC-associated loci in NACC had affect sizes in the same direction in ROSMAP, which suggests that these loci may be associated with B-ASC but suffer from the “winner’s curse” in the NACC/ADGC analyses.

*ELOVL4* codes for the elongation of very long chain fatty acids-4 protein, an elongase that synthesizes very long chain saturated and unsaturated fatty acids. Most research on *ELOVL4*has focused on its association with mendelian diseases affecting the visual and nervous systems, such as Stargardt-like macular dystrophy and spinocerebellar ataxia [[36](#ref-hopiavuori2019)]. In GTEx, ELOVL4 is more highly expressed in the brain (median transcripts per kilobase million [TPM] 5.5-43.9) relative to most other tissues. We searched the GWAS Catalog and found two recent studies that found genome-wide significant associations between variants mapped to *ELOVL4* and multiple body weight-related phenotypes, including waist circumference adjusted for body mass index (BMI) (rs76567515, P = 2 x 10-12), waist-to-hip ratio (rs76567515, P = 3 x 10-10), and waist-to-hip ratio adjusted for BMI (rs1849275, P = 1 x 10-9) [[10](#ref-buniello2019), [37](#ref-zhu2020), [38](#ref-kichaev2019)]. These variants are significant eQTLs for ELOVL4 in GTEx, but are each located >200 Kb from the locus identified in our study and are not in LD with the lead variant (r2 < 0.05). Additionally, another suggestive variant in NACC/ADGC, rs387083, is located between the free fatty acid receptor genes *FFAR1* and *FFAR3* on chromosome nineteen. These results suggest that fatty acid metabolism and signaling may play a role in B-ASC risk, and that ELOVL4 may potentially be related to B-ASC risk through its effects on BMI-related phenotypes.

The variant suggestively associated with B-ASC in the NACC/ADGC primary analyses that was subsequently validated in ROSMAP, rs7902929, is located approximately 212 Kb from the 3’ end of the gene SORCS3. While the genes on which regulator intergenic variants exert their effects are not necessarily the closest genes, SORCS3 is the only protein-coding gene within a 1 megabase window from rs7902929, increasing the likelihood that the locus is functionally tied to it. SORCS3 codes for the sortilin-related VPS10 domain-containing receptor 3, a vacuolar protein expressed in the brain [[39](#ref-reitz2013), [40](#ref-wang2020)]. Previous studies using candidate gene designs have provided tentative evidence that genetic variation in SORCS3 may be associated with AD [[39](#ref-reitz2013), [40](#ref-wang2020)]. In the GWAS Catalog, SORCS3-mapped variants are significantly associated with multiple phenotypes, including depressive symptoms (rs1021363, P = 1 x 10-13), self-reported educational attainment (rs11599236, P = 1 x 10-13), and systolic blood pressure (rs191784289, P = 3 x 10-13) [[38](#ref-kichaev2019), [41](#ref-baselmans2019), [42](#ref-lee2018)].

Variants identified as suggestively associated with B-ASC among participants aged 80 years or older at death were largely different than those identified in our primary analyses. One intronic locus of FAM19A1 was identified as suggestively associated with B-ASC in the NACC/ADGC dataset. This gene is highly and preferentially expressed in the brain [[14](#ref-thegeno2013)]. FAM19A1 belongs to the family with sequence similarity 19 member A gene family, a five-member gene family with largely unknown function, though the family shares structural similarity to the CC-chemokine family [[43](#ref-wangyingbao2018)]. A locus near the 3’ end of another member of this family, FAM19A5, was identified as suggestively associated with B-ASC in the NACC/ADGC primary ordinal analysis. A recent study found that FAM19A5 is an adipokine that is capable of inhibiting post-injury neointima formation in injured blood vessels, and that FAM19A5 expression is downregulated in obesity.49 Another study found that FAM19A3, another gene homologue of FAM19A1, is highly expressed in microglia and helps attenuate cerebral ischemia [[44](#ref-shao2015)].

Two loci suggestively associated with B-ASC in the analyses on older participants colocalized with gene expression with HLA-A and NRMAL2P in GTEx. HLA-A is a Major Histone Compatibility Complex (MHC) gene and is an integral component of the immune system. HLA-A has been found to be significantly associated with 120 complex traits in the GWAS Atlas [[45](#ref-watanabe2019)]. While not identified in GWAS, certain HLA-A alleles have evidence of association with Alzheimer’s disease (AD) risk [[46](#ref-cifuentes2014)]. In one recent study, authors used ADNI data to investigate the association between AD HLA-A found that HLA-A variants were associated with atrophy of the left parahippocampus, right hippocampus, and right amygdala [[47](#ref-wang2017)]. NRMAL2P is a long non-coding RNA (LncRNA) pseudogene that has previously been linked to gastric, colon, and esophageal cancers [[[48](#ref-feng2020)]; [[49](#ref-johnson2017)]; [[50](#ref-mizumoto2019)]; ].

None of the effects of top loci in the NACC/ADGC GWAS were mediated by diabetes or hypertension status in NACC. This finding fails to help clarify the somewhat inconsistent evidence of association between these clinical risk factors and B-ASC pathology. However, this analysis was limited by a substantially smaller sample size (n = 1679) than our other analysis and the diagnostic criteria used, as the diagnostic variables in NACC used consisted of a mixture of patient self-reports, physician reporting, and use of hypertension or diabetes medications. Future studies employing causal inference methods such as two-sample Mendelian randomization may be able to provide clearer evidence for or against the genetic mediation of B-ASC risk by clinical risk factors.

In closing, we used autopsy-derived endophenotypes linked to genetic data in multiple cohorts to identify genetic loci associated with B-ASC pathology. We found one locus significantly associated with B-ASC that colocalizes with ELOVL4 gene expression in GTEx. This study provides first GWAS of autopsy-verified brain arteriolosclerosis pathology and contributes a growing body of literature that recognizes the importance of using autopsy-based cohorts to augment clinical diagnoses to study the genetics of dementia and cognitive impairment.

# Code Stash

# References

1. Ighodaro ET, Abner EL, Fardo DW, et al (2017) Risk factors and global cognitive status related to brain arteriolosclerosis in elderly individuals. Journal of Cerebral Blood Flow & Metabolism 37:201–216. <https://doi.org/10.1177/0271678X15621574>

2. Grinberg LT, Thal DR (2010) Vascular pathology in the aged human brain. Acta Neuropathologica 119:277–290. <https://doi.org/10.1007/s00401-010-0652-7>

3. Buchman AS, Yu L, Boyle PA, et al (2013) Microvascular brain pathology and late-life motor impairment. Neurology 80:712–718. <https://doi.org/10.1212/WNL.0b013e3182825116>

4. Chou S-Y, Shulman JM, Keenan BT, et al (2013) Genetic susceptibility for ischemic infarction and arteriolosclerosis based on neuropathologic evaluations. Cerebrovascular Diseases (Basel, Switzerland) 36:181–188. <https://doi.org/10.1159/000352054>

5. Arvanitakis Z, Capuano AW, Leurgans SE, et al (2016) Relation of cerebral vessel disease to alzheimer’s disease dementia and cognitive function in older persons: A cross-sectional study. The Lancet Neurology 15:934–943. <https://doi.org/10.1016/S1474-4422(16)30029-1>

6. Neltner JH, Abner EL, Baker S, et al (2014) Arteriolosclerosis that affects multiple brain regions is linked to hippocampal sclerosis of ageing. Brain 137:255–267. <https://doi.org/10.1093/brain/awt318>

7. Wu J, Chen X, Xie Y, et al (2005) Characteristics and risk factors of intrarenal arterial lesions in patients with IgA nephropathy. Nephrology Dialysis Transplantation 20:719–727. <https://doi.org/10.1093/ndt/gfh716>

8. Ikee R, Honda K, Ishioka K, et al (2010) Postprandial hyperglycemia and hyperinsulinemia associated with renal arterio-arteriolosclerosis in chronic kidney disease. Hypertension Research: Official Journal of the Japanese Society of Hypertension 33:499–504. <https://doi.org/10.1038/hr.2010.22>

9. Cameron JS (2006) The discovery of diabetic nephropathy: from small print to centre stage. Journal of Nephrology 19 Suppl 10:S75–87

10. Buniello A, MacArthur JAL, Cerezo M, et al (2019) The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Research 47:D1005–D1012. <https://doi.org/10.1093/nar/gky1120>

11. Traylor M, Zhang CR, Adib-Samii P, et al (2016) Genome-wide meta-analysis of cerebral white matter hyperintensities in patients with stroke. Neurology 86:146–153. <https://doi.org/10.1212/WNL.0000000000002263>

12. Malik R, Chauhan G, Traylor M, et al (2018) Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. Nature Genetics 50:524–537. <https://doi.org/10.1038/s41588-018-0058-3>

13. Chauhan G, Adams HHH, Satizabal CL, et al (2019) Genetic and lifestyle risk factors for MRI-defined brain infarcts in a population-based setting. Neurology 92:e486–e503. <https://doi.org/10.1212/WNL.0000000000006851>

14. (2013) The genotype-tissue expression (GTEx) project. Nature genetics 45:580–585. <https://doi.org/10.1038/ng.2653>

15. Welcome to alzheimer’s disease genetics consortium | ADGC

16. Besser LM, Kukull WA, Teylan MA, et al (2018) The Revised National Alzheimer’s Coordinating Center’s Neuropathology Form-Available Data and New Analyses. Journal of Neuropathology and Experimental Neurology 77:717–726. <https://doi.org/10.1093/jnen/nly049>

17. Buchman AS, Leurgans SE, Nag S, et al (2011) Cerebrovascular disease pathology and parkinsonian signs in old age. Stroke; a journal of cerebral circulation 42:3183–3189. <https://doi.org/10.1161/STROKEAHA.111.623462>

18. Chang CC, Chow CC, Tellier LC, et al (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience 4: <https://doi.org/10.1186/s13742-015-0047-8>

19. Chang CC, Purcell S Plink 1.9

20. Purcell S, Neale B, Todd-Brown K, et al (2007) PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. The American Journal of Human Genetics 81:559–575. <https://doi.org/10.1086/519795>

21. Manichaikul A, Mychaleckyj JC, Rich SS, et al (2010) Robust relationship inference in genome-wide association studies. Bioinformatics 26:2867–2873. <https://doi.org/10.1093/bioinformatics/btq559>

22. Marees AT, de Kluiver H, Stringer S, et al (2018) A tutorial on conducting genome-wide association studies: Quality control and statistical analysis. International Journal of Methods in Psychiatric Research 27:e1608. <https://doi.org/10.1002/mpr.1608>

23. (2012) An integrated map of genetic variation from 1,092 human genomes. Nature 491:56–65. <https://doi.org/10.1038/nature11632>

24. R Core Team (2020) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria

25. Schlegel B, Steenbergen M (2020) Brant: Test for parallel regression assumption

26. Leeuw CA de, Mooij JM, Heskes T, Posthuma D (2015) MAGMA: Generalized Gene-Set Analysis of GWAS Data. PLOS Computational Biology 11:e1004219. <https://doi.org/10.1371/journal.pcbi.1004219>

27. Merico D, Isserlin R, Stueker O, et al (2010) Enrichment Map: A Network-Based Method for Gene-Set Enrichment Visualization and Interpretation. PLoS ONE 5:e13984. <https://doi.org/10.1371/journal.pone.0013984>

28. Aguet F, Brown AA, Castel SE, et al (2017) Genetic effects on gene expression across human tissues. Nature 550:204–213. <https://doi.org/10.1038/nature24277>

29. Giambartolomei C, Vukcevic D, Schadt EE, et al (2014) Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. PLOS Genetics 10:e1004383. <https://doi.org/10.1371/journal.pgen.1004383>

30. GTEx portal

31. Chen H, Conomos MP (2020) GMMAT: Generalized linear mixed model association tests

32. Conomos MP, Gogarten SM, Brown L, et al (2019) GENESIS: GENetic EStimation and inference in structured samples (GENESIS): Statistical methods for analyzing genetic data from samples with population structure and/or relatedness

33. Gogarten SM, Sofer T, Chen H, et al Genetic association testing using the GENESIS R/Bioconductor package. Bioinformatics. <https://doi.org/10.1093/bioinformatics/btz567>

34. Zheng X, Levine D, Shen J, et al (2012) A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics 28:3326–3328. <https://doi.org/10.1093/bioinformatics/bts606>

35. Baron RM, Kenny DA The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations. 10

36. Hopiavuori BR, Anderson RE, Agbaga M-P (2019) ELOVL4: Very long-chain fatty acids serve an eclectic role in mammalian health and function. Progress in Retinal and Eye Research 69:137–158. <https://doi.org/10.1016/j.preteyeres.2018.10.004>

37. Zhu Z, Guo Y, Shi H, et al (2020) Shared genetic and experimental links between obesity-related traits and asthma subtypes in UK Biobank. The Journal of Allergy and Clinical Immunology 145:537–549. <https://doi.org/10.1016/j.jaci.2019.09.035>

38. Kichaev G, Bhatia G, Loh P-R, et al (2019) Leveraging Polygenic Functional Enrichment to Improve GWAS Power. American Journal of Human Genetics 104:65–75. <https://doi.org/10.1016/j.ajhg.2018.11.008>

39. Reitz C, Tosto G, Vardarajan B, et al (2013) Independent and epistatic effects of variants in VPS10-d receptors on alzheimer disease risk and processing of the amyloid precursor protein (APP). Translational Psychiatry 3:e256. <https://doi.org/10.1038/tp.2013.13>

40. Wang M, Wang S, Li Y, et al (2020) Integrated analysis and network pharmacology approaches to explore key genes of xingnaojing for treatment of alzheimer’s disease. Brain and Behavior 10: <https://doi.org/10.1002/brb3.1610>

41. Baselmans BML, Jansen R, Ip HF, et al (2019) Multivariate genome-wide analyses of the well-being spectrum. Nature Genetics 51:445–451. <https://doi.org/10.1038/s41588-018-0320-8>

42. Lee JJ, Wedow R, Okbay A, et al (2018) Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. Nature Genetics 50:1112–1121. <https://doi.org/10.1038/s41588-018-0147-3>

43. Wang Yingbao, Chen Dixin, Zhang Yan, et al (2018) Novel adipokine, FAM19A5, inhibits neointima formation after injury through sphingosine-1-phosphate receptor 2. Circulation 138:48–63. <https://doi.org/10.1161/CIRCULATIONAHA.117.032398>

44. Shao Y, Deng T, Zhang T, et al (2015) FAM19A3, a novel secreted protein, modulates the microglia/macrophage polarization dynamics and ameliorates cerebral ischemia. FEBS Letters 589:467–475. <https://doi.org/10.1016/j.febslet.2015.01.003>

45. Watanabe K, Stringer S, Frei O, et al (2019) A global overview of pleiotropy and genetic architecture in complex traits. Nature Genetics 51:1339–1348. <https://doi.org/10.1038/s41588-019-0481-0>

46. Cifuentes RA, Murillo-Rojas J (2014) Alzheimer’s Disease and HLA-A2: Linking Neurodegenerative to Immune Processes through an In Silico Approach

47. Wang Z-X, Wang H-F, Tan L, et al (2017) HLA-A2 Alleles Mediate Alzheimer’s Disease by Altering Hippocampal Volume. Molecular Neurobiology 54:2469–2476. <https://doi.org/10.1007/s12035-016-9832-3>

48. Feng H, Liu X (2020) Interaction between ACOT7 and LncRNA NMRAL2P via Methylation Regulates Gastric Cancer Progression. Yonsei Medical Journal 61:471–481. <https://doi.org/10.3349/ymj.2020.61.6.471>

49. Johnson GS, Li J, Beaver LM, et al (2017) A functional pseudogene, NMRAL2P, is regulated by Nrf2 and serves as a coactivator of NQO1 in sulforaphane-treated colon cancer cells. Molecular Nutrition & Food Research 61: <https://doi.org/10.1002/mnfr.201600769>

50. Mizumoto A, Ohashi S, Kamada M, et al (2019) Combination treatment with highly bioavailable curcumin and NQO1 inhibitor exhibits potent antitumor effects on esophageal squamous cell carcinoma. Journal of Gastroenterology 54:687–698. <https://doi.org/10.1007/s00535-019-01549-x>