Genome-wide association study of brain arteriolosclerosis

Lincoln MP Shade[1] [iD]**, Yuriko Katsumata[1,2],**
Timothy J Hohman[3], Kwangsik Nho[4], Andrew J Saykin[4] [iD]**,**
Shubhabrata Mukherjee[5], Kevin L Boehme[6], John SK Kauwe[7],
Lindsay A Farrer[8], Gerard D Schellenberg[9],
Jonathan L Haines[10], Richard P Mayeux[11], Julie A Schneider[12],
Peter T Nelson[2,13] and David W Fardo[1,2]

Abstract

Brain arteriolosclerosis (B-ASC) is characterized by pathologically altered brain parenchymal arterioles. B-ASC is associated with cognitive impairment and increased likelihood of clinical dementia. To date, no study has been conducted on genome-wide genetic risk of autopsy-proven B-ASC. We performed a genome-wide association study (GWAS) of the B-ASC phenotype using multiple independent aged neuropathologic cohorts. Included in the study were participants with B-ASC autopsy and genotype data available from the NACC, ROSMAP, ADNI, and ACT data sets. Initial Stage 1 GWAS ($n = 3382$) and Stage 2 mega-analysis ($n = 4569$) were performed using data from the two largest cohorts (NACC and ROSMAP). Replication of top variants and additional Stage 3 mega-analysis were performed incorporating two smaller cohorts (ADNI and ACT). Lead variants in the top two loci in the Stage 2 mega-analysis (rs7902929, $p = 1.8 \times 10^{-7}$; rs2603462, $p = 4 \times 10^{-7}$) were significant in the ADNI cohort (rs7902929, $p = 0.012$; rs2603462, $p = 0.012$). The rs2603462 lead variant colocalized with *ELOVL4* expression in the cerebellum (posterior probability = 90.1%). Suggestive associations were also found near *SORCS1* and *SORCS3*. We thus identified putative loci associated with B-ASC risk, but additional replication is needed.

Keywords

VCID, SVD, arteriosclerosis, neuropathology, dementia, aging

Received 23 June 2021; Revised 18 September 2021; Accepted 14 October 2021

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 collagen.[1–4] B-ASC is commonly found in autopsied

[1]Department of Biostatistics, College of Public Health, University of Kentucky, Lexington, KY, USA
[2]Sanders-Brown Center on Aging and Alzheimer's Disease Research Center, University of Kentucky, Lexington, KY, USA
[3]Vanderbilt Memory & Alzheimer's Center, Department of Neurology, Vanderbilt University Medical Center, Nashville, TN, USA
[4]Department of Radiology & Imaging Sciences, Indiana University School of Medicine, Indianapolis, IN, USA
[5]Department of Medicine, University of Washington, Seattle, WA, USA
[6]ARUP Laboratories, Salt Lake City, UT, USA
[7]Office of the President, Brigham Young University–Hawaii, Laie, HI, USA
[8]Department of Medicine, Boston University, Boston, MA, USA
[9]Penn Neurodegeneration Genomics Center, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA
[10]Institute for Computational Biology, Case Western Reserve University, Cleveland, OH, USA
[11]Department of Neurology, Columbia University, New York, NY, USA
[12]Departments of Neurology and Pathology, Rush University Medical Center, Chicago, IL, USA
[13]Pathology and Laboratory Medicine, University of Kentucky, Lexington, KY, USA

Corresponding author:
David W Fardo, 725 Rose Street, Room 201, Lexington, KY 40536-0082, USA.
Email: david.fardo@uky.edu

elderly individuals, with more than 50% of participants showing some B-ASC pathology in large autopsy studies.[1,3,5] B-ASC is often comorbid with multiple neuropathologies including limbic-predominant age-related TDP-43 encephalopathy (LATE), Alzheimer's disease (AD), hippocampal sclerosis (HS), micro-infarcts, and large-vessel infarcts.[1,6,7] B-ASC is also associated with cognitive decline, even after adjusting for age and the presence of other neuropathologies.[1,6] Despite the clinical importance of B-ASC, its risk factors – other than age and sex – remain largely uncharacterized.[1] Hypertension (HTN) and diabetes mellitus (DM) 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 at death (<80 years vs. ≥80 years), HTN was significantly associated with B-ASC in the younger group only, and DM was not associated with B-ASC in either group.[1,8–10]

Genome-wide association studies (GWAS) have proven useful for investigating genomic risk of complex diseases via analysis of millions of genetic variants and have to date identified many thousands of genetic variants associated with disease phenotypes.[11] For example, GWAS have identified genetic risk loci for stroke and for imaging-based phenotypes of cerebrovascular disease such as white matter hyperintensities (WMH) and brain infarcts.[5,12] Studies of monogenic cerebrovascular disorders have identified genes associated with collagen and basement membrane formation, vascular smooth muscle, inflammation, and potassium ion channels.[4] GWAS of HS have identified multiple ion channel-associated genes[12] that have since been replicated in gene-based studies.[13]

There has been no published GWAS for autopsy-proven B-ASC to date.[14–16] 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 genetic risk for B-ASC may not be wholly shared with other cerebrovascular phenotypes.[4] In one previous study of Religious Orders Study/Memory and Aging Project (ROSMAP) participants, 167 independent genetic variants meeting genome-wide significance threshold ($p < 5 \times 10^{-8}$) in GWAS of stroke or stroke risk factors were tested for association with B-ASC pathology.[5] The authors found six variants nominally associated with B-ASC at the 5% significance level.[5] Another study found that an HS risk variant located in *ABCC9* was nominally associated with B-ASC.[1] We hypothesized that investigating genetic risk factors of B-ASC at the genome-wide level would provide important insights into its pathophysiological development as well as its relationship to neuroimaging and other neuropathological phenotypes.

In the present study, we conducted a GWAS using B-ASC pathology as an endophenotype in four independent autopsy cohorts with genotype data available.

Methods

An outline of our study design is shown in Figure 1. In Stage 1, a GWAS was performed on neuropathology data from the National Alzheimer's Coordinating Center (NACC) Neuropathology data set linked to GWAS genotype data from the Alzheimer's Disease Genetics Consortium (ADGC). In Stage 2, top variants from the NACC GWAS were first analyzed in ROSMAP, and then the two data sets were merged
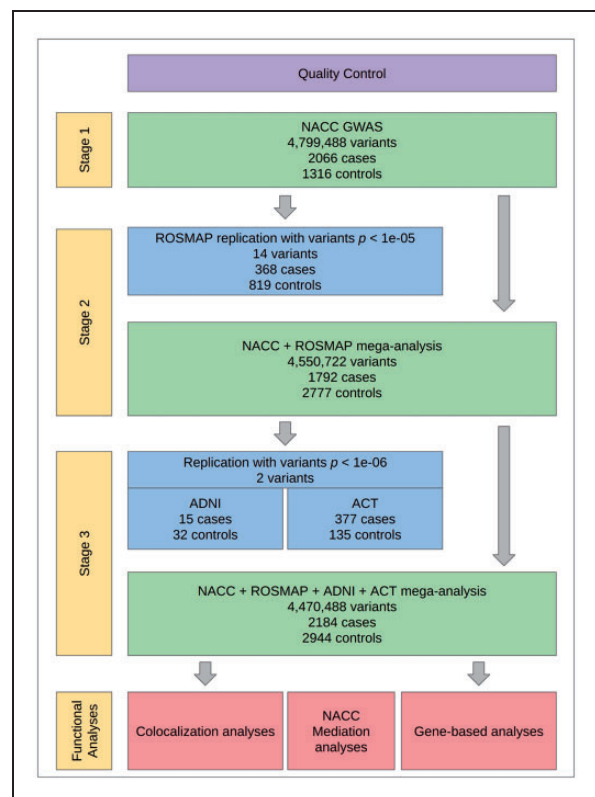

Figure 1. Overall study design. First, quality control (QC) was performed on all neuropathological and genotype data sets used (see Methods). GWAS were performed across three stages. In Stage 1, GWAS was performed on NACC participants (n = 3382). In Stage 2, all variants with $p < 1 \times 10^{-5}$ (k = 14) were analyzed separately in ROSMAP (n = 1187), and then NACC and ROSMAP participant data were merged for mega-analysis (n = 4569). In Stage 3, top variants from Stage 2 mega-analysis with $p < 1 \times 10^{-6}$ (k = 2) were analyzed in two replication cohorts, ADNI (n = 47) and ACT (n = 512), and then NACC, ROSMAP, ADNI, and ACT data were merged for mega-analysis. Downstream functional analyses, consisting of colocalization and gene-based analyses, were then performed on Stage 3 mega-analysis results. Mediation analyses for HTN and DM were also performed using a subset of NACC participants with clinical data available (n = 1726).

and mega-analyzed. In Stage 3, we analyzed the top variants from Stage 2 in two smaller autopsy cohorts from the Alzheimer's Disease Neuroimaging Initiative (ADNI) and the Adult Changes in Thought (ACT) study. We then mega-analyzed all four cohorts together. To investigate potential biologically functional correlates to disease risk, we then performed 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.[17] To investigate the potential of HTN or DM acting as mediators on B-ASC risk, we performed mediation analyses on a subset of NACC participants with available clinical data.

Study participants

All study participants were de-identified and deceased, so that archival samples were exclusively used. Thus our study did not fall under the definition of 'Human Subjects Research', according to the University of Kentucky Institutional Review Board, because of NIH Exemption #4–involving the collection/study of data or specimens if publicly available, and/or recorded such that subjects cannot be identified.

For the Stage 1 GWAS, neuropathology data from participants in >30 National Institute on Aging-funded Alzheimer's Disease Research Centers (ADRCs) studies in the NACC Neuropathology Data Set were linked to ADGC genotype data.[18,19] Each ADRC has its own study population, design, and participant recruitment strategies. Data are uniformly collected data and subsequently aggregated by NACC. Participants diagnosed with any of 19 unusual neurological conditions were then excluded because B-ASC may represent something different in the presence of these rare diseases (see Supplementary Table S1 for full exclusion criteria). Similar exclusion criteria were not applied to other studies' participants due to lack of data availability. A total of 3501 participants had both B-ASC neuropathology and genotype data available and passed initial inclusion 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).[20] 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 1213 ROSMAP participants had both autopsy and genotype data available.

The ADNI (adni.loni.usc.edu) was launched in 2003 as a public-private research partnership, led by Principal Investigator Michael W. Weiner, MD.

The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early AD. A subset of ADNI participants undergo autopsy and receive neuropathological phenotyping. A total of 60 ADNI participants had both B-ASC and genotype data available.

The ACT study began in 1994 and recruited residents in the greater Seattle area aged 65 years and older without dementia at time of enrollment.[21] The goal of the study has expanded to include three cohorts and continuous enrollment using the same enrollment criteria and has a current total of 4,960 participants across all three cohorts. A total of 518 ACT participants had both B-ASC and genotype data available.

Definitions of B-ASC variables used

In the NACC Neuropathology Data Set, B-ASC was graded as an ordinal variable with possible values of 0 (none), 1 (mild), 2 (moderate), or 3 (severe). Grading was performed by trained neuropathologists at each ADRC. The NACC NP v10 Coding Guidebook (https://files.alz.washington.edu/documentation/np10-guidebook.pdf) instructs examiners to grade B-ASC based on subcortical white or gray matter and to exclude arteriolar thickening due secondary to cerebral amyloid angiopathy. On page 13, the Guidebook defines B-ASC as, "concentric hyaline thickening of the media of arterioles. Intimal fibrosis may also accompany this change. The term 'lipo-hyalinosis' is sometimes used to refer to the same pathologic change." The B-ASC variable in ROSMAP was 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.[22] ADNI and ACT autopsy procedures followed the NACC Neuropathology codebook.[23]

Identifying ethnic outliers

In all cohorts, principal component analysis (PCA) in PLINK v1.9[24] was performed with a pruned subset of independent (linkage disequilibrium (LD) $r^2 < 0.05$) variants from each data set merged to data from the 1000 Genomes Project Phase 3 (1000 Genomes, n = 2504).[25] All variants were checked to ensure that major and minor alleles matched in the study data sets and 1000 Genomes. The standardized first and second principal components (PCs) were then plotted for each participant using the ggplot2 R package in R version

$4.0.4.$[26] All participants within a Euclidean distance of 0.35 from the centroid of 1000 Genomes participants in the EUR superpopulation were considered to be of European descent and included in analyses. We then re-ran PCA for the included participants and included the first five PCs as covariates in regression models.

Quality control (QC) of participants and genotype data

Imputation for ADGC was performed using the Haplotype Reference Consortium (HRC) panel[27] (workflow can be found at https://github.com/KBoeh me/ADGC_HRC_MERGE). The ROSMAP genotype QC workflow has been previously described.[28] Briefly, imputation was performed for ROSMAP on the Michigan Imputation Server[29] using the HRC panel. Imputation for ACT was performed on the Michigan Imputation Server using the Trans-Omics for Precision Medicine (TOPMed) reference panel.[30,31] We performed additional standard QC procedures on all genotyping data using PLINK v1.9 and KING.[24,32–35] 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 $p < 1 \times 10^{-6}$ using all participants in a cohort.

Participants were removed if they had a genotype missingness rate above 5% or a heterozygosity rate more than 3 standard deviations above or below the mean of their cohort. All participants were checked for duplicate genotype information across studies. All participants with duplicate information within one cohort were removed, while participants with duplicate information across cohorts were preferentially kept in the data set analyzed first and removed from cohorts used later in the study (*i.e.* NACC > ROSMAP > ADNI and ACT). Participant relatedness, as measured by proportion identity by descent (π) was estimated using PLINK. For clusters of closely related participants ($\pi \geq 0.18$), preference was given to the participant with the highest genotyping rate and others were removed. In NACC analyses, 4799488 variants and 3382 participants passed QC protocols. In the ROSMAP replication data set, 4889494 variants and 1187 participants passed QC protocols. Thirteen ADNI participants were identified as NACC duplicates and removed, leaving 47 participants for analysis. A total of 512 ACT participants passed QC measures.

Statistical analyses

Single-variant analyses. Single-variant association analyses for NACC were performed using logistic regression in PLINK v1.9 and ordinal regression in R using the MASS package.[26,32] To create a dichotomous outcome variable for logistic regression, participants with none or mild B-ASC were treated as controls and those with moderate or severe B-ASC as cases. This decision was made because previous studies have found that moderate-to-severe B-ASC is associated with worse cognitive functioning.[1,6] Covariates in the regression models included age at death, sex, ADGC genotyping cohort, and the first five PCs. An additive mode of inheritance was assumed in all analyses. A Bonferroni-corrected threshold of $p < 5 \times 10^{-8}$ was used for genome-wide significance and a predetermined threshold of $p < 1 \times 10^{-5}$ for "suggestive" association meriting further investigation. Then, using PLINK v1.9, variants meeting the suggestive threshold were clumped to create a set of independent variants (defined as LD $r^2 < 0.05$). For suggestive variants in ordinal regression analyses, the proportional odds assumption was tested using the Brant test in the brant R package.[36] Finally, we examined single variants previously found to be putatively associated with B-ASC in Chou et al. (2013)[5] to determine if any were validated in NACC at the $p < 0.05$ significance level.

In the ROSMAP analyses, single-variant regression analyses were performed using variants identified as suggestively significant in the NACC analyses, using a significance threshold of $p < 0.05$ for these variants. Covariates included age at death, sex, study (ROS vs. MAP) and the first five PCs. We then performed a mega-analysis on both cohorts. We also performed both fixed- and random-effects meta-analyses in PLINK v1.9 for NACC and ROSMAP logistic GWAS results.

To seek replication of our results, variants with NACC + ROSMAP mega-analytic $p < 1 \times 10^{-6}$ were analyzed separately in the ADNI and ACT data sets. The ADNI and ACT participants were then added to the combined NACC and ROSMAP data set to perform a combined mega-analyss. A total of 4470488 variants were shared between the NACC, ROSMAP, ADNI, and ACT data sets that passed QC measures.

To test if variants associated with B-ASC were also associated with large-vessel cerebrovascular disease, we tested the association of variants with $p < 1 \times 10^{-6}$ in the Stage 2 mega-analysis with intracranial atherosclerosis in a recent GWAS performed on ROSMAP participants.[37]

Gene-based analyses. Following single-variant analyses, we performed gene-based analyses using MAGMA.[38] Variants were mapped to genes within 1000 kilo-base pairs (kb) using the Genome Reference Consortium Human Build 37 (GRCh37/hg19) positions available on the MAGMA website (https://ctg.cncr.nl/software/ magma), producing a total of 18473 genes with at least

one annotated variant. We then performed gene-based analyses in MAGMA using a Bonferroni-corrected significance threshold of $p < 2.5 \times 10^{-6}$.

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 GTEx project V8 publicly available data (accessible at https://www.gtexportal.org/home/datasets).[17,39–41] QTL are calculated in GTEx by performing single-variant analysis of gene expression for all variants within 1000 kb of the transcription start and end sites for each gene. First, we systematically checked to determine if suggestive variants ($p < 1 \times 10^{-5}$) from the Stage 3 mega-analysis were significant expression QTL (eQTL) or splicing QTL (sQTL) across 52 tissues using GTEx summary statistics of participants of European descent. We then performed colocalization analysis on each significant phenotype/tissue combination using default prior probabilities in the coloc package ($p_1 = p_2 = 1 \times 10^{-4}$ and $p_{12} = 1 \times 10^{-5}$). These parameter choices indicate that each variant used has a 1/10,000 prior probability of being a B-ASC risk variant or a QTL, and a variant that is either a B-ASC risk variant or a QTL has a 1/10 prior probability of being both.[40] Because the coloc package can currently only analyze dichotomous or continuous variables, we did not analyze variants from the ordinal regression analyses. A posterior probability of colocalization (PPH4) of 50% or greater was chosen to indicate evidence for colocalization.

Mediation analyses. To test the hypotheses that variants associated with B-ASC risk may be mediated by HTN or DM, we performed mediation analyses using R on the subset of NACC participants with clinical variables available.[1,42] Participants were labeled as DM or HTN cases if they had at least one of 1) self-reported diagnosis, 2) clinician-reported diagnoses, or 3) reported use of DM or HTN medication on their most recent clinical visit prior to death.

Sensitivity analyses. Given previously identified potential differences in clinical risk factors for B-ASC in participants stratified by age at death, we re-analyzed our analyses in NACC and ROSMAP using only participants with an age at death of 80 or above. To further assess the robustness of our results under different model assumptions, we performed several sensitivity analyses in our NACC data set 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.[43–46] To overcome issues with computing PCs with samples with related participants, we used the PC-AiR method.[45]

Results

Of the 3382 NACC participants that met inclusion criteria for analysis, 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 (Table 1). ROSMAP participants that met inclusion for analysis had comparatively less B-ASC pathology ($p = 0.002$): 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

Table 1. Cohorts used in study.

Variable	Labels	NACC	ROSMAP	ADNI	ACT
B-ASC					
	None	935 (27.6%)	414 (34.9%)	1 (2.1%)	7 (1.4%)
	Mild	1023 (30.2%)	405 (34.1%)	31 (66%)	128 (25%)
	Moderate	1043 (30.8%)	284 (23.9%)	13 (27.7%)	257 (50.2%)
	Severe	381 (11.3%)	84 (7.1%)	2 (4.3%)	120 (23.4%)
Sex					
	Female	1680 (49.7%)	800 (67.4%)	8 (17%)	286 (55.9%)
	Male	1702 (50.3%)	387 (32.6%)	39 (83%)	226 (44.1%)
Age of death					
	Mean (SD)	82 (9.4)	89.6 (6.5)	83.3 (7.1)	89 (6.5)
	Median [Min, Max]	83 [47, 111]	90 [66, 108.3]	84 [59, 97]	90 [70, 106]

Note: Age of Death variable is integer for NACC, ADNI, and ACT but continuous in ROSMAP.
SD: standard deviation; Min: minimum; Max: maximum.

were also significantly more likely to be male (50% vs.33%, $p < 0.001$) and had younger ages at death on average (mean age at death 82 vs. 89.6, $p < 0.001$) compared to ROSMAP participants. ADNI had the highest proportion of male participants of any cohort at 83%, while ACT had the lowest percentage at participant with no B-ASC with 7 (1.4%) participants.

Single-variant analyses

In the NACC and ROSMAP cohorts, 1424 (42.1%) and 368 (31%) of participants had either moderate or severe B-ASC and were counted as cases, respectively. In the Stage 1 NACC GWAS ($n = 3382$), one locus on Chromosome 6q14.1 with lead variant rs2603462 was genome-wide significantly associated with B-ASC (odds ratio [OR] = 1.45, $p = 2.5 \times 10^{-8}$). We identified 13 other loci that met our suggestive association threshold of $p < 1 \times 10^{-5}$, described in Table 2. Of the 14 loci that met the suggestive threshold, rs7902929 on Chromosome 10q25.1 was significant at $p < 0.05$ level in ROSMAP (NACC OR = 1.57, $p = 7.8 \times 10^{-6}$; ROSMAP OR = 1.61, $p = 0.0069$). For the other 13 loci, two variants did not pass QC in ROSMAP, two had non-significant ROSMAP ORs < 1, and 9 had non-significant effect sizes in the same direction as in NACC. Results in the ordinal regressions were broadly similar to the logistic analyses in both cohorts. Of the top 14 loci in the NACC logistic regression analysis, four also met the suggestive threshold in the ordinal regression analysis, and 10 had $p < 1 \times 10^{-4}$ (Supplementary Table S2) Of the six variants found to be nominally associated with B-ASC by Chou et al. (2013),[5] no variants were validated at the $p < 0.05$ significance level in the NACC cohort.

In the Stage 2 NACC and ROSMAP mega-analysis ($n = 4569$), 11 loci met the suggestive significance threshold, with rs7902929 (OR = 1.58, $p = 1.8 \times 10^{-7}$) and rs2603462 (OR = 1.34, $p = 4 \times 10^{-7}$) producing the smallest p-values. Findings from the Stage 2 mega-analysis are illustrated in Figure 2. Because the ADNI ($n = 47$) and ACT ($n = 512$) sample sizes were much smaller than those NACC and ROSMAP, we limited the replication attempts in these cohorts to variants meeting a more stringent threshold of $p < 1 \times 10^{-6}$, which two variants, rs2603462 and rs7902929, met. In the ADNI neuropathology cohort, effects of both rs2603462 (OR [95% confidence interval (CI)] =4.75 [1.42–15.91], $p = 0.012$) and rs7902929 (OR [95% CI] =26.49 [2.07–338.8], $p = 0.012$) were replicated at the $p < 0.05$ significance level (Table 3). In the ACT neuropathology cohort, neither variant was significant, and both variants had non-significant effects in the opposite direction than in the other cohorts. rs2603462 was associated with cerebral atherosclerosis in ROSMAP ($p = 0.02$) while rs7902929 was not ($p = 0.56$).

In the Stage 3 mega-analysis combining the NACC, ROSMAP, ADNI, and ACT cohorts, no variants met the genome-wide significance threshold. A total of nine loci met the threshold for suggestive significance (Table 4), of which five were identified in Stage 2, while four were novel loci. Both rs2603462 and rs7902929 met the suggestive threshold.

Gene-based analysis

In the gene-based analysis, no genes achieved Bonferroni-adjusted significance ($p < 2.5 \times 10^{-6}$). *SORCS1* ($p = 5.5 \times 10^{-5}$) achieved the smallest p-value of any

Table 2. Stage 1 GWAS results.

SNP	CHR	BP	Gene	A1/A2	NACC OR [95% CI]	NACC P	ROSMAP OR [95% CI]	ROSMAP P
rs2418491	2	59,535,434	*FANCL*	G/A	1.33 [1.19–1.49]	4.8e-07	1.02 [0.84–1.23]	0.84
rs3774902	4	23,890,782	*PPARGC1A*	A/G	1.77 [1.42–2.21]	4.5e-07	–	–
rs4470773	5	177,507,080	*N4BP3*	G/C	1.28 [1.15–1.42]	3.9e-06	1.07 [0.89–1.29]	0.47
rs28362345	6	31,165,836	*BCKDHB*	T/C	1.27 [1.15–1.41]	6.3e-06	0.94 [0.78–1.14]	0.53
rs115980554	6	34,456,632	*PACSIN1*	C/T	1.63 [1.33–2]	3.4e-06	0.95 [0.65–1.37]	0.77
rs2603462	6	81,418,667	*HCG27*	C/A	1.45 [1.27–1.66]	2.5e-08	1.05 [0.84–1.33]	0.65
rs12700439	7	23,472,061	*IGF2BP3*	T/C	1.26 [1.14–1.4]	8.0e-06	1.07 [0.89–1.29]	0.47
rs7902929	10	107,237,532	*SORCS3*	C/T	1.57 [1.29–1.92]	7.8e-06	1.61 [1.14–2.26]	0.0069
rs10790707	11	124,460,485	*OR8A1/PANX3*	A/G	1.26 [1.14–1.4]	8.1e-06	1.03 [0.85–1.24]	0.76
rs61944465	13	27,272,704	*WASF3*	G/A	1.33 [1.18–1.5]	3.6e-06	1.16 [0.93–1.46]	0.19
rs6574718	14	26,395,832	*NOVA1*	C/T	1.28 [1.15–1.42]	5.3e-06	–	–
rs9895518	17	49,645,803	*CA10*	A/G	1.28 [1.15–1.42]	8.3e-06	1.09 [0.9–1.32]	0.39
rs387083	19	35,847,115	*FFAR1/FFAR3*	G/A	1.28 [1.15–1.42]	5.4e-06	1.02 [0.85–1.22]	0.86
rs2069126	20	19,749,957	*SLC24A3*	A/G	1.27 [1.14–1.4]	6.7e-06	1.11 [0.92–1.34]	0.27

Key: SNP, single nucleotide polymorphism; CHR, Chromosome; BP, base pair; Gene, closest protein-coding gene; A1/A2, effect/non-effect allele; OR, odds ratio; 95% CI, 95% confidence interval. A1 set so that NACC OR >= 1.
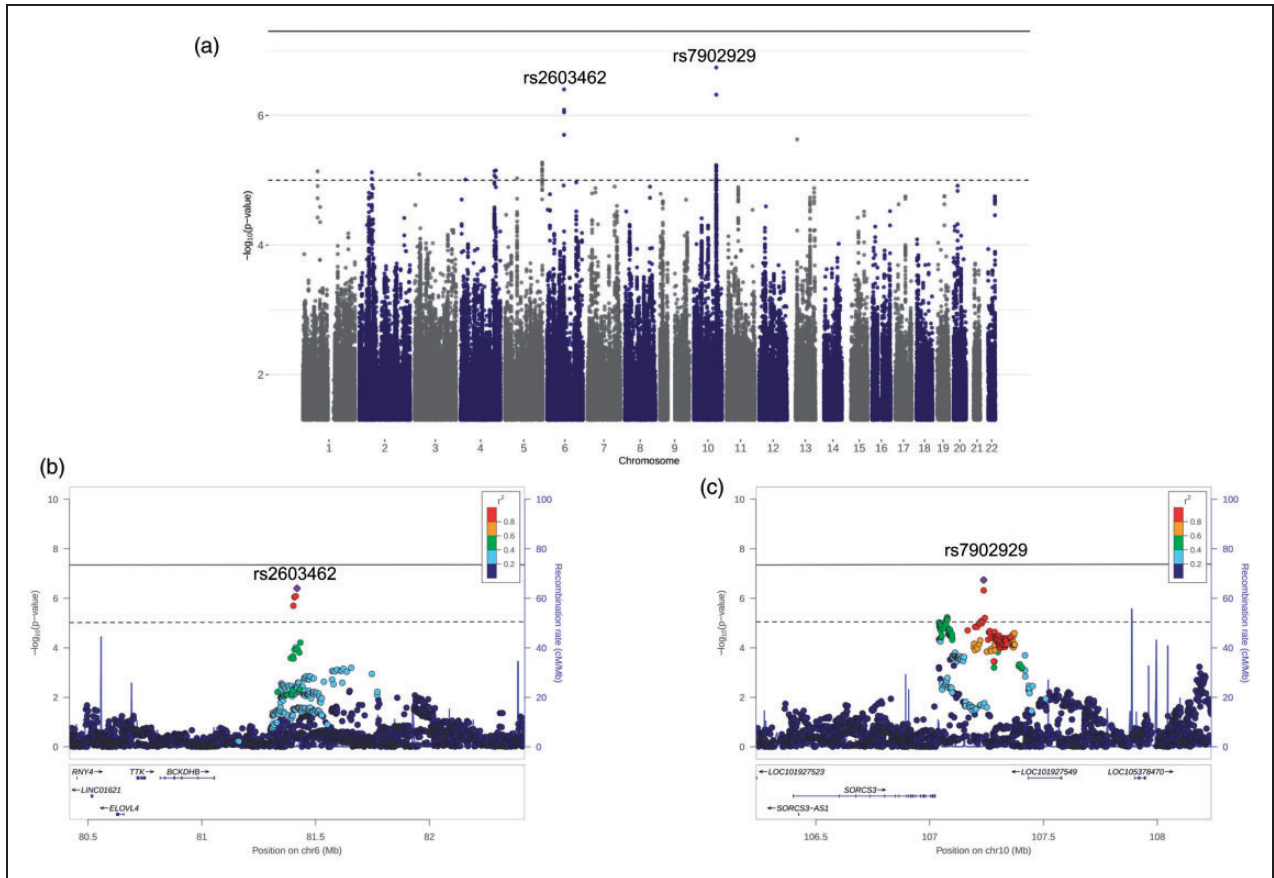
Figure 2. Stage 2 mega-analysis results. (a) Manhattan plot of Stage 2 GWAS mega-analysis of NACC and ROSMAP. The horizontal dashed line in each plot is the suggestive threshold ($p < 1 \times 10^{-5}$) while the solid line is the threshold for genome-wide significance ($p < 5 \times 10^{-8}$). (b & c) Regional plots of rs2603462 and rs7902929 +/− 1000 kb, respectively.

Table 3. Stage 2 mega-analysis and replication results.

SNP	CHR	BP	Gene	Stage 2 mega-analysis		ADNI		ACT	
				OR [95% CI]	P	OR [95% CI]	P	OR [95% CI]	P
rs61776730	1	71,057,739	*BC041441*	1.46 [1.24–1.72]	7.3e-06	–	–	–	–
rs2418491	2	59,535,434	*FLJ30838*	1.24 [1.13–1.37]	7.6e-06	–	–	–	–
rs13082422	3	22,338,009	*ZNF385D*	1.22 [1.12–1.34]	8.1e-06	–	–	–	–
rs3774902	4	23,890,782	*PPARGC1A*	1.55 [1.27–1.88]	9.7e-06	–	–	–	–
rs7675509	4	161,518,462	*FSTL5*	1.24 [1.13–1.36]	7.2e-06	–	–	–	–
rs10050232	4	168,333,572	*SPOCK3*	1.25 [1.14–1.38]	7.0e-06	–	–	–	–
rs6898408	5	58,406,648	*PDE4D*	1.23 [1.12–1.35]	9.3e-06	–	–	–	–
rs4370294	5	177,510,515	*N4BP3*	1.24 [1.13–1.35]	5.3e-06	–	–	–	–
rs2603462	6	81,418,667	*BCKDHB*	1.34 [1.2–1.5]	4.0e-07	4.75 [1.42–15.91]	0.012	0.77 [0.52–1.13]	0.18
rs7902929	10	107,237,532	*SORCS3*	1.58 [1.33–1.87]	1.8e-07	26.49 [2.07–338.8]	0.012	0.81 [0.47–1.4]	0.45
rs61944465	13	27,272,704	*WASF3*	1.29 [1.16–1.44]	2.3e-06	–	–	–	–

Note: A1 set so that Stage 2 mega-analysis OR >= 1.
SNP: single nucleotide polymorphism; CHR: Chromosome; BP: base pair; Gene: closest protein-coding gene; A1/A2: effect/non-effect allele; OR: odds ratio; 95% CI: 95% confidence interval.

Table 4. Stage 3 mega-analysis results.

SNP	CHR	BP	Gene	A1/A2	OR [95% CI]	P
rs61776730	1	71,057,739	*BC041441*	A/G	1.44 [1.23–1.69]	4.6e-06
rs4370294	5	177,510,515	*N4BP3*	G/A	1.23 [1.13–1.34]	3.3e-06
rs2603462	6	81,418,667	*BCKDHB*	C/A	1.3 [1.17–1.45]	1.9e-06
rs10738370	9	14,582,342	*ZDHHC21*	T/C	1.25 [1.13–1.38]	8.9e-06
rs7902929	10	107,237,532	*SORCS3*	C/T	1.51 [1.28–1.78]	7.1e-07
rs76828179	12	32,594,268	*FGD4*	A/G	1.36 [1.19–1.56]	5.2e-06
rs61944465	13	27,272,704	*WASF3*	G/A	1.29 [1.16–1.42]	1.1e-06
rs415468	19	44,456,672	*ZNF221*	C/G	1.22 [1.11–1.32]	9.3e-06
rs16981092	20	19,745,011	*SLC24A3*	G/A	1.22 [1.11–1.32]	8.5e-06

Note: A1 set so that OR >= 1.
SNP: single nucleotide polymorphism; CHR: Chromosome; BP: base pair; Gene: closest protein-coding gene; A1/A2: effect/non-effect allele; OR: odds ratio; 95% CI: 95% confidence interval.

Table 5. Expression quantitative trait loci (eQTL) Colocalizing with B-ASC risk.

SNP	CHR	Phenotype	Tissue	PPH4
rs4608937	5	NSD1	Heart: Atrial Appendage	94.5
rs2603462	6	ELOVL4	Brain: Cerebellar Hemisphere	90.1
rs61944465	13	WASF3	Brain: Caudate basal ganglia	94.8
rs61944465	13	WASF3a	Cells: Cultured fibroblasts	85.1
rs61944465	13	WASF3	Brain: Cerebellar Hemisphere	81.6
rs61944465	13	WASF3	Brain: Cerebellum	79.0
rs16981092	20	RIN2	Thyroid	80.9

SNP: single-nucleotide polymorphism; CHR: Chromosome; PPH$: posterior probability of colocalization.
[a]Denotes sQTL phenotype.

gene, while the adjacent *SORCS3* gene *p*-value was somewhat larger ($p = 0.0083$, Supplementary Table S3).

Colocalization analyses

In the Stage 3 mega-analysis, four suggestive variants were significant QTLs for four genes across eight tissues in GTEx, leading us to perform colocalization analyses for a total of 10 phenotype-tissue combinations. Of these, seven met our evidential threshold of PPH4 > 50% for colocalization. The strongest evidence for colocalization was found for *WASF3* expression, which colocalized with B-ASC risk across three brain regions, including the basal ganglia (PPH4 = 94.8%). Preferential splicing of *WASF3* in fibroblasts colocalized with PPH4 = 85.1% (Table 5). *NSD1* colocalized with B-ASC risk in the atrial appendage (PPH4 = 94.5%). *ELOVL4* expression in the cerebellar hemisphere colocalized with B-ASC risk with a PPH4 of 90.1%. Finally, *RIN2* expression colocalized with B-ASC risk in the Thyroid (PPH4 = 94.5%).

Mediation analyses

In the subset of our NACC participants who had clinical data for DM ($n = 1725$) and HTN ($n = 1726$) status, none of the suggestive variants were associated

with HTN or DM at the $p < 0.05$ level, and all variant effect sizes in regressions including HTN or DM were within the 95% confidence intervals in the base model excluding them.

Sensitivity analyses

In the analysis using only participants with an age at death of 80 years or older, no variants achieved genome-wide significance. In the NACC analysis, we identified 12 independent loci that met the suggestive threshold. No suggestive variants from the NACC analyses were validated in the ROSMAP analyses at the $p < 0.05$ significance threshold (Supplementary Table S2). Performing fixed-effects meta-analysis on the NACC 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.

Discussion

The present study constituted the first GWAS of autopsy-proven B-ASC using neuropathology and

genotype data from four large autopsy cohorts. A significant association was found between one locus on Chromosome 6 and B-ASC in the Stage 1 NACC GWAS ($p = 2.5 \times 10^{-8}$). In the Stage 3 mega-analysis of all four cohorts, this locus colocalized with *ELOVL4* gene expression in GTEx (PPH4 = 90.1%), providing evidence that this locus may affect B-ASC risk through mediating *ELOVL4* expression (Table 5). Another locus on Chromosome 10 near *SORCS3* suggestively associated with B-ASC was replicated in the ROSMAP cohort. These variants' effects were also replicated in the ADNI cohort, but failed to replicate in the ACT cohort (Table 3). 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 the GTEx data.

ELOVL4, SORCS1, and SORCS3 loci

ELOVL4 codes for the elongation of very long chain fatty acids-4 protein, an "elongase" enzyme that catalyzes the synthesis of very long chain fatty acids (VLC-FA, $\geq$28 C) (see Supplementary Figure S1 for gene interaction network). Much of the 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.[47] In mice, *ELOVL4* knockouts showed significantly different distributions of VLCFA in brain tissue compared to wild-types.[48] Mutant *ELOVL4* mice show significantly increased brain metabolic activity compared to wild-types, increasing glucose uptake by 3-fold without concurrent changes to blood-brain barrier integrity.[47] Results also show that mutant mice had absent VLC-FA in synaptic vesicles, resulting in more erratic release of neurotransmitters.[47] These studies demonstrate that *ELOVL4* plays a critical role in VLC-FA metabolism, which are in turn necessary for proper central nervous system functioning.

While the *ELOVL4* mutations studied in basic science work have largely been loss-of-function mutations that cause highly penetrant monogenic disease, recent genomics works expands the role of *ELOVL4* as a contributor to polygenic traits. 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. Two recent studies 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, ($p = 2 \times 10^{-12}$), waist-to-hip ratio (rs76567515, $p = 3 \times 10^{-10}$), and waist-to-hip ratio adjusted for BMI (rs1849275, $p = 1 \times 10^{-9}$).[11,49,50] These variants were also

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 ($r^2 < 0.05$). Interestingly, the lead GWAS variant in our study was also nominally associated with cerebral atherosclerosis in a recent GWAS performed with ROSMAP participants ($p = 0.02$).[37] This result could indicate that *ELOVL4* may affect large-vessel disease risk in addition to B-ASC.

Another gene, *BCKDHB*, is located closer to the lead GWAS variant than *ELOVL4*. *BCKDHB* is one of three genes responsible for the monogenic metabolism disease Maple Syrup Urine Disease.[51] Additionally, another suggestive variant in NACC, rs387083, is located between the free fatty acid receptor genes *FFAR1* and *FFAR3* on Chromosome 19. This locus acts as an eQTL of *FFAR3* in the visceral omentum adipose. Collectively, these results suggest that fatty acid metabolism and signaling may play a role in B-ASC risk.

The suggestive NACC GWAS variant subsequently validated in ROSMAP, rs7902929, is located approximately 212 kb from the 3' end of the gene *SORCS3*. While the genes on which 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 (with *SORCS1* just outside of this range), 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 (see Supplementary Figure S1 for gene interaction network).[52,53] Previous studies using candidate gene designs have provided tentative evidence that genetic variation in *SORCS3* may be associated with AD.[52,53] In a recent GWAS of dementia in women enrolled in the the Women's Health Initiative,[54] a *SORCS3* intronic rare variant, rs76590698, was significantly associated with dementia ($p = 2 \times 10^{-8}$.[55] In the GWAS Catalog, *SORCS3*-mapped variants are significantly associated with multiple cognitively relevant phenotypes, including depressive symptoms (rs1021363, $p = 1 \times 10^{-13}$), self-reported educational attainment (rs11599236, $p = 1 \times 10^{-13}$), and systolic blood pressure (rs191784289, $p = 3 \times 10^{-13}$).[50,56,57]

Additionally, a member of the VPS10 family, *SORCS1*, achieved the smallest *p*-value in the gene-based analysis. We performed a *post-hoc* analysis of this region and identified two independent loci with lead *p*-values just above the suggestive threshold driving this effect: one locus in LD with rs7902929 and one not ($p = 1.2 \times 10^{-5}$). We examined the lead variant in the second locus and identified a consistent effect across all four cohorts that was nominally significant in NACC ($p = 1 \times 10^{-4}$) and ACT ($p = 0.001$), though

not in ROSMAP ($p = 0.25$) or ADNI ($p = 0.07$). Previous studies have identified *SORCS1* as a potential risk gene for AD.[52,53] In mouse models, *SORCS1* and *SORCS3* regulate protein transport in neurons and affect metabolism through the regulation of brain-derived neurotropic factor.[58] *SORCS1* and *SORCS3* are also associated with altered amyloid precursor protein (APP) processing,[52] and amylodosis causes subsequent downregulation of both genes.[59] It is possible that the VPS10 family influences B-ASC risk through AD risk, as B-ASC is associated with AD neuropathology,[6] though further work should be done to clarify the possible role of the VPS10 family on both phenotypes.

Limitations and sources of bias

Statistical modeling indicated that none of the effects of top loci in the NACC GWAS appeared to be mediated by HTN or DM status in NACC. This finding has some caveats, including the inconsistent evidence of association between these clinical risk factors and B-ASC pathology.[1,4] While HTN and DM may be associated with B-ASC in younger individuals, this association seems to be attenuated or non-existent in older individuals (80 years or older at time of death).[1] Indeed, in the present study, neither HTN nor DM were associated with B-ASC in the individuals used for mediation analysis, making the resulting lack of mediation of genetic risk not surprising. Moreover, this analysis was limited by the substantially smaller sample size ($n = 1726$) than that of the Stage 1 GWAS. Additionally, the diagnostic variables in NACC consisted of a mixture of patient self-reports, physician reporting, and use of HTN or DM medications. Future studies employing causal inference methods such as Mendelian randomization may be able to provide clearer evidence for or against the roles of HTN and DM in B-ASC and the mediation of B-ASC genetic risk by clinical risk factors.

Most genetic loci identified preliminarily as being associated with B-ASC in NACC were not replicated in ROSMAP. This could be potentially explained via the significantly different demographic attributes and study designs between the two cohorts, as ROSMAP participants were older at death and had lower average levels of diagnosed B-ASC pathology (Table 1). Thus, there may be unaccounted-for 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. In contrast, the NACC participants are recruited from primarily clinic-based populations across the ADRCs. Differences in study design are also seen in the ADNI cohort, whose recruitment practices resemble those of ADRCs, and in ACT, whose

participants are recruited from the aged dementia-free community in the Seattle, WA area.

Heterogeneity in neuropathological grading of B-ASC in each cohort may also contribute to impairing study power and limiting reproducibility between studies. In one study investigating the inter-rater reliability of cerebrovascular neuropathological diagnoses (i.e., how often neuropathologists agree on diagnosing a given slide), B-ASC had the second-lowest reliability (after microinfarcts), with reliability below 0.8 in five of twelve brain regions graded.[60] This relatively low inter-rater diagnostic reliability under near-ideal circumstances (different trained experts reviewing the exact same stained slides) indicates there may be substantial measurement variance in B-ASC between neuropathologists both within and between study centers. Furthermore, B-ASC may be graded in different brain regions in different centers and studies. Of note, in the Stage 3 mega-analysis including all four cohorts, the strongest evidence for colocalization (*WASF3*, PPH4 = 94.8%) was found in the caudate lobe of the basal ganglia, providing preliminary support for this brain region as a potential biological target for genetic B-ASC risk.

An important limitation of our study is the relatively low sample size for a GWAS. NACC, the largest cohort used, had only 3382 participants available for GWAS, and the total sample size of the Stage 3 mega-analysis with all four cohorts was 5128. In contrast, a recent GWAS meta-analysis of AD exceeded 90,000 participants,[61] and GWAS of other phenotypes have exceeded 1 million participants.[57] Nevertheless, smaller GWAS of other neuropathologic endophenotypes have identified risk loci,[13] demonstrating that smaller sample sizes do not necessarily prevent discovery, especially for large effect-size phenomena. Despite sample size limitations, one suggestive locus on Chromosome 10 near *SORCS3* identified in NACC was replicated in ROSMAP, and both this locus and a locus near *ELOVL4* were replicated in ADNI. Furthermore, nine of eleven loci suggestively associated with B-ASC 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 suffer from regression towards the mean in the ROSMAP analysis. Finally, the use of downstream analyses can provide additional functional evidence for risk loci that do not meet the conservative Bonferroni-corrected significance thresholds.

Acknowledgements

Declaration of conflicting interests

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article:Andrew J. Saykin receives [F18] Flortaucipir (AV-1451) precursor support from Avid Radiopharmaceuticals. Julie A. Schneider reports personal fees from Avid Radiopharmaceuticals (Eli Lilly Inc.), Alnylam Pharmaceuticals, and the National Hockey League. All other authors declare that they have no conflict of interest.

Authors' contributions

Lincoln M.P. Shade, Yuriko Katsumata, Peter T. Nelson, and David W. Fardo conceived of and carried out the research and drafted the manuscript. Timothy J. Hohman, Julie A. Schneider, Andrew J. Saykin, Shubabrata Mukherjee, Kevin L. Boehme, John S.K. Kauwe, Lindsay A. Farrer, Gerard D. Schellenberg, Jonathan L. Haines, and Richard P. Mayeux provided data. Shubabrata Mukherjee assisted with and provided consultation for

ACT analyses. Kwangsik Nho and Andrew J. Saykin assisted with and provided consultation for ADNI analyses.

ORCID iDs

Lincoln MP Shade ⓘD https://orcid.org/0000-0001-6093-4823
Andrew J Saykin ⓘD https://orcid.org/0000-0002-1376-8532

Supplemental material

Supplemental material for this article is available online.

References

1. Ighodaro ET, Abner EL, Fardo DW, et al.; Alzheimer's Disease Neuroimaging Initiative (ADNI). Risk factors and global cognitive status related to brain arteriolosclerosis in elderly individuals. *J Cereb Blood Flow Metab* 2017; 37: 201–216.

2. Grinberg LT and Thal DR. Vascular pathology in the aged human brain. *Acta Neuropathol* 2010; 119: 277–290.

3. Buchman AS, Yu L, Boyle PA, et al. Microvascular brain pathology and late-life motor impairment. *Neurology* 2013; 80: 712–718.

4. Blevins BL, Vinters HV, Love S, et al. Brain arteriolosclerosis. *Acta Neuropathol* 2021; 141: 1–24.

5. Chou S-Y, Shulman JM, Keenan BT, et al. Genetic susceptibility for ischemic infarction and arteriolosclerosis based on neuropathologic evaluations. *Cerebrovasc Dis* 2013; 36: 181–188.

6. Arvanitakis Z, Capuano AW, Leurgans SE, et al. Relation of cerebral vessel disease to Alzheimer's disease dementia and cognitive function in older persons: a cross-sectional study. *Lancet Neurol* 2016; 15: 934–943.

7. Neltner JH, Abner EL, Baker S, et al. Arteriolosclerosis that affects multiple brain regions is linked to hippocampal sclerosis of ageing. *Brain* 2014; 137: 255–267.

8. Wu J, Chen X, Xie Y, et al. Characteristics and risk factors of intrarenal arterial lesions in patients with IgA nephropathy. *Nephrol Dial Transplant* 2005; 20: 719–727.

9. Ikee R, Honda K, Ishioka K, et al. Postprandial hyperglycemia and hyperinsulinemia associated with renal arterio-arteriolosclerosis in chronic kidney disease. *Hypertens Res* 2010; 33: 499–504.

10. Cameron JS. The discovery of diabetic nephropathy: from small print to centre stage. *J Nephrol* 2006; 19 Suppl 10: S75–S87.

11. Buniello A, MacArthur JAL, Cerezo M, et al. The NHGRI-EBI GWAS catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res* 2019; 47: D1005–D1012.

12. Beecham GW, Hamilton K, Naj AC, et al.; Alzheimer's Disease Genetics Consortium (ADGC). Genome-Wide association Meta-analysis of neuropathologic features of Alzheimer's disease and related dementias. *PLOS Genet* 2014; 10: e1004606.

13. Katsumata Y, Nelson PT, Ellingson SR, et al. Gene-based association study of genes linked to hippocampal sclerosis of aging neuropathology: GRN, TMEM106B, ABCC9, and KCNMB2. *Neurobiol Aging* 2017; 53: 193. e17–193–e25.

14. Traylor M, Zhang CR, Adib-Samii P, et al. Genome-wide meta-analysis of cerebral white matter hyperintensities in patients with stroke. *Neurology* 2016; 86: 146–153.

15. Malik R, Chauhan G, Traylor M, et al.; MEGASTROKE Consortium. Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nat Genet* 2018; 50: 524–537.

16. Chauhan G, Adams HHH, Satizabal CL, et al. Genetic and lifestyle risk factors for MRI-defined brain infarcts in a population-based setting. *Neurology* 2019; 92: e486–e503.

17. The GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* 2013; 45: 580–585.

18. Welcome to Alzheimer's disease genetics consortium, ADGC, www.adgenetics.org/. (accessed 29 November 2021).

19. Besser LM, Kukull WA, Teylan MA, et al. The revised national Alzheimer's coordinating center's neuropathology form – available data and new analyses. *J Neuropathol Exp Neurol* 2018; 77: 717–726.

20. Bennett DA, Buchman AS, Boyle PA, et al. Religious orders study and rush memory and aging project. *J Alzheimers Dis* 2018; 64: S161–S189.

21. Kukull WA, Higdon R, Bowen JD, et al. Dementia and Alzheimer disease incidence: a prospective cohort study. *Arch Neurol* 2002; 59: 1737.

22. Buchman AS, Leurgans SE, Nag S, et al. Cerebrovascular disease pathology and parkinsonian signs in old age. *Stroke* 2011; 42: 3183–3189.

23. Franklin EE, Perrin RJ, Vincent B, et al.; Alzheimer's Disease Neuroimaging Initiative. Brain collection, standardized neuropathologic assessment, and comorbidity in Alzheimer's disease neuroimaging initiative 2 participants. *Alzheimers Dement* 2015; 11: 815–822.

24. Chang CC, Chow CC, Tellier LC, et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* 2015; 4: 7.

25. The 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature* 2012; 491: 56–65.

26. R Core Team. *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing, 2020.

27. McCarthy S, Das S, Kretzschmar W, et al.; Haplotype Reference Consortium. A reference panel of 64,976

haplotypes for genotype imputation. *Nat Genet* 2016; 48: 1279–1283.

28. Dumitrescu L, Mahoney ER, Mukherjee S, et al.; The Alzheimer's Disease Neuroimaging Initiative (ADNI). Genetic variants and functional pathways associated with resilience to Alzheimer's disease. *Brain* 2020; 143: 2561–2575.

29. Das S, Forer L, Schönherr S, et al. Next-generation genotype imputation service and methods. *Nat Genet* 2016; 48: 1284–1287.

30. Fuchsberger C, Abecasis GR and Hinds DA. minimac2: faster genotype imputation. *Bioinformatics* 2015; 31: 782–784.

31. Taliun D, Harris DN, Kessler MD, et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* 2021; 590: 290–299.

32. Chang CC and Purcell S. *Plink 1.9*, www.cog-genomics. org/plink/1.9/. (accessed 29 November 2021).

33. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007; 81: 559–575.

34. Manichaikul A, Mychaleckyj JC, Rich SS, et al. Robust relationship inference in genome-wide association studies. *Bioinformatics* 2010; 26: 2867–2873.

35. Marees AT, de Kluiver H, Stringer S, et al. A tutorial on conducting genome-wide association studies: quality control and statistical analysis. *Int J Methods Psychiatr Res* 2018; 27: e1608.

36. Schlegel B and Steenbergen M. *Brant: Test for parallel regression assumption*, https://CRAN.R-project.org/package=brant (2020, accessed 29 November 2021).

37. Vattathil SM, Liu Y, Harerimana NV, et al. A genetic study of cerebral atherosclerosis reveals novel associations with NTNG1 and CNOT3. *Genes* 2021; 12: 815.

38. Leeuw CA, de Mooij JM, Heskes T, et al. MAGMA: generalized gene-set analysis of GWAS data. *PLOS Comput Biol* 2015; 11: e1004219.

39. Battle A, Brown CD, Engelhardt BE, et al.; eQTL manuscript working group. Genetic effects on gene expression across human tissues. *Nature* 2017; 550: 204–213.

40. Giambartolomei C, Vukcevic D, Schadt EE, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLOS Genet* 2014; 10: e1004383.

41. GTEx portal, www.gtexportal.org/home/gene/ENSG00000183662.6#geneExpression (accessed 29 November 2021).

42. Baron RM and Kenny DA. The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations. *J Pers Soc Psychol* 1986; 51: 1173–1182.

43. Chen H and Conomos MP. *GMMAT: generalized linear mixed model association tests*, https://CRAN.R-project. org/package=GMMAT (2020, accessed 29 November 2021).

44. Conomos MP, Gogarten SM, Brown L, et al. *GENESIS: GENetic EStimation and inference in structured samples (GENESIS): Statistical methods for analyzing genetic data from samples with population structure and/or relatedness*, https://github.com/UW-GAC/GENESIS (2019, accessed 29 November 2021).

45. Gogarten SM, Sofer T, Chen H, et al. Genetic association testing using the GENESIS R/bioconductor package. *Bioinformatics* 2019; 35: 5346–5348.

46. Zheng X, Levine D, Shen J, et al. A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* 2012; 28: 3326–3328.

47. Hopiavuori BR, Anderson RE and Agbaga M-P. ELOVL4: very long-chain fatty acids serve an eclectic role in mammalian health and function. *Prog Retin Eye Res* 2019; 69: 137–158.

48. Vasireddy V, Sharon M, Salem N, et al. Role of ELOVL4 in Fatty Acid Metabolism. In: Anderson RE, LaVail MM and Hollyfield JG (eds). *Recent Advances in Retinal Degeneration*. New York, NY: Springer, pp. 283–290.

49. Zhu Z, Guo Y, Shi H, et al. Shared genetic and experimental links between obesity-related traits and asthma subtypes in UK biobank. *J Allergy Clin Immunol* 2020; 145: 537–549.

50. Kichaev G, Bhatia G, Loh P-R, et al. Leveraging polygenic functional enrichment to improve GWAS power. *Am J Hum Genet* 2019; 104: 65–75.

51. Strauss KA, Puffenberger EG and Carson VJ. Maple Syrup Urine Disease. GeneReviews®, 2006. http://www.ncbi.nlm. nih.gov/books/NBK1319/ (accessed 28 May 2021).

52. Reitz C, Tosto G, Vardarajan B, et al.; the Alzheimer's Disease Genetics Consortium (ADGC). Independent and epistatic effects of variants in VPS10-d receptors on Alzheimer disease risk and processing of the amyloid precursor protein (APP). *Transl Psychiatry* 2013; 3: e256–e256.

53. Wang M, Wang S, Li Y, et al. Integrated analysis and network pharmacology approaches to explore key genes of Xingnaojing for treatment of Alzheimer's disease. *Brain Behav* 2020; 10: e01610.

54. Design of the women's health initiative clinical trial and observational study. *Control Clin Trials* 1998; 19: 61–109.

55. Blue EE, Thornton TA, Kooperberg C, et al. Non-coding variants in MYH11, FZD3, and SORCS3 are associated with dementia in women. *Alzheimers Dement* 2021; 17: 215–225.

56. Baselmans BML, Jansen R, Ip HF, et al.; Social Science Genetic Association Consortium. Multivariate genome-wide analyses of the well-being spectrum. *Nat Genet* 2019; 51: 445–451.

57. Lee JJ, Wedow R, Okbay A, et al.; Social Science Genetic Association Consortium. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat Genet* 2018; 50: 1112–1121.

58. Subkhangulova A, Malik AR, Hermey G, et al. SORCS1 and SORCS3 control energy balance and orexigenic peptide production. *EMBO Rep* 2018; 19: e44810.

59. Hermey G, Hoffmeister-Ullerich SA, Merz B, et al. Amyloidosis causes downregulation of SorLA, SorCS1 and SorCS3 expression in mice. *Biol Chem* 2019; 400: 1181–1189.

60. Skrobot OA, Attems J, Esiri M, et al. Vascular cognitive impairment neuropathology guidelines (VCING): the contribution of cerebrovascular pathology to cognitive impairment. *Brain* 2016; 139: 2957–2969.

61. Kunkle BW, Grenier-Boley B, Sims R, et al.; Genetic and Environmental Risk in AD/Defining Genetic, Polygenic and Environmental Risk for Alzheimer's Disease Consortium (GERAD/PERADES). Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates a?, tau, immunity and lipid processing. *Nat Genet* 2019; 51: 414–430.