# Title page

**Title**

Genome-wide association study of multiple neuropathology endophenotypes identifies novel risk loci and provides insights into genetic risk of dementia

**Authors**

Lincoln M.P. Shade (1), Yuriko Katsumata (1, 2), Erin L. Abner (2, 3), Steven A. Claas (1), Timothy J. Hohman (4), Shubhabrata Mukherjee (5), Margaret A. Pericak-Vance (6), Richard P. Mayeux (7), Lindsay A. Farrer (8), Gerard D. Schellenberg (9), Jonathan L. Haines (10), Walter A. Kukull (11), Kwagsik Nho (12), Andrew J. Saykin (12), David A. Bennett (13), Julie A. Schneider (14), Mark T. W. Ebbert (2, 15, 16), Peter T. Nelson (16, 17), David W. Fardo (1, 2), the National Alzheimer’s Coordinating Center, and the Alzheimer’s Disease Genetics Consortium

**Add Qi and Khine as coauthors**

**Author Affiliations**

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. Department of Epidemiology and Environmental Health, College of Public Health, University of Kentucky, Lexington, KY, USA;
4. Vanderbilt Memory & Alzheimer’s Center, Department of Neurology, Vanderbilt University Medical Center, Nashville, TN, USA;
5. Department of Medicine, University of Washington, Seattle, WA, USA;
6. John P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL, USA;
7. Department of Neurology, Columbia University, New York, NY, 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. Cleveland Institute for Computational Biology, Case Western Reserve University, Cleveland, OH, USA;
11. Department of Epidemiology, University of Washington, Seattle, WA, USA;
12. Department of Radiology & Imaging Sciences, Indiana University School of Medicine, Indianapolis, IN, USA;
13. Rush Alzheimer’s Disease Center and Department of Neurological Sciences, Rush Medical College, Chicago, IL, USA;
14. Rush Alzheimer’s Disease Center and Departments of Neurology and Pathology, Rush University Medical Center, Chicago, IL, USA;
15. Division of Biomedical Informatics, Department of Internal Medicine, University of Kentucky, Lexington, KY, USA;
16. Department of Neuroscience, University of Kentucky College of Medicine, Lexington, KY, USA;
17. Department of Pathology and Laboratory Medicine, University of Kentucky, Lexington, KY, USA

**Corresponding Author**

David W. Fardo

Healthy Kentucky Research Building

760 Press Ave, Room 372

Lexington, KY 40536-0679

[david.fardo@uky.edu](mailto:david.fardo@uky.edu)

859-218-2070

# Key words

GWAS, Alzheimer’s disease, neuropathology, atherosclerosis, neurofibrillary tangles, cerebral amyloid angiopathy, TDP-43

# Main

Late-onset amnestic dementia diagnosed as Alzheimer’s Disease (AD/LOAD) is increasingly recognized as a heterogeneous clinical syndrome that can reflect multiple underlying independent or co-occurring pathological processes that may have independent or pleiotropic genetic risk factors.[1](#ref-farfel2016)–[3](#ref-karanth2020) Because of the complex nature of amnestic dementia and the practical challenge of constructing large cohorts with harmonized pathology data, LOAD genome-wide association studies (GWAS) to date have primarily utilized clinical diagnosis or proxy phenotypes based on family history of dementia to help unravel the genetic etiology of disease.[4](#ref-jansen2019)–[7](#ref-kunkle2019) While these LOAD GWAS have been immensely successful at identifying dementia-associated genetic loci, with over 80 such loci identified thus far,[6](#ref-bellenguez2022) the clinical phenotypes used in most studies complicate the interpretation of their results and belies a more complex reality of mixed neuropathologies being the norm in aged individuals.[3](#ref-karanth2020),[8](#ref-escott-price2022) To build on the successes of previous LOAD GWAS, and to begin decrypting the complexities of this heterogeneous clinical syndrome, we believe GWAS using neuropathology endophenotypes (NPE) is an essential next step–not only to broadly understand LOAD’s genetic etiology, but to identify alleles that are drive specific LOAD-associated pathologies.[9](#ref-karran2022)

Neuritic plaques and neurofibrillary tangles (NFT) are present in the brains of the majority of people diagnosed with clinical LOAD, but ~20% of diagnosed do not have substantial AD neuropathologic change (ADNC), and another ~50% have both ADNC and comorbid non-AD pathologies.[3](#ref-karanth2020),[10](#ref-nelson2009) For example, TAR DNA-binding protein 43-kDa (TDP-43) pathology is found in 40-50% of autopsied individuals with dementia and contributes to an amnestic syndrome, limbic-predominant age-related TDP-43 encephalopathy (LATE), with a characteristic pattern of neuropathological change called LATE-NC.[3](#ref-karanth2020),[11](#ref-nelson2019),[12](#ref-nelson2022) Hippocampal sclerosis (HS) of aging, a pathology characterized by neuronal death, gliosis, and atrophy of the hippocampus beyond normal ranges based on levels of AD pathology, commonly co-occurs with LATE-NC and is associated with severe cognitive impairment.[11](#ref-nelson2019),[13](#ref-brenowitz2014) It is already known that the genetic risk factors of ADNC and LATE-NC differ,[14](#ref-dugan2021) but the genetic underpinnings of common comorbid (ADNC+LATE-NC) phenotypes are not well understood.

Cerebrovascular pathologies, which contribute to cognitive decline and dementia, are also common in elderly autopsied research participants.[15](#ref-skrobot2016) Cerebral amyloid angiopathy (CAA) is a cerebrovascular pathology characterized by amyloid-beta deposition in cerebral blood vessels.[16](#ref-weber2018) CAA often co-occurs with LOAD but can independently contribute to cerebral injury by causing hemorrhages in the brain parenchyma.[15](#ref-skrobot2016),[16](#ref-weber2018) Infarcts of both small but grossly visible lacunar arteries and microscopically examined vessels (the latter referred to as microinfarcts) are also common in aged individuals and contribute to cognitive decline.[15](#ref-skrobot2016),[17](#ref-smith2012) Cerebral atherosclerosis and sclerosis of small cerebral blood vessels, called brain arteriolosclerosis, predispose individuals to infarcts and HS[18](#ref-arvanitakis2017),[19](#ref-neltner2014); moreover, they contribute to cognitive decline even after adjusting for the presence of related pathologies.[20](#ref-arvanitakis2016),[21](#ref-ighodaro2017) Collectively, these factors reveal an increasingly complex and synergistic web of pathologies which contribute to cognitive impairment and dementia and further demonstrate the need to understand how existing (and potentially new) GWAS loci are involved in LOAD-associated pathologies.

Studying the genomic risk factors of individual underlying neuropathologies can provide a complementary approach to large GWAS of clinical- and family history-based outcomes for the study of AD and dementia risk. As proof of principle, prior GWAS of neuropathology endophenotypes (NPE) have confirmed known LOAD risk loci and have identified potential new neuropathology risk loci.[22](#ref-beecham2014)–[26](#ref-reddy2021) Some NPE, particularly LATE-NC, have yet to be studied using GWAS; statistical methodological development for GWAS and downstream functional analyses along with continued recruitment in autopsy cohorts provide additional opportunity for in-depth investigation into multiple neuropathologies. Here, we perform GWAS on eleven neuropathology endophenotypes in the largest harmonized cohort, to date, using three high-quality data sources with autopsy and genotype data. For previously studied NPE, the sample sizes used here represent 50-150% increases relative to those used in prior studies. We also perform downstream functional analyses to explore potential biological functional mechanisms of newly identified risk loci and provide insight into known AD risk loci.

# Results

## Participants and distributions of neuropathology endophenotypes

We used genotype and neuropathology data from three autopsy data sources: (1) the National Alzheimer’s Coordinating Center (NACC, N = 5,940), (2) the Religious Orders Study and the Rush Memory and Aging Project (ROSMAP, N = 1,183), and (3) the Adult Changes in Thought (ACT, N = 681) study. In total, 7,804 unique participants were included in our analyses. NACC participants were more likely to be diagnosed with either dementia (83%) or AD/mild cognitive impairment (MCI, 77.5%) than ROSMAP (55% and 66%) or ACT (13% and 10%) participants. NACC participants died at a younger age (mean age at death 81 years) compared to ROSMAP (90 years) and ACT (89 years) participants and had more balanced participation between the sexes, with 50% of NACC participants being female vs. 68% and 58% in ROSMAP and ACT, respectively. NACC participants were also more likely to have an *APOE* allele (54%) vs. ROSMAP (25%) or ACT (28%; Chi-square test for all comparisons). These differences reflect heterogeneity in the recruitment for these studies. NACC is based on data collected from over 30 Alzheimer’s Disease Research Centers which often recruit from clinic patients and their families. In contrast, ACT originally recruited aged participants in the Seattle, WA area who did not have dementia at time of enrollment, and ROS recruited members of the Catholic Church clergy.

Available sample sizes differed among the eleven studied NPE, which included the AD-related pathologies of Consortium to Establish a Registry for Alzheimer’s Disease (CERAD) score for neuritic plaques ($N = 7,786, Braak staging of NFT (N = 7,776), and diffuse amyloid-beta plaques; the non-AD proteinopathies LATE-NC (N = 3,112) and Lewy body pathology (N = nlewy); hippocampal sclerosis (N = 7,164); and the cerebrovascular pathologies CAA (N = 7,381), gross infarcts (N = 7,398), microinfarcts (N = 7,480), atherosclerosis (N = 7,340), and arteriolosclerosis (N = 6,668). Relative to the Beecham et al. (2014) GWAS of neuropathology endophenotypes, our updated sample represents an overall 52% increase in sample size.[22](#ref-beecham2014) The sample size increase for individual neuropathologies is greater, ranging from 57% for Braak stage to 150% for CAA, demonstrating both an increased total number of autopsy participants and more complete neuropathological profiling for each participant. HS, microinfarcts, and gross infarcts were coded as binary case/control outcomes. All other phenotypes were ordinal variables with either four or seven levels (Methods and Supplementary Note describe our phenotype definitions and harmonization approach).

To explore the co-occurrence of neuropathology endophenotypes in our pooled data, we estimated polychoric correlations (an approach which assumes that observed ordinal or binary variables reflect latent normally distributed variables) between neuropathology endophenotype pairs followed by hierarchical clustering using the polycor, psych, and pheatmap R packages.[27](#ref-revelle2022)–[29](#ref-kolde2019) Inspection of the resulting dendrogram identified three positively correlated clusters of endophenotypes: a “vascular” cluster consisting of gross infarcts, microinfarcts, arteriolosclerosis, and atherosclerosis; an “Alzheimer’s disease” cluster consisting of Braak NFT stage, neuritic plaques, diffuse amyloid-beta plaques, and CAA; and a “LATE” cluster consisting of LATE-NC and HS (Supplementary Figure 1).

## GWAS meta-analysis of eleven neuropathology endophenotypes (NPE)

An overview of our study design is shown in **Figure 1**. Genetic association analyses were performed with logistic or proportional-odds logistic regression mixed-effects models as appropriate (Methods). We performed GWAS on 11 NPE using NACC participants, ROSMAP, and ACT separately. We then performed fixed-effect meta-analysis using METAL on variants with minor allele frequencies in each data source.[30](#ref-metal:f)

In total, 13 NPE-genomic locus associations had lead variants that met the genome-wide significance threshold of . Three NPE (Braak stage, CERAD score, and CAA) were associated with two or more independent loci () in the *APOE* region, though we report only one locus in the *APOE* region for each phenotype unless secondary loci remained significant after conditioning on *APOE* diplotype (see conditional *APOE*-adjusted analysis below). The lead variant rs429358 in the *APOE* region was significantly associated with four NPE: Braak stage (, ), CAA (, ), CERAD score (, ), and LATE-NC ( , ). Variants within *TMEM106B* were associated with both HS (rs7805419; , ) and LATE-NC (2043539; , ). A *GRN* locus was associated with HS (rs5848; , ). One *BIN1* locus was associated with Braak stage (rs6733839 , ).

Novel associations identified in the NPE GWAS meta-analysis included a variant 12 kbp upstream of *COL4A1* associated with cerebral atherosclerosis (rs2000660; , ), a *PIK3R5* intronic locus associated with Braak stage (rs72844606 , ), and an intronic *LZTS1* locus associated with arteriolosclerosis (rs78909048; , ). One intronic locus of *SPATA48* was close to genome-wide signicantly associated with atherosclerosis (rs62447817; , )

## Conditional *APOE*-adjusted analysis

To check whether any loci in the *APOE* region (defined as less than 500 kbp from the start or end site of *APOE* transcription) were associated with NPE risk independently of the known effects of *APOE* alleles, we re-analyzed the *APOE* region of Chromosome 19 while adjusting for *APOE* diplotypes. We limited this conditional analysis to NPE with which variants within the *APOE* region met the genome-wide significance threshold of in Stage 3 (Braak NFT stage, neuritic plaque CERAD score, diffuse amyloid-beta plaques, cerebral amyloid angiopathy, and LATE-NC). One locus between *APOC4-APOC2* with lead variant rs7247551 remained significantly associated with CAA after adjusting for *APOE* diplotypes (, , Table 1 and Figure 3a). No variants remained genome-wide significantly associated with any other APOE-associated NPE, and the P-values of the lead variant rs429358 from genome-wide meta-analysis unadjusted for APOE were all above 0.05 in the re-analysis. Using independent data from a recent GWAS of CAA in 815 participants with dementia in the Mayo Clinic Brain Bank, we replicated the association of rs7247551 with CAA while adjusting for APOE diplotype (, ). These results provide evidence for a novel locus in the *APOE* region other than the known *APOE* haplotypes that affects CAA pathology burden, and that the genetic risk basis of CAA in the *APOE* region may differ from that AD neuropathologies of neuritic plaques and NFT.

| Table 1: Significant NPE-Associated Loci in Stage 3 pooled GWAS | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Phenotype | Genea | Variant | Chromosome | Positionb | Minor/major allele | ORc | 95% CI | P |
| Braak NFT Stage | *BIN1* | rs6733839 | 2 | 127,135,234 | T/C | 1.21 | 1.21 [1.14-1.29] | 1.6e-09 |
| LATE-NC | *TMEM106B* | rs2043539 | 7 | 12,214,254 | A/G | 0.70 | 0.7 [0.63-0.78] | 5.8e-11 |
| HS | *TMEM106B* | rs7805419 | 7 | 12,242,825 | C/T | 0.65 | 0.65 [0.58-0.73] | 3.2e-13 |
| Atherosclerosis | *SPATA48* | rs62447817 | 7 | 50,096,036 | A/G | 1.35 | 1.35 [1.21-1.5] | 5.5e-08 |
| Arteriolosclerosis | *LZTS1* | rs78909048 | 8 | 20,279,428 | G/A | 0.44 | 0.44 [0.34-0.57] | 5.7e-10 |
| Atherosclerosis | *COL4A1* | rs2000660 | 13 | 110,136,094 | A/G | 0.73 | 0.73 [0.66-0.82] | 2.7e-08 |
| Braak NFT Stage | *PIK3R5* | rs72844606 | 17 | 8,930,274 | T/C | 0.69 | 0.69 [0.6-0.79] | 4e-08 |
| HS | *GRN* | rs5848 | 17 | 44,352,876 | T/C | 1.40 | 1.4 [1.24-1.57] | 3.2e-08 |
| Braak NFT Stage | *APOE* | rs429358 | 19 | 44,908,684 | C/T | 2.06 | 2.06 [1.92-2.21] | 9.7e-89 |
| CAA | *APOE* | rs429358 | 19 | 44,908,684 | C/T | 2.49 | 2.49 [2.32-2.67] | 4.4e-138 |
| CERAD Score | *APOE* | rs429358 | 19 | 44,908,684 | C/T | 2.42 | 2.42 [2.23-2.62] | 4.7e-103 |
| LATE-NC | *APOE* | rs429358 | 19 | 44,908,684 | C/T | 1.70 | 1.7 [1.48-1.95] | 1.7e-14 |
| CAAd | *APOC4-APOC2/APOC2* | rs7247551 | 19 | 44,951,509 | G/A | 0.81 | 0.81 [0.76-0.86] | 8e-12 |
| aClosest protein-coding gene according to GENCODE release 40. bGenome positions are based on build HG38. cORs are with respect to minor allele. dResult from APOE diplotype-adjusted analysis. | | | | | | | | |

## Associations of clinical and proxy AD risk loci with NPE

We checked whether AD-associated loci identified in the recent Bellenguez et al. (2022) GWAS of AD and related dementia (ADD) were associated with NPE.[6](#ref-bellenguez2022) Bellenguez et al. identified a total of 83 non-*APOE* independent loci (39 previously identified, 44 novel) associated with ADD (hereafter, “ADD loci”), 80 of which had lead variants that were included in our study. Each of the 11 NPE used as outcomes in the present study was checked for association with variants in these 80 ADD loci. Benjamini-Hochberg adjustments were then applied to *P*-values for each NPE separately to control the false-discovery rate (FDR). In total, 22 NPE-locus associations had adjusted *P*-values (*Q*-values) and were considered significant; of these 22 associations, 21 had concordant directions of effect with Bellenguez et al. (Table 2 and Supplementary Table 5). The pathognomonic AD pathologies, Braak stage and CERAD, had 70 and 60 of 80 concordant directions of effect with ADD loci, respectively, regardless of statistical significance. Three ADD loci (*BIN1, MME,* and *HLA*) were significantly associated with Braak NFT stage after multiple test correction. Ten AD loci (*CR1*, *BIN1, INPP5D,* ZCWPW1/NYAP1*, PTK2B, PICALM*, *SORL1*, *FERMT2*, *SNX1,* andABCA7) were significantly associated with CERAD score after multiple test correction. Four AD loci (*TMEM106B, IL34, GRN*, and *MAPT*) were significantly associated with HS, all of which except for *L34* (HS , ; AD , ) were concordant in effect direction. Two ADD loci (*TMEM106B* and *GRN*) were significantly associated with LATE-NC, both concordant. Finally, one ADD locus (*PLCG2*) was significantly associated with microinfarcts. These results indicate that NPE studies largely corroborate the findings of large AD GWAS based on clinical and proxy phenotypes. However, several ADD loci, particularly *TMEM106B* and *GRN*, are associated with non-AD pathology (HS and LATE-NC) but not neuritic plaques or NFT.

| Table 2: Associations between NPE and known ADD loci | | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| NPE | Chromosome | Locusa | Positionb | Variant | EAFc | Effect/Other Allele | NPE Beta | NPE OR | NPE P-value | NPE Q-value | ADD Beta | ADD OR | ADD P-value | NPE-ADD Concordant Effect Direction |
| Braak NFT Stage | 2 | *BIN1* | 127,135,234 | rs6733839 | 0.42 | T/C | 0.19 | 1.21 | 1.6e-09 | 1.3e-07 | 0.17 | 1.18 | 6.5e-90 | Yes |
| Braak NFT Stage | 3 | *MME* | 155,069,722 | rs16824536 | 0.05 | A/G | -0.28 | 0.75 | 0.00016 | 0.0066 | -0.09 | 0.92 | 3.8e-06 | Yes |
| Braak NFT Stage | 6 | *HLA* | 32,615,322 | rs6605556 | 0.84 | A/G | 0.16 | 1.17 | 0.00024 | 0.0066 | 0.10 | 1.10 | 1e-17 | Yes |
| CAA | 11 | *PICALM* | 86,157,598 | rs3851179 | 0.35 | T/C | -0.11 | 0.89 | 0.00087 | 0.042 | -0.11 | 0.90 | 6.5e-36 | Yes |
| CAA | 12 | *TPCN1* | 113,281,983 | rs6489896 | 0.93 | T/C | -0.20 | 0.82 | 0.001 | 0.042 | -0.07 | 0.93 | 2.5e-06 | Yes |
| CERAD Score | 1 | *CR1* | 207,577,223 | rs679515 | 0.20 | T/C | 0.13 | 1.14 | 0.0018 | 0.025 | 0.12 | 1.13 | 5.2e-33 | Yes |
| CERAD Score | 2 | *BIN1* | 127,135,234 | rs6733839 | 0.41 | T/C | 0.18 | 1.19 | 3.2e-07 | 2.6e-05 | 0.17 | 1.18 | 6.5e-90 | Yes |
| CERAD Score | 2 | *INPP5D* | 233,117,202 | rs10933431 | 0.78 | C/G | 0.13 | 1.14 | 0.0017 | 0.025 | 0.09 | 1.09 | 1e-17 | Yes |
| CERAD Score | 7 | *ZCWPW1/NYAP1* | 100,334,426 | rs7384878 | 0.71 | T/C | 0.11 | 1.12 | 0.0032 | 0.032 | 0.08 | 1.08 | 2.1e-18 | Yes |
| CERAD Score | 8 | *PTK2B* | 27,362,470 | rs73223431 | 0.37 | T/C | 0.11 | 1.12 | 0.0012 | 0.024 | 0.07 | 1.07 | 5.3e-15 | Yes |
| CERAD Score | 11 | *PICALM* | 86,157,598 | rs3851179 | 0.35 | T/C | -0.17 | 0.84 | 1.1e-06 | 4.4e-05 | -0.11 | 0.90 | 6.5e-36 | Yes |
| CERAD Score | 11 | *SORL1* | 121,564,878 | rs11218343 | 0.96 | T/C | 0.25 | 1.28 | 0.0057 | 0.046 | 0.17 | 1.18 | 1e-14 | Yes |
| CERAD Score | 14 | *FERMT2* | 52,924,962 | rs17125924 | 0.91 | A/G | -0.17 | 0.84 | 0.0038 | 0.035 | -0.09 | 0.92 | 5.8e-10 | Yes |
| CERAD Score | 15 | *SNX1* | 64,131,307 | rs3848143 | 0.78 | A/G | -0.12 | 0.89 | 0.0031 | 0.032 | -0.05 | 0.95 | 1.1e-06 | Yes |
| CERAD Score | 19 | *ABCA7* | 1,050,875 | rs12151021 | 0.33 | A/G | 0.14 | 1.15 | 0.0005 | 0.014 | 0.11 | 1.11 | 4.1e-30 | Yes |
| HS | 7 | *TMEM106B* | 12,229,967 | rs13237518 | 0.42 | A/C | -0.42 | 0.66 | 6e-11 | 4.8e-09 | -0.04 | 0.96 | 5.1e-07 | Yes |
| HS | 16 | *IL34* | 70,660,097 | rs4985556 | 0.11 | A/C | -0.30 | 0.74 | 0.0019 | 0.039 | 0.06 | 1.06 | 5.6e-06 | No |
| HS | 17 | *GRN* | 44,352,876 | rs5848 | 0.30 | T/C | 0.33 | 1.40 | 3.2e-08 | 1.3e-06 | 0.06 | 1.07 | 1.8e-12 | Yes |
| HS | 17 | *MAPT* | 46,779,275 | rs199515 | 0.79 | C/G | 0.26 | 1.30 | 0.00034 | 0.0092 | 0.06 | 1.06 | 6e-09 | Yes |
| LATE-NC | 7 | *TMEM106B* | 12,229,967 | rs13237518 | 0.43 | A/C | -0.38 | 0.69 | 6.3e-08 | 5.1e-06 | -0.04 | 0.96 | 5.1e-07 | Yes |
| LATE-NC | 17 | *GRN* | 44,352,876 | rs5848 | 0.30 | T/C | 0.28 | 1.32 | 1.3e-06 | 5.2e-05 | 0.06 | 1.07 | 1.8e-12 | Yes |
| Microinfarct | 16 | *PLCG2* | 81,739,398 | rs12446759 | 0.61 | A/G | 0.14 | 1.15 | 0.00055 | 0.044 | 0.06 | 1.06 | 3.6e-12 | Yes |
| ADD = Alzheimer's Disease and related dementias; EAF = Effect allele frequency. NPE P-values, betas, and OR are from meta-analysis. NPE Q-values are produced by applying Benjamini-Hochberg adjustments for each endophenotype separately. ADD P-values, betas, and OR are from Bellenguez et al (2022) stage I GWAS (N = 487,511). OR are with respect to the Bellenguez effect allele. aEither known locus or closest protein-coding gene according to GENCODE release 40. bPosition of lead variant using GRCh38 assembly. cEffect allele frequency in NPE meta-analysis. | | | | | | | | | | | | | | |

## Gene-prioritization and enrichment analyses

We used gene-based, pathway, and enrichment analyses implemented in FUMA to identify potential functional genes and biological pathways involved in neuropathology development (results available in Supplementary Data zip file).[31](#ref-watanabe2017) We used a minimum threshold of association of in the Stage 3 analysis for variant annotation. Gene-level analyses implemented in MAGMA corroborated our single-variant analyses,[32](#ref-leeuw2015) identifying gene associations with neuropathology endophenotypes in regions with significant variant-phenotype associations. *APOE* was significantly associated with NFT, diffuse plaques, CAA, neuritic plaques, and LATE-NC. *TMEM106B* was associated with both HS and LATE-NC. Gene-set enrichment analyses were performed using MAGMA and Gene Ontology (GO) and curated gene sets from MSigDB.[33](#ref-liberzon2011) After Bonferroni correction genes involved in polysaccharide digestion were significantly associated with microinfarcts (). We also found evidence that genes involved in substrate recognition for ubiquination were associated with cerebral atherosclerosis (), and genes overexpressed in developing B cells were associated with Braak NFT stage (). FUMA variant annotation using CADD score, RegulomeDB score, eQTL, and chromatin interaction annotations indicated that the rs2000660 lead variant associated with cerebral atherosclerosis in Stage 3 is located within an enhancer region located 13 Kbp upstream of *COL4A1* transcription start site.

## Colocalization analysis

We investigated whether loci associated with multiple NPE colocalized using a Bayesian colocalization analysis approach implemented in the coloc R package.[34](#ref-giambartolomei2014) Because CAA was the only NPE with an associated locus in the *APOE* region with a signal independent of *APOE* diplotype, we excluded this region from NPE-NPE colocalization analysis. Pairs of NPE with the same variant reaching a P-value threshold of with concordant effect directions included (1) Braak NFT stage and CERAD score in the *BIN1* region on Chromosome 2 and (2) HS and LATE-NC in the *TMEM106B* region on Chromosome 7 (**Figure 4**). Both loci exhibited high posterior probability of colocalization (), indicating that the same genetic variants drive association with both endophenotypes.

A locus within *GRN* associated with HS () strongly colocalized with *GRN* expression in multiple tissues (). The *TMEM106B* locus associated with HS and LATE-NC colocalized with *TMEM106B* expression in multiple tissues, including the cerebellar hemisphere (). Two CpG sites located either within *TMEM106B* (cg09613507) or downstream (cg23422036) colocalized with both HS and late (cg09613507-HS , cg09613507-LATE-NC , cg23422036-HS , cg23422036-LATE-NC ).

The Chromosome 19 CAA-associated locus identified in the *APOE*-adjusted analysis colocalized with expression of multiple genes in GTEx: *APOE* expression in the wall of the aorta; *8CLPTM1* expression in the skin of the leg and suprapubic region; and *APOC2* expression in 14 different tissues, including the brain cortex, caudate, nucleus accumbens, and cerebellum (). The same locus colocalized with mQTL for four CpG sites in ROSMAP, cg09555818, cg04401876, cg10169327, and cg13119609 (, see **Figure 3b**).

Multiple suggestive NPE loci showed evidence of colocalization with eQTL in GTEx. In total, 44 NPE loci (lead variant ) colocalized with QTL with , with 533 total colocalizing NPE-QTL pairs across 44 tissues (Supplementary Table S4).

## Multiple eQTL and mQTL colocalizing with NPE are validated with methylation and RNA-seq in ROSMAP

Using ROSMAP participants with available DNA methylation and/or RNA-Seq data available from the DLPFC, we further investigated associations between QTL phenotypes found to colocalize with NPE in brain tissues of GTEx and/or ROSMAP. We tested methylation levels at each of the four CpG sites – cg09555818, cg04401876, cg10169327, and cg13119609 – that colocalized with the Chromosome 19 CAA risk locus for association with to confirm association with lead variant rs7247551. We first confirmed that all four CpG sites were significantly associated () with rs7247551 with directions of effect consistent with those previously reported for ROSMAP.[35](#ref-ng2017) We then tested for association between methylation at these sites with CAA pathology. Hypomethylation at cg09555818 (odds ratio , ) and cg13119609 (, ) were significantly associated with worse CAA pathology (Figure 3C).

Next, as *APOC2* expression in multiple brain tissues colocalized with CAA in GTEx but not ROSMAP, we investigated to see whether there was a nominal association between *APOC2* expression in the DLPFC and rs7247551. We found that the G allele of rs7247551 was significantly associated with increased *APOC2* expression in the DLPFC (, , Figure 3D); however, the direction of effect was opposite of that found in brain tissues in GTEx (i.e., the G allele of rs7247551 was associated with decreased *APOC2* expression in GTEx, Figure 3E). Expression of *APOC2* in the DLPFC was not associated with CAA in ROSMAP (, ). We performed an additional *post-hoc* analysis for nominal *APOE* eQTL activity of rs7247551 in ROSMAP (rs7247551 was lead eQTL in colocalizing *APOE* expression loci in the aorta, thyroid, and prostate, though APOE expression in no brain tissues colocalized with CAA). We confirmed that rs7247551 was not associated with *APOE* expression in the DLPFC in ROSMAP (, Figure 3D).

We then investigated the associations betwen HS and LATE-NC with the *GRN* and *TMEM106B* expression and methylation in the DLPFC. We first confirmed that the lead *TMEM106B* variants for HS (rs7805419) and LATE-NC (rs2043539) were in LD () and that rs2043539 was associated with DNA methylation levels at cg09613507 and cg23422036 as previously reported. We also found that Neither TMEM106B nor GR*N* expression were associated with HS (), while decreased *TMEM106B* expression was associated with more severe LATE-NC pathology (). Of the two CpG sites that colocalized with HS and LATE-NC, hypermethylation of cg09613507 was associated with more severe LATE-NC pathology ().

# Discussion

The present study of eleven neuropathology endophenotypes with a maximum sample size of 7,786 participants serves as an autopsy-based complement to previous studies, expanding on the findings of previous genetic association studies of NPE, Alzheimer’s, and related dementias to attempt to better understand the complex associations between different neuropathologies and genetic risk factors.[2](#ref-katsumata2022),[6](#ref-bellenguez2022),[14](#ref-dugan2021),[22](#ref-beecham2014)–[24](#ref-vattathil2021),[26](#ref-reddy2021) We identified a novel *APOE* -independent CAA risk locus that also affects *APOC2* brain expression and dementia-associated CpG methylation sites. Two of these CpG sites were in turn also associated with CAA risk. We also discovered intriguing new loci mapped to *COL4A1*, *PIK3R5*, and *LZTS1* associated with atherosclerosis in the circle of Willis, Braak stage for neurofibrillary tangles, and brain arteriolosclerosis, respectively. Lastly, we investigated known loci in *BIN1*, *APOE*, and *TMEM106B* to provide additional context on their associations with multiple NPE.

Multiple variants in the *APOE* region were associated with pathognomonic AD NPE, including Braak stage, CERAD score, diffuse amyloid plaques, and CAA. Variants in the *APOE* region were also associated with LATE-NC, which is consistent with previous genetic association studies of NPE.[14](#ref-dugan2021),[22](#ref-beecham2014) Notably, *APOE* variants were not significantly associated with vascular pathology except for CAA. When we re-ran Chromosome 19 analyses on these phenotypes while adjusting for *APOE* genotype, only one CAA-associated locus with lead variant rs7247551 remained significant (, ). We confirmed that this locus was indeed novel and not in LD with a recent study of CAA in *APOE* carriers by Reddy et al. (2021) by confirming that between rs7247551 and Reddy et al. lead variant rs5117.[26](#ref-reddy2021) Futhermore, we were able to replicate the associated between rs7247551 and CAA pathology while adjusting for *APOE* diplotype in 815 Mayo Clinic Brain Bank participants used in Reddy et al. but not used in the present CAA GWAS (, ), providing additional evidence for the APOE locus being important for CAA pathology beyond the known effects of APOE haplotypes.

Several variants in the novel CAA locus were lead eQTL for *APOC2* and brain expression in GTEx Colocalization analysis confirmed that the CAA risk locus shares a functional variant with both *APOC2* eQTL ( in brain tissues) and nearby mQTL (). This locus also colocalized with *APOE* expression in the wall of the aorta () and prostate () but not in any brain tissues. Furthermore, a lead QTL in this locus, rs4803779, was associated with *APOC2* expression but not *APOE* expression in the DLPFC in ROSMAP.

Variants in the same locus were also mQTL in the DLPFC in ROSMAP for four methylation sites (rs4803779 for top site cg09555818; Figure 3). We confirmed that two of the CpG sites affected by the CAA risk locus, cg09555818 (, ) and cg13119609 (, ), were in turn significantly associated with CAA pathology. Both of these CpG sites are located within the *APOC4-APOC2* readthrough transcript region overlapping *APOC4* and *APOC2*. Notably, we found that methylation levels at cg09555818 and cg13119609 are highly positively correlated (, ) in ROSMAP. Our results are consistent with the hypothesis that rs7247551 locus affects CAA risk through hypomethylation of CpG sites in the *APOC4-APOC2* region. These results are also consistent with previous studies in other human cohorts that implicate hypomethylation at cg09555818 and cg13119609 as potentially associated with Alzheimer’s disease.[36](#ref-walker2021)–[38](#ref-shao2018) Collectively, these results provide circumstantial evidence indicates that *APOC2* may be the target gene of the rs7247551 CAA risk locus as it colocalizes with *APOC2* expression in multiple brain tissues in GTEx and the associated CpG sites are located in exon 3 of *APOC4-APOC2*. However, more research must be done to verify that *APOC2* is the target gene.

One intronic locus of *TMEM106B* was significantly associated with both HS (, ) and LATE-NC (, ). Additionally, a locus within *GRN* was associated with HS (, ). Both of these genes are associated with frontotemporal lobar degeneration,[39](#ref-rollinson2011),[40](#ref-ciani2019) HS,[41](#ref-katsumata2017),[42](#ref-nelson2015) and ADD.[5](#ref-wightman2021),[6](#ref-bellenguez2022) We found that HS, LATE-NC, and ADD all colocalize at these two loci (, Figure 5A). These results indicate that HS, LATE-NC, and LOAD likely share causal loci for these genes. When NPE were regressed on known AD variants, Braak NFT stage was significantly associated with *TMEM106B* with a direction of effect opposite of both recent AD GWAS and other neuropathologies,[5](#ref-wightman2021),[6](#ref-bellenguez2022) while neuritic plaques were not associated with *TMEM106B* in our study. Evidence from RNA-Seq analyses supported the hypothesis that *TMEM106B* may have discordant effects on TDP-43 and NFT pathology, and both the genetic and gene expression associations between *TMEM106B* and NFT were strengthened when adjusting for LATE-NC stage. While preliminary, these results support the possibility that the associations between *TMEM106B* and AD found in recent GWAS are driven by non-AD pathologies, namely TDP-43, in dementia cases, and may mask a more complex association between *TMEM106B* and NFT.

A locus approximately 30 kbp downstream of *BIN1* on Chromosome 2q14 was significantly associated with Braak stage (, ) and suggestively associated with CERAD score for neuritic plaques (, ). We verified through colocalization analysis that the same locus drives association signals with Braak NFT stage and CERAD score. In prior GWAS, this locus is second only to *APOE* for strength of association with LOAD.[6](#ref-bellenguez2022) Interestingly, the lead variant in this locus, rs6733839, was associated with neither diffuse amyloid plaques (, ) nor CAA (, ). Previous research supports the hypothesis that *BIN1* is associated with LOAD through its effect on NFT rather than amyloid pathology.[43](#ref-holler2014),[44](#ref-franzmeier2019) As neuritic plaques contain dying nerve cell processes with aberrant tau fibrils identical to those seen in neurofibrillary tangles,[10](#ref-nelson2009) our findings are also consistent with the hypothesis that *BIN1* influences AD risk primarily through tau rather than amyloid pathogenic processes.

One locus on Chromosome 12q34 with lead variant rs2000660 (minor allele frequency in pooled data set) located 12 kbp upstream of *COL4A1* was significantly associated with atherosclerosis in the circle of Willis (, ). In previous GWAS, researchers have reported the *COL4A1/COL4A2* locus to be associated with numerous other vascular phenotypes, including peripheral artery disease, coronary artery disease, stroke, and arteriolar stiffness.[47](#ref-steffensen2018) The *COL4A1/COL4A2* locus has also been found to be associated with rare familial cerebrovascular diseases and lacunar ischemic stroke[48](#ref-blevins2021),[49](#ref-rannikmae2017) In a recent GWAS, rs2000660 in particular was a lead risk variant for migraines.[50](#ref-hautakangas2022) The relevance of the *COL4A1* locus to cerebral vascular traits is thus highly supported by previous research, and the biological role of collagen IV in vascular disease is possibly related to disruption of the extracellular matrix.[47](#ref-steffensen2018) *COL4A1* in the brain is preferentially expressed in endothelial cells (Figure 1**)** and codes for a component of collagen IV, an important component of basal lamina. rs2000660 was not nominally associated with any other vascular NPE in our study, and a previous GWAS of cerebral atherosclerosis using ROSMAP participants did not identify the *COL4A1* as a risk locus.[24](#ref-vattathil2021) However, the sample size of the previous study was significantly smaller than the one used here (1,325 vs 7,340). Indeed, in the present study, rs2000660 reached genome-wide significance only in the pooled Stage 3 analysis, though it was nominally significant in the ROSMAP-only replication analysis in Stage 2 (). rs2000660 was not a QTL in GTEx or ROSMAP. However, rs650724, a variant in high LD with rs2000660 (), is a synonymous coding variant (p.Ser1600Ser in ENST00000375820.10; p.Ser319Ser in ENST00000650424.1) for *COL4A1*, presenting a possible functional variant driving association in this locus.

Another novel genome-wide significant locus located on Chromosome 17p13 was associated with Braak NFT stage and had a lead variant of rs72807981 (, , ), an intronic variant within *PIK3R5*, which codes for a phosphatidylinositol 3-kinase involved in cell growth, motility, and survival. This variant was suggestively associated with Braak NFT stage in NACC (, ) and was nominally replicated in ACT (, ). The same variant was also nominally associated with neuritic plaques in the pooled analysis (, ) but was not nominally associated with any other AD NPE. rs72807981 is not a QTL in GTEx or ROSMAP, nor were any of the variants in high LD with it in our study. One variant in LD () had a CADD score of 11.65, indicating a potential functional variant. There is previous research suggesting that *PIK3R5* is more highly expressed in aged adults with Braak NFT stages V and VI vs. non-demented controls (false-discovery rate ).[51](#ref-guennewig2021) *PIK3R5* is expressed preferentially in microglial cells in humans (Figure 3F),[52](#ref-zhang2016) suggesting that its association with neurofibrillary pathology may be immune-mediated.

Add discussion of novel arteriolosclerosis locus.

In conclusion, we identified several promising novel loci associated with NPE and replicated multiple known risk loci for AD using NPE. We also provided additional context and consideration for relationships between specific risk loci and different NPE. Our study demonstrates the importance of studying genetic risk factors of neuropathology endophenotypes as a complement to studies of clinical and proxy phenotypes of Alzheimer’s disease.

# Methods

## Participants

**NACC:** The present study used National Alzheimer’s Coordinating Center (NACC) data from 36 National Institute on Aging-funded Alzheimer’s Disease Research Centers (ADRCs). Individual ADRCs use different recruitment strategies and perform autopsies on-site, but neuropathology data at each ADRC are collected using a standard form (<https://files.alz.washington.edu/documentation/np11-form.pdf>) and submitted to NACC where they are aggregated and anonymized. The NACC Neuropathology (NP) data set based on the first version of this form was originally implemented in 2001,[53](#ref-besser2018) and this analysis uses data from then through the December 2021 freeze. Participants were excluded if they did not have autopsy data available or if they were noted in the NP data set to have at least one of 19 conditions that could potentially bias results. These conditions include brain tumors, severe head trauma, and fronto-temporal lobar degeneration. (See **Supplementary Table S1** for full list of variables used for exclusion criteria.)

**ROSMAP:** ROSMAP consists of harmonized data from two longitudinal cohort studies: The Religious Orders Study (ROS) and the Rush Memory and Aging Project (MAP).[54](#ref-bennett2018) ROS and MAP were both approved by an Institutional Review Board of Rush University Medical Center. All participants signed an Anatomic Gift Act, as well as informed and repository consents. ROS began in 1994 and has recruited over 1500 Catholic priests, nuns, and brothers across the United States. MAP started in 1997 and has enrolled more than 2300 community members in the greater Chicago area of northeastern Illinois. The ROSMAP NP data used in this study were received from Rush University Medical Center in January 2020.

**ACT:** The Adult Changes in Thought (ACT) study began in 1994 and recruited residents in the greater Seattle area aged 65 years and older without dementia at time of enrollment.[55](#ref-kukull2002) The study has expanded to include three cohorts with continued enrollment using the original enrollment criteria and has a current total of 4,960 participants across all three cohorts. The ACT NP data used in this study were obtained from Kaiser Permanente in May 2021.

**ADNI:** The Alzheimer’s Disease Neuroimaging Initiative (ADNI; <https://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, positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment and early AD. A subset of ADNI participants undergo autopsy and receive neuropathological phenotyping. The ADNI NP data used in this study was downloaded from the ADNI website in October 2021.

Some participants in ROSMAP and ADNI also had neuropathology and genotype data available in NACC. In these cases, records in the NACC were preferentially kept to maximize sample size of the initial Stage 1 analysis using only NACC participants (see Statistical Analyses subsection). An overview of our study design is presented in Figure 1.

All study participants were deceased and the resulting data were de-identified for University of Kentucky investigators, and we exclusively used archival samples. Therefore, our study does 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, or/or recorded such that subjects cannot be identified.”

## Genotype data and quality control

Genotype data for all cohorts underwent imputation using the Trans-’Omics for Precision Medicine (TOPMed) Imputation Server and the TOPMed reference panel.[56](#ref-taliun2021) NACC and ACT raw genotype data were obtained from the September 2020 freeze of the Alzheimer’s Disease Genetics Consortium (ADGC) and subsequently imputed, while pre-imputed ROSMAP and ADNI genotype data were received from collaborators in the Hohman lab at Vanderbilt University in March 2021 and December 2021, respectively. Genetic variants with MAF < 0.1% and imputation quality scores of <0.8 were removed prior to further quality control measures. Due to small samples sizes of participants with substantial non-European ancestry, especially in replication cohorts, these participants were excluded from analysis. Standard GWAS quality control procedures were followed for variant and participant inclusion (see Supplementary Note).

## Defining and harmonizing neuropathology endophenotypes for analysis

In total, we combined and/or harmonized 11 neuropathology endophenotypes for analysis across the four studies. We note that there are differences in the way that some neuropathological data were collected across studies, and our strategy for harmonizing was informed by practical considerations for maximizing available samples sizes given the available endophenotypes. Thus, several synthetic NPE were created by merging existing NPE within a cohort or by harmonizing categorical variables from one cohort and continuous variables from another. Hippocampal sclerosis (HS), microinfarcts, and gross infarcts were recorded as binary case-control phenotypes. Ateriolosclerosis, atherosclerosis, cerebral amyloid angiopathy (CAA), Consortium to Establish a Registry for Alzheimer’s Disease (CERAD) score for neuritic plaques, diffuse amyloid plaques, limbic-predominant age-related TDP-43 encephalopathy neuropathologic change (LATE-NC), and Lewy body pathology were recorded as four-stage ordinal variables that either measured progressive severity of pathology (“none” < “mild” < “moderate” < “severe”) or progressing anatomical distribution of pathology. Braak NFT was recorded as a seven-stage ordinal variable that followed the anatomical distributional stages originally characterized by Braak and Braak.[57](#ref-braak1991) We provide deeper description of our harmonization approach in the Supplementary Note, and a detailed listing of variables harmonized across data sources to construct neuropathology endophenotypes for analysis is available in **Supplementary Table 3**.

## DNA methylation data

Pre-processed and quality-controlled DNA methylation data for 740 ROSMAP participants were downloaded from Synapse.org (Synapse IDs: syn3157275 and syn3191087). DNA methylation preparation and quality control measures have been previously described.[58](#ref-yu2015),[59](#ref-dejager2018) Briefly, approximately 50mg of frozen gray matter tissue from the dorsolateral prefrontal cortex (DLPFC) were sampled from each participant. DNA was then extracted and processed using the Illumina Infinium HumanMethylation450 BeadChip. Quality control measures included removing low-quality probes; removing participants with poor bisulfite-conversion efficiency; and adjusting methylation levels by age, sex, and batch, which adequately controlled for batch effects.[59](#ref-dejager2018) Missing methylation levels were imputed using 100-nearest neighbors.[58](#ref-yu2015),[59](#ref-dejager2018)

## RNA-Seq data

Pre-processed and quality-controlled bulk-tissue RNA-Seq data from the DLPFC of ROSMAP participants were downloaded from Synapse.org (Synapse IDs: syn21088596, syn21323366, syn3505732, syn3505724). As previously described, samples were prepared by sectioning approximately 100 mg of gray matter tissue from the DLPFC and RNA extracted using the Qiagen MiRNeasy Mini (cat no. 217004) protocol and then submitted for transcriptome library construction using the dUTP protocol and Illumina sequencing.[@dejager2018] A total of 634 participants in seven batches were sequenced with an average sequencing depth of 50 million paired reads per sample.[@dejager2018] Subsequent quality control and batch corrections were performed, and the final output of the RNA-Seq pipeline was fragments per kilobase of transcript per million mapped reads (FPKM).[59](#ref-dejager2018)

## Statistical Analyses

### Single-variant GWAS

We analyzed ordinal endophenotypes using proportional odds logistic mixed-effects models implemented in the POLMM R package[60](#ref-bi2021) and analyzed binary variables similarly with logistic mixed-effects models implemented in the SAIGE R package.[61](#ref-zhou2018) Fixed-effect covariates included age at death, sex, cohort, and the first 10 genetic principal components (PCs) created using the PCA in Related Samples (PC-AiR) method in the GENESIS R package.[62](#ref-conomos2015) We included a dense genetic relationship matrix (GRM) as a random effect to account for relatedness between participants. An additive mode of inheritance was assumed in all analyses.

Analyses of individual data sources proceeded in two stages. In stage one, GRM were constructed using a pruned set of independent variants, defined as having a pairwise within moving windows of 15 kilobase pairs (kbp). Null models, which including fixed covariates and the GRM, were then fitted using either the POLMM or SAIGE R packages. In stage 2, score tests were performed on each variant with a saddle-point approximation used to calculate p-values. We considered all variants with a to be genome-wide significant. To identify independent risk loci, we clumped results using the “--clump” flag in PLINK 1.9 with the pairwise linkage-disequilibrium (LD) threshold set to .[63](#ref-chang),[64](#ref-chang2015) Following analyses of individual cohorts, we performed fixed-effects meta-analyses using the METAL software using inverse-variance weighting on variants with MAF >1% in each cohort.[30](#ref-metal:f)

### Conditional analysis of the *APOE* region

The region surrounding the *APOE* gene on Chromosome 19 is consistently the single strongest genetic risk factor for LOAD in GWAS. Three common forms of the *APOE* gene – , , and – are present in our study populations (see Supplementary Table 2 for distribution of *APOE* alleles in participants), and the and alleles are associated with lower and higher risk of LOAD, respectively, relative to the allele.[65](#ref-reiman2020) We therefore expected that variants in the *APOE* region, defined as the region within 200 kbp from the start and end transcription sites of *APOE*, would be associated with multiple NPE in our study. Moreover, we hypothesized that genetic variants in the *APOE* region may influence neuropathology risk independently of the effects of *APOE* alleles. To test this hypothesis, we re-analyzed variants in Chromosome 19 while adjusting for *APOE* genotype. We limited re-analysis to endophenotypes with at least one genome-wide significant association signal within the *APOE* locus in the Stage 3 GWAS. *APOE* genotypes were determined using the rs7412 and rs429358 variants according to the SNPedia online reference.[66](#ref-apoe-s) Both variants had very high imputation quality scores in our data (, respectively). The genotype was used as reference, and we included fixed-effect indicator variables to adjust for , , , and genotypes. We chose this approach rather than adjusting for counts of and alleles because it is robust to potential non-linear effects of genotypes.

### Replication of known AD risk loci in NPE

We used the 83 AD and proxy-AD loci presented in a recent large GWAS to investigate whether AD-associated loci were associated with NPE.[6](#ref-bellenguez2022) We restricted our comparison to AD loci with lead variants with MAF , which excluded 3 loci, leaving 80 loci for comparison. We controlled the false-discovery rate for each NPE using the Benjamini-Hochberg procedure.[67](#ref-benjamini1995) Variants with an adjusted Q-value were considered significant.

### FUMA annotation, gene-prioritization, and functional enrichment pipeline

We mapped variants to genes and performed subsequent gene and gene-set analyses using the FUMA pipeline.[31](#ref-watanabe2017) Variants were mapped to genes if they had in the Stage 3 analysis and were located within 10 Kbp of a protein-coding gene’s transcription start or end sites. Gene-based analyses were performed using MAGMA. The top variant PCs that accounted for 99.9% of variance in a gene’s region were used to test for significance using an F-test. We considered genes with resulting P-values of to be significantly associated with NPE. Gene-set enrichment analyses were performed using MAGMA gene-set analysis of Gene Ontology (GO) and curated gene sets from MSigDB.[33](#ref-liberzon2011) Bonferroni P-value corrections were made for each NPE individually.

### Colocalization analyses

We used multiple sources of publicly available summary statistics from external studies as data sources for Bayesian colocalization analyses. First, we downloaded Genotype-Tissue Expression Project (GTEx) v8 European ancestry quantitative trait loci (QTL) analysis summary statistics, which contains summary statistics for significant gene expression and splicing QTL variants (eQTL and sQTL, respectively) in 48 body tissues.[68](#ref-gtexconsortium2017) We also used gene expression and DNA methylation QTL (mQTL) analysis summary statistics from studies using tissue from the dorsolateral prefrontal cortex of ROSMAP participants.[35](#ref-ng2017) These studies examined the associations of genetic variants with molecular traits and provide curated lists of significant QTL variants. Finally, we downloaded the summary statistics from a recent GWAS of LOAD for a targeted *post hoc* colocalization analysis in the *TMEM106B* and *GRN* genes.[6](#ref-bellenguez2022)

For each neuropathology endophenotype outcome in our study, we first compiled a list of genetic variants with p-values in our Stage 3 GWAS. We then queried the lists of significant QTL variants in GTEx and ROSMAP to identify neuropathology-associated QTL variants. For each genetic locus associated with neuropathology endophenotypes that had at least one significant QTL in either GTEx or ROSMAP, we performed colocalization analysis using the “coloc.abf” function in the coloc R package.[34](#ref-giambartolomei2014) For ordinal variables, we chose dichotomizing cut points to determine case-control proportions. We used coloc’s default prior probability of colocalization (PrC) of and considered a posterior as a threshold for evidence of colocalization.

To investigate whether shared GWAS signals drive association among multiple neuropathology endophenotypes, we also performed colocalization analysis on loci with variants exceeding a p-value threshold of and concordant effect direction for at least two NPE in the Stage 3 analysis.

### Association analyses using DLPFC DNA methylation and bulk RNA-seq data from ROSMAP

ROSMAP participants had post-mortem bulk tissue samples collected from the dorsolateral prefrontal cortex (DLPFC) which underwent DNA methylation quantification using Illumina DNAMethylation450 chip and gene expression and RNA-Seq using Illumina HiSeq2000.[59](#ref-dejager2018) In total, 708 ROSMAP participants had DNA methylation data available for analysis. We restricted analyses involving DNA methylation or RNA-Seq data to NPE-associated loci that reached the genome-wide significance threshold in the meta-analysis and also colocalized with mQTL or eQTL in a brain tissue in either GTEx or ROSMAP.

In our *APOE* -adjusted genetic association analysis, one locus near *APOE* remained significantly associated with cerebral amyloid angiopathy. This locus colocalized with DNA methylation levels at four CpG sites in ROSMAP. To investigate whether these CpG sites were in turn associated with CAA pathology, we combined individual-level DNA methylation and neuropathological data in ROSMAP for analysis. We used cumulative logit models using the “clm” function implemented in the “ordinal” R package with the semi-quantitative CAA variable described above as the outcome for analysis. We performed four analyses, with one of each of the four CpG sites tested as the independent variable of interest in each analysis. We adjusted for age, sex, ROS vs MAP study, bisulfite conversion efficiency, post-mortem interval, and *APOE* genotype in each analysis. Similar models were used to test associations between HS and LATE-NC and methylation levels at CpG sites cg09613507 and cg23422036. Wald tests were performed on the resulting parameter estimates to test for statistical significance.

For genes with significant eQTL in GTEx or ROSMAP that colocalized with NPE, we performed additional targeted analyses to assess the association between gene expression and NPE. We first assessed the association between NPE lead variants and gene expression in ROSMAP to confirm nominal eQTL status. We then performed generalized linear regression models between square-root or log-transformed mRNA expression and NPE outcomes adjusting for age at death, sex, PMI, and RNA integrity number.

### Replication of CAA locus Using Mayo Clinic Neuropathology GWAS

We used data from Mayo Clinic Brain Bank participants available from the Reddy et al. (2021) study of the genetic risk factors of CAA (data set heretofore referred to as MC-CAA) to attempt to replicate a novel CAA locus in the present study in an independent sample.[26](#ref-reddy2021) Neuropathology and genetic variant data were downloaded from Synapse (Synapse IDs: syn10930250, syn21499318, syn21522653, and syn21547862). Eight participants were identified as duplicates between batches or with NACC participants and removed. Whereas CAA is graded on a four-level ordinal scale in the present study, CAA in MC-CAA is graded as an average of CAA burden across five brain regions.[26](#ref-reddy2021) We therefore used linear regression with the outcome variable as sqrt(CAA) with the independent variable of interest being the number of G alleles of variant rs7247551. Covariates included *APOE* diplotype (, , , , or ), sex, age at death (truncated at 90 years), and the first 3 genetic PCs.

# List of abbreviations

ACT, the Adult Change in Thought Study; AD, Alzheimer’s disease; ADGC, Alzheimer’s Disease Genetics Consortium; ADNI, AD Neuroimaging Initiative; ADRC, Alzheimer’s Disease Research Center; CAA, cerebral amyloid angiopathy; eQTL, expression QTL; GRM, genetic relationship matrix; GTEx, Genotype-Tissue Expression Project; GWAS, genome-wide association study; HS, hippocampal sclerosis; kbp, kilobase pairs; LATE-NC, LATE neuropathologic change; LATE, limbic-predominant age-related TDP-43 encephalopathy; LD, linkage disequilibrium; LOAD, late-onset Alzheimer’s disease; MAP, Rush Memory and Aging Project; mQTL, methylation QTL; NACC, National Alzheimer’s Coordinating Center; NFT, neurofibrillary tangle; NP, neuropathology; NPE, neuropathology endophenotype; OR, odds ratio; PC-AiR, PCA in Related Samples; PC1, first principal component; PCA, principal components analysis; PET, positron emission tomography; PrC, probability of colocalizationQTL, quantitative trait locus; ROS, Religious Orders Study; ROSMAP, the Religious Orders Study and the Rush Memory and Aging Project; sQTL, splicing QTL; TOPMed, Trans-’Omics for Precision Medicine.

# Declarations

## Ethics approval and consent to participate

All study participants were deceased and the resulting data de-identified, and we exclusively used archival samples. Therefore, our study does 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, or/or recorded such that subjects cannot be identified.”

## Consent for publication

Not applicable.

## Availability of data and materials

All code used for data preparation and analysis is available at <https://www.github.com/lincoln-shade/np_phewas>. ROSMAP data can be requested at <https://www.radc.rush.edu> and <https://www.synapse.org>. ADGC data is can be requested from NIAGADS at <https://www.niagads.org/resources/related-projects/alzheimers-disease-genetics-consortium-adgc-collection>. NACC neuropathology data can be requested at <https://naccdata.org/>. ACT data can be requested at <https://actagingresearch.org/>. ADNI data can be downloaded at <https://adni.loni.usc.edu/>. The results published here are in whole or in part based on data obtained from the AD Knowledge Portal.

## Competing interests

JAS reported personal fees from Observational Study Monitoring Board Framingham, Observational Study Monitoring Board Discovery (National Institute of Neurological Disorders and Stroke), and Takeda Pharma. AJS reported support from Avid Radiopharmaceuticals, a subsidiary of Eli Lilly (in kind contribution of PET tracer precursor); Bayer Oncology (Scientific Advisory Board); Eisai (Scientific Advisory Board); Siemens Medical Solutions USA, Inc. (Dementia Advisory Board); NIH NHLBI (MESA Observational Study Monitoring Board); and Springer-Nature Publishing (Editorial Office Support as Editor-in-Chief, Brain Imaging and Behavior). All other authors declare that they have no competing interests.

## Funding

**The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article:** R56AG057191; F30NS124136; P30AG028383; the University of Kentucky Center for Clinical and Translational Science TL-1 Fellowship [grant number TL1TR001997]; the National Center for Advancing Translational Sciences [grant number UL1TR001998]; and the Dean of the College of Medicine, University of Kentucky. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH or the University of Kentucky.

Genotyping was supported by the Alzheimer’s Disease Genetics Consortium through the National Institute of Aging [grant numbers U01 AG032984, RC2AG036528]

Samples from the National Cell Repository for Alzheimer’s Disease (NCRAD), which receives government support under a cooperative agreement grant (U24 AG21886) awarded by the National Institute on Aging (NIA), were used in this study. We thank contributors who collected samples used in this study, as well as patients and their families, whose help and participation made this work possible;

Data for this study were prepared, archived, and distributed by the National Institute on Aging Alzheimer’s Disease Data Storage Site (NIAGADS) at the University of Pennsylvania (U24-AG041689-01).

We thank the study participants and staff of the Rush Alzheimer’s Disease Center. The Religious Orders Study and the Rush Memory and Aging Project are supported by grants from the National Institutes of Health: [grant number P30AG10161, P30AG72975, R01AG15819, R01AG17917, R01AG22018, R01AG33678, R01AG34374, R01AG36042, R01AG40039, R01AG042210, U01AG46152, U01AG61356, R01AG47976, R01AG43379, RF1AG54057, R01AG56352, R01NS78009, and UH2NS100599], and the Illinois Department of Public Health.

This research was funded by the National Institute on Aging (U19AG066567). Data collection for this work was additionally supported, in part, by prior funding from the National Institute on Aging (U01AG006781). All statements in this report, including its findings and conclusions, are solely those of the authors and do not necessarily represent the views of the National Institute on Aging or the National Institutes of Health. We thank the participants of the Adult Changes in Thought (ACT) study for the data they have provided and the many ACT investigators and staff who steward that data. You can learn more about ACT at: [https://actagingstudy.org/](https://actagingstudy.org/%22).

Data collection and sharing for this project was funded by the Alzheimer’s Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer’s Association; Alzheimer’s Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health ([www.fnih.org](https://fnih.org/)). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer’s Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

## Authors’ contributions

L.M.P.S. conceptualized study design, prepared data, performed analyses, and was a major contributor in writing the manuscript. Y.K. provided feedback on analyses and contributed to the manuscript. SC and ME helped prepare figures and provided extensive feedback on manuscript preparation. E.L.A. provided guidance on interpretation of *BIN1* results and provided extensive feedback on manuscript preparation. T.J.H. performed imputation and quality control on ROSMAP genotype data. K.N. and A.J.S. provided imputed and quality-controlled ADNI genotype data used in an earlier version of the manuscript and provided feedback on manuscript preparation. D.A.B. and J.A.S. provided ROSMAP neuropathology data and made critical revisions to the manuscript. P.T.N. provided guidance on defining neuropathology endophenotypes and contributed to the manuscript. D.W.F. conceptualized study design and provided feedback on manuscript preparation. All authors read and approved the final manuscript.

**ADD QI AND KHINE CONTRIBUIONS**

## Acknowledgements

The NACC database is funded by NIA/NIH Grant U24 AG072122. NACC data are contributed by the NIA-funded ADRCs: P30 AG062429 (PI James Brewer, MD, PhD), P30 AG066468 (PI Oscar Lopez, MD), P30 AG062421 (PI Bradley Hyman, MD, PhD), P30 AG066509 (PI Thomas Grabowski, MD), P30 AG066514 (PI Mary Sano, PhD), P30 AG066530 (PI Helena Chui, MD), P30 AG066507 (PI Marilyn Albert, PhD), P30 AG066444 (PI John Morris, MD), P30 AG066518 (PI Jeffrey Kaye, MD), P30 AG066512 (PI Thomas Wisniewski, MD), P30 AG066462 (PI Scott Small, MD), P30 AG072979 (PI David Wolk, MD), P30 AG072972 (PI Charles DeCarli, MD), P30 AG072976 (PI Andrew Saykin, PsyD), P30 AG072975 (PI David Bennett, MD), P30 AG072978 (PI Neil Kowall, MD), P30 AG072977 (PI Robert Vassar, PhD), P30 AG066519 (PI Frank LaFerla, PhD), P30 AG062677 (PI Ronald Petersen, MD, PhD), P30 AG079280 (PI Eric Reiman, MD), P30 AG062422 (PI Gil Rabinovici, MD), P30 AG066511 (PI Allan Levey, MD, PhD), P30 AG072946 (PI Linda Van Eldik, PhD), P30 AG062715 (PI Sanjay Asthana, MD, FRCP), P30 AG072973 (PI Russell Swerdlow, MD), P30 AG066506 (PI Todd Golde, MD, PhD), P30 AG066508 (PI Stephen Strittmatter, MD, PhD), P30 AG066515 (PI Victor Henderson, MD, MS), P30 AG072947 (PI Suzanne Craft, PhD), P30 AG072931 (PI Henry Paulson, MD, PhD), P30 AG066546 (PI Sudha Seshadri, MD), P20 AG068024 (PI Erik Roberson, MD, PhD), P20 AG068053 (PI Justin Miller, PhD), P20 AG068077 (PI Gary Rosenberg, MD), P20 AG068082 (PI Angela Jefferson, PhD), P30 AG072958 (PI Heather Whitson, MD), P30 AG072959 (PI James Leverenz, MD).

The results published here are in part based on data obtained from the AD Knowledge Portal.

# References

1. Farfel, J. M. *et al.* [Relation of genomic variants for alzheimer disease dementia to common neuropathologies](https://doi.org/10.1212/WNL.0000000000002909). *Neurology* **87**, 489–496 (2016).

2. Katsumata, Y. *et al.* [Multiple gene variants linked to Alzheimer’s-type clinical dementia via GWAS are also associated with non-Alzheimer’s neuropathologic entities](https://doi.org/10.1016/j.nbd.2022.105880). *Neurobiology of Disease* **174**, 105880 (2022).

3. Karanth, S. *et al.* [Prevalence and Clinical Phenotype of Quadruple Misfolded Proteins in Older Adults](https://doi.org/10.1001/jamaneurol.2020.1741). *JAMA Neurology* **77**, 1299 (2020).

4. Jansen, I. E. *et al.* [Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer’s disease risk](https://doi.org/10.1038/s41588-018-0311-9). *Nature Genetics* **51**, 404–413 (2019).

5. Wightman, D. P. *et al.* [A genome-wide association study with 1,126,563 individuals identifies new risk loci for Alzheimer’s disease](https://doi.org/10.1038/s41588-021-00921-z). *Nature Genetics* **53**, 1276–1282 (2021).

6. Bellenguez, C. *et al.* New insights into the genetic etiology of Alzheimer’s disease and related dementias. *Nature Genetics* 1–25 (2022) doi:[10.1038/s41588-022-01024-z](https://doi.org/10.1038/s41588-022-01024-z).

7. Kunkle, B. W. *et al.* [Genetic meta-analysis of diagnosed Alzheimer’s disease identifies new risk loci and implicates A?, tau, immunity and lipid processing](https://doi.org/10.1038/s41588-019-0358-2). *Nature Genetics* **51**, 414 (2019).

8. Escott-Price, V. & Hardy, J. [Genome-wide association studies for alzheimer’s disease: Bigger is not always better](https://doi.org/10.1093/braincomms/fcac125). *Brain Communications* **4**, fcac125 (2022).

9. Karran, E. & De Strooper, B. [The amyloid hypothesis in Alzheimer disease: new insights from new therapeutics](https://doi.org/10.1038/s41573-022-00391-w). *Nature Reviews. Drug Discovery* **21**, 306–318 (2022).

10. Nelson, P. T., Braak, H. & Markesbery, W. R. [Neuropathology and cognitive impairment in alzheimer disease: A complex but coherent relationship](https://doi.org/10.1097/NEN.0b013e3181919a48). *Journal of neuropathology and experimental neurology* **68**, 1–14 (2009).

11. Nelson, P. T. *et al.* [Limbic-predominant age-related TDP-43 encephalopathy (LATE): consensus working group report](https://doi.org/10.1093/brain/awz099). *Brain* **142**, 1503–1527 (2019).

12. Nelson, P. T. *et al.* [Frequency of LATE neuropathologic change across the spectrum of alzheimer’s disease neuropathology: Combined data from 13 community-based or population-based autopsy cohorts](https://doi.org/10.1007/s00401-022-02444-1). *Acta neuropathologica* **144**, 27–44 (2022).

13. Brenowitz, W. D., Monsell, S. E., Schmitt, F. A., Kukull, W. A. & Nelson, P. T. [Hippocampal sclerosis of aging is a key alzheimer’s disease mimic: Clinical-pathologic correlations and comparisons with both alzheimer’s disease and non-tauopathic frontotemporal lobar degeneration](https://doi.org/10.3233/JAD-131880). *Journal of Alzheimer’s disease : JAD* **39**, 691–702 (2014).

14. Dugan, A. J. *et al.* [Analysis of genes (TMEM106B, GRN, ABCC9, KCNMB2, and APOE) implicated in risk for LATE-NC and hippocampal sclerosis provides pathogenetic insights: A retrospective genetic association study](https://doi.org/10.1186/s40478-021-01250-2). *Acta Neuropathologica Communications* **9**, 152 (2021).

15. Skrobot, O. A. *et al.* [Vascular cognitive impairment neuropathology guidelines (VCING): The contribution of cerebrovascular pathology to cognitive impairment](https://doi.org/10.1093/brain/aww214). *Brain* **139**, 2957–2969 (2016).

16. Weber, S. A., Patel, R. K. & Lutsep, H. L. [Cerebral amyloid angiopathy: diagnosis and potential therapies](https://doi.org/10.1080/14737175.2018.1480938). *Expert Review of Neurotherapeutics* **18**, 503–513 (2018).

17. Smith, E. E., Schneider, J. A., Wardlaw, J. M. & Greenberg, S. M. [Cerebral microinfarcts: The invisible lesions](https://doi.org/10.1016/S1474-4422(11)70307-6). *Lancet Neurology* **11**, 272–282 (2012).

18. Arvanitakis, Z. *et al.* [The Relationship of Cerebral Vessel Pathology to Brain Microinfarcts](https://doi.org/10.1111/bpa.12365). *Brain Pathology (Zurich, Switzerland)* **27**, 77–85 (2017).

19. Neltner, J. H. *et al.* [Arteriolosclerosis that affects multiple brain regions is linked to hippocampal sclerosis of ageing](https://doi.org/10.1093/brain/awt318). *Brain* **137**, 255–267 (2014).

20. Arvanitakis, Z., Capuano, A. W., Leurgans, S. E., Bennett, D. A. & Schneider, J. A. [Relation of cerebral vessel disease to alzheimer’s disease dementia and cognitive function in older persons: A cross-sectional study](https://doi.org/10.1016/S1474-4422(16)30029-1). *The Lancet. Neurology* **15**, 934–943 (2016).

21. Ighodaro, E. T. *et al.* [Risk factors and global cognitive status related to brain arteriolosclerosis in elderly individuals](https://doi.org/10.1177/0271678X15621574). *Journal of Cerebral Blood Flow & Metabolism* **37**, 201–216 (2017).

22. Beecham, G. W. *et al.* [Genome-Wide Association Meta-analysis of Neuropathologic Features of Alzheimer’s Disease and Related Dementias](https://doi.org/10.1371/journal.pgen.1004606). *PLOS Genetics* **10**, e1004606 (2014).

23. Shade, L. M. *et al.* Genome-wide association study of brain arteriolosclerosis. *Journal of Cerebral Blood Flow & Metabolism* 0271678X211066299 (2022) doi:[10.1177/0271678X211066299](https://doi.org/10.1177/0271678X211066299).

24. Vattathil, S. M. *et al.* [A Genetic Study of Cerebral Atherosclerosis Reveals Novel Associations with NTNG1 and CNOT3](https://doi.org/10.3390/genes12060815). *Genes* **12**, 815 (2021).

25. Farrell, K. *et al.* [Genome-wide association study and functional validation implicates JADE1 in tauopathy](https://doi.org/10.1007/s00401-021-02379-z). *Acta Neuropathologica* **143**, 33–53 (2022).

26. Reddy, J. S. *et al.* [Genome-wide analysis identifies a novel LINC-PINT splice variant associated with vascular amyloid pathology in alzheimer’s disease](https://doi.org/10.1186/s40478-021-01199-2). *Acta Neuropathologica Communications* **9**, 93 (2021).

27. Revelle, W. *Psych: Procedures for psychological, psychometric, and personality research*. <https://CRAN.R-project.org/package=psych> (2022).

28. Fox, J. [*Polycor: Polychoric and polyserial correlations*](https://CRAN.R-project.org/package=polycor). (2022).

29. Kolde, R. [*Pheatmap: Pretty heatmaps*](https://CRAN.R-project.org/package=pheatmap). (2019).

30. [METAL: Fast and efficient meta-analysis of genomewide association scans. - PubMed - NCBI](https://www.ncbi.nlm.nih.gov/pubmed/20616382).

31. Watanabe, K., Taskesen, E., Bochoven, A. van & Posthuma, D. [Functional mapping and annotation of genetic associations with FUMA](https://doi.org/10.1038/s41467-017-01261-5). *Nature Communications* **8**, 1826 (2017).

32. Leeuw, C. A. de, Mooij, J. M., Heskes, T. & Posthuma, D. [MAGMA: Generalized Gene-Set Analysis of GWAS Data](https://doi.org/10.1371/journal.pcbi.1004219). *PLOS Computational Biology* **11**, e1004219 (2015).

33. Liberzon, A. *et al.* [Molecular signatures database (MSigDB) 3.0](https://doi.org/10.1093/bioinformatics/btr260). *Bioinformatics* **27**, 1739–1740 (2011).

34. Giambartolomei, C. *et al.* [Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics](https://doi.org/10.1371/journal.pgen.1004383). *PLOS Genetics* **10**, e1004383 (2014).

35. Ng, B. *et al.* [An xQTL map integrates the genetic architecture of the human brain’s transcriptome and epigenome](https://doi.org/10.1038/nn.4632). *Nature Neuroscience* **20**, 1418–1426 (2017).

36. Walker, R. M. *et al.* [Identification of epigenome-wide DNA methylation differences between carriers of APOE ε4 and APOE ε2 alleles](https://doi.org/10.1186/s13073-020-00808-4). *Genome Medicine* **13**, 1 (2021).

37. Walker, R. M. *et al.* [Epigenome-wide analyses identify DNA methylation signatures of dementia risk](https://doi.org/10.1002/dad2.12078). *Alzheimer’s & Dementia: Diagnosis, Assessment & Disease Monitoring* **12**, e12078 (2020).

38. Shao, Y. *et al.* [DNA methylation of TOMM40-APOE-APOC2 in alzheimer’s disease](https://doi.org/10.1038/s10038-017-0393-8). *Journal of human genetics* **63**, 459–471 (2018).

39. Rollinson, S. *et al.* [Frontotemporal lobar degeneration genome wide association study replication confirms a risk locus shared with amyotrophic lateral sclerosis](https://doi.org/10.1016/j.neurobiolaging.2010.12.005). *Neurobiology of Aging* **32**, 758.e1–758.e7 (2011).

40. Ciani, M., Benussi, L., Bonvicini, C. & Ghidoni, R. [Genome wide association study and next generation sequencing: A glimmer of light toward new possible horizons in frontotemporal dementia research](https://doi.org/10.3389/fnins.2019.00506). *Frontiers in Neuroscience* **13**, 506 (2019).

41. Katsumata, Y., Nelson, P. T., Ellingson, S. R. & Fardo, D. W. [Gene-based association study of genes linked to hippocampal sclerosis of aging neuropathology: GRN, TMEM106B, ABCC9, and KCNMB2](https://doi.org/10.1016/j.neurobiolaging.2017.01.003). *Neurobiology of Aging* **53**, 193.e17–193.e25 (2017).

42. Nelson, P. T. *et al.* [Reassessment of risk genotypes (GRN, TMEM106B, and ABCC9 variants) associated with hippocampal sclerosis of aging pathology](https://doi.org/10.1097/NEN.0000000000000151). *Journal of Neuropathology and Experimental Neurology* **74**, 75–84 (2015).

43. Holler, C. J. *et al.* [Bridging Integrator 1 (BIN1) Protein Expression Increases in the Alzheimer’s Disease Brain and Correlates with Neurofibrillary Tangle Pathology](https://doi.org/10.3233/JAD-132450). *Journal of Alzheimer’s Disease* **42**, 1221–1227 (2014).

44. Franzmeier, N., Rubinski, A., Neitzel, J. & Ewers, M. [The BIN1 rs744373 SNP is associated with increased tau-PET levels and impaired memory](https://doi.org/10.1038/s41467-019-09564-5). *Nature Communications* **10**, 1766 (2019).

45. Klarin, D. *et al.* [Genome-wide association study of peripheral artery disease in the million veteran program](https://doi.org/10.1038/s41591-019-0492-5). *Nature medicine* **25**, 1274–1279 (2019).

46. Nikpay, M. *et al.* [A comprehensive 1000 genomes-based genome-wide association meta-analysis of coronary artery disease](https://doi.org/10.1038/ng.3396). *Nature genetics* **47**, 1121–1130 (2015).

47. Steffensen, L. B. & Rasmussen, L. M. [A role for collagen type IV in cardiovascular disease?](https://doi.org/10.1152/ajpheart.00070.2018) *American Journal of Physiology-Heart and Circulatory Physiology* **315**, H610–H625 (2018).

48. Blevins, B. L. *et al.* [Brain arteriolosclerosis](https://doi.org/10.1007/s00401-020-02235-6). *Acta Neuropathologica* **141**, 1–24 (2021).

49. Rannikmae, K. *et al.* [COL4A2 is associated with lacunar ischemic stroke and deep ICH: Meta-analyses among 21,500 cases and 40,600 controls](https://doi.org/10.1212/WNL.0000000000004560). *Neurology* **89**, 1829–1839 (2017).

50. Hautakangas, H. *et al.* [Genome-wide analysis of 102,084 migraine cases identifies 123 risk loci and subtype-specific risk alleles](https://doi.org/10.1038/s41588-021-00990-0). *Nature Genetics* **54**, 152–160 (2022).

51. Guennewig, B. *et al.* [Defining early changes in Alzheimer’s disease from RNA sequencing of brain regions differentially affected by pathology](https://doi.org/10.1038/s41598-021-83872-z). *Scientific Reports* **11**, 4865 (2021).

52. Zhang, Y. *et al.* [Purification and Characterization of Progenitor and Mature Human Astrocytes Reveals Transcriptional and Functional Differences with Mouse](https://doi.org/10.1016/j.neuron.2015.11.013). *Neuron* **89**, 37–53 (2016).

53. Besser, L. M. *et al.* [The Revised National Alzheimer’s Coordinating Center’s Neuropathology Form-Available Data and New Analyses](https://doi.org/10.1093/jnen/nly049). *Journal of Neuropathology and Experimental Neurology* **77**, 717–726 (2018).

54. Bennett, D. A. *et al.* [Religious orders study and rush memory and aging project](https://doi.org/10.3233/JAD-179939). *Journal of Alzheimer’s disease : JAD* **64**, S161–S189 (2018).

55. Kukull, W. A. *et al.* [Dementia and Alzheimer Disease Incidence: A Prospective Cohort Study](https://doi.org/10.1001/archneur.59.11.1737). *Archives of Neurology* **59**, 1737 (2002).

56. Taliun, D. *et al.* [Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program](https://doi.org/10.1038/s41586-021-03205-y). *Nature* **590**, 290–299 (2021).

57. Braak, H. & Braak, E. [Neuropathological stageing of Alzheimer-related changes](https://doi.org/10.1007/BF00308809). *Acta Neuropathologica* **82**, 239–259 (1991).

58. Yu, L. *et al.* [Association of Brain DNA Methylation in SORL1, ABCA7, HLA-DRB5, SLC24A4, and BIN1 With Pathological Diagnosis of Alzheimer Disease](https://doi.org/10.1001/jamaneurol.2014.3049). *JAMA Neurology* **72**, 15–24 (2015).

59. De Jager, P. L. *et al.* [A multi-omic atlas of the human frontal cortex for aging and Alzheimer’s disease research](https://doi.org/10.1038/sdata.2018.142). *Scientific Data* **5**, 180142 (2018).

60. Bi, W. *et al.* [Efficient mixed model approach for large-scale genome-wide association studies of ordinal categorical phenotypes](https://doi.org/10.1016/j.ajhg.2021.03.019). *The American Journal of Human Genetics* **108**, 825–839 (2021).

61. Zhou, W. *et al.* [Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies](https://doi.org/10.1038/s41588-018-0184-y). *Nature Genetics* **50**, 1335–1341 (2018).

62. Conomos, M. P., Miller, M. B. & Thornton, T. A. [Robust Inference of Population Structure for Ancestry Prediction and Correction of Stratification in the Presence of Relatedness](https://doi.org/10.1002/gepi.21896). *Genetic Epidemiology* **39**, 276–293 (2015).

63. Chang, C. C. & Purcell, S. [*Plink 1.9*](https://www.cog-genomics.org/plink/1.9/).

64. Chang, C. C. *et al.* [Second-generation PLINK: rising to the challenge of larger and richer datasets](https://doi.org/10.1186/s13742-015-0047-8). *GigaScience* **4**, (2015).

65. Reiman, E. M. *et al.* [Exceptionally low likelihood of Alzheimer’s dementia in APOE2 homozygotes from a 5,000-person neuropathological study](https://doi.org/10.1038/s41467-019-14279-8). *Nature Communications* **11**, 667 (2020).

66. [APOE - SNPedia](https://www.snpedia.com/index.php/APOE).

67. Benjamini, Y. & Hochberg, Y. [Controlling the false discovery rate: A practical and powerful approach to multiple testing](https://www.jstor.org/stable/2346101). *Journal of the Royal Statistical Society. Series B (Methodological)* **57**, 289–300 (1995).

68. GTEx Consortium *et al.* [Genetic effects on gene expression across human tissues](https://doi.org/10.1038/nature24277). *Nature* **550**, 204–213 (2017).

69. Rj, P. *et al.* [LocusZoom: regional visualization of genome-wide association scan results](https://doi.org/10.1093/bioinformatics/btq419). *Bioinformatics (Oxford, England)* **26**, (2010).

# Figure captions

**Figure 1: Overview of study design.** White boxes represent sources of data or summary statistics. Purple boxes represent genetic association analysis, while green boxes represent downstream functional analyses. Stage 1 GWAS were performed using National Alzheimer’s Coordinating Center (NACC) neuropathology (NP) data set. Variants reaching a suggestive threshold of association () in Stage 1 then underwent attempted nominal replication () separately in three independent studies: Religious Orders Study and the Rush Memory and Aging Project (ROSMAP), the Adult Changes in Thought (ACT) study, and the Alzheimer’s Disease Neuroimaging Initiative (ADNI). In Stage 3, data from NACC, ROSMAP, ACT, and ADNI were then merged for pooled GWAS. Stage 3 summary statistics were then used in additional analyses, including attempted replication of Alzheimer’s disease risk loci in neuropathology endophenotypes, gene-based and pathway analysis, and colocalization analysis. Significant Stage 3 loci that showed evidence of colocalization were then investigated using DNA methylation and RNA-Seq data from ROSMAP.

**Figure 2: Manhattan plots of Stage 3 results for LATE-NC, Braak stage, cerebral atherosclerosis, and HS.** *P*-values are derived using saddle-point approximation and are truncated at . Genome-wide significant loci are highlighted in orange and annotated with mapped gene (known loci) or nearest protein-coding gene (novel loci). The solid black line represents the genome-wide significance level (), and the black dotted line represents the suggestive significance level ().

**Figure 3: Novel associations identified between *COL4A1* and *PIK3R5* loci with cerebral atherosclerosis and Braak stage, respectively.** **A)** Regional LocusZoom plot of atherosclerosis-associated *COL4A1* locus.[69](#ref-rj2010) Mb, megabase; uses hg19. **B)** Forest plot for individual and pooled cohorts for lead variant rs2000660 odds ratio and 95% confidence intervals. Lead variant reached a suggestive level of significance in NACC analysis (, ) and was nominally replicated in ROSMAP (OR = 0.68, p = 0.0079). ADNI results are not shown due to wide confidence interval (OR = 1.25, 95% CI = 0.08-18.84). **C)** Human brain cell-type expression profile of *COL4A1* in Zang et al. (2016).[52](#ref-zhang2016) *COL4A1* is preferentially expressed in fetal astrocytes and endothelial cells with lower expression in neurons. FPKM, Fragments Per Kilobase of transcript per Million mapped reads. **D)** Regional LocusZoom plot of Braak-associated *PIK3R5* locus. Mb, megabase; uses hg19. **E)** Forest plot for individual and pooled cohorts for rs72807981 odds ratio and 95% confidence intervals. Lead variant rs72807981 reached a suggestive level of significance in NACC analysis (, ) and was nominally replicated in ACT (, ). **F)** Human brain cell-type expression profile of *PIK3R5* in Zang et al. (2016).[52](#ref-zhang2016) *PIK3R5* is primarily expressed in microglia. FPKM, Fragments Per Kilobase of transcript per Million mapped reads.

**Figure 4: Locus in *APOE/APOC2* region associates with CAA and colocalizes with mQTL in ROSMAP. A)** Regional LocusZoom of associated locus. Mb, megabase; uses hg19. Eighteen genes are not shown due to space considerations. **B)** Regional plot showing CAA associations and colocalization with four mQTL in ROSMAP with CpG sites: cg04401876, cg09555818, cg10169327, and cg13119609. Posterior probability of colocalization equals 97% with each of the four mQTL. The darkly bordered box indicates the region with SNPs most associated with CAA and the four colocalizing mQTL. **C)** Associations between the same four CpG sites and CAA pathology in ROSMAP (N = 708). Two sites, cg09555818 (*p* = 0.004) and cg13119609 (*p* = 0.0007) were significantly associated with CAA pathology. **D)** Schematic showing regression beta estimates between SNP and CpG sites, SNP and CAA, and CpG sites and CAA.