Proposed journal: ***Molecular Neurodegeneration***

# Title page

## Title

Genome-wide association study of multiple neuropathology endophenotypes identifies novel risk loci and provides insights into known Alzheimer’s risk loci

## Authors

Lincoln M.P. Shade1

Yuriko Katsumata1,2

Steven A. Claas1

Mark T. W. Ebbert2,3,4

Erin L. Abner2,5

Timothy J. Hohman6

Shubhabrata Mukherjee7

Kwagsik Nho8

Andrew J. Saykin8

David A. Bennett9

Julie A. Schneider10

Peter T. Nelson2,11

David W. Fardo1,2

for the Alzheimer’s Disease Neuroimaging Initiative\*

## Author Affiliations

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

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

3Division of Biomedical Informatics, Department of Internal Medicine, University of Kentucky, Lexington, KY

4Department of Neuroscience, University of Kentucky College of Medicine, Lexington, KY

5Department of Epidemiology and Environmental Health, College of Public Health, University of Kentucky, Lexington, KY

6Vanderbilt Memory & Alzheimer’s Center, Department of Neurology, Vanderbilt University Medical Center, Nashville, Tennessee

7Department of Medicine, University of Washington, Seattle, WA

8Department of Radiology & Imaging Sciences, Indiana University School of Medicine, Indianapolis, IN

9Rush Alzheimer’s Disease Center and Department of Neurological Sciences, Rush Medical College, Chicago, IL

10Rush Alzheimer’s Disease Center and Departments of Neurology and Pathology, Rush University Medical Center, Chicago, IL

11Department of Pathology and Laboratory Medicine, University of Kentucky, Lexington, KY

\*Data used in preparation of this article were obtained from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: <http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf>

## 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

# Abstract

## Background

Alzheimer’s disease is highly heritable and exhibits neuropathological hallmarks of neurofibrillary tau tangles and neuritic amyloid plaques. Previous genome-wide association studies (GWAS) have identified over 70 genomic risk loci of clinically diagnosed Alzheimer’s disease. However, upon autopsy, many Alzheimer’s disease patients have multiple comorbid neuropathologies that may have independent or pleiotropic genomic risk factors. Autopsy data combined with GWAS provides the opportunity to study the genetic risk factors of individual neuropathologies.

## Methods

We studied the genome-wide risk factors of eleven Alzheimer’s disease-related neuropathology endophenotypes. We used four sources of neuropathological data: National Alzheimer’s Coordinating Center, Religious Orders Study and Rush Memory and Aging Project, Adult Changes in Thought study, and Alzheimer’s Disease Neuroimaging Initiative. We used generalized linear mixed models to identify risk loci, followed by Bayesian colocalization analyses to identify potential functional mechanisms by which genetic loci influence neuropathology risk.

## Results

We identified two novel loci associated with neuropathology: one *PIK3R5* locus (lead variant rs72807981) with neurofibrillary pathology, and one *COL4A1* locus (lead variant rs2000660) with cerebral atherosclerosis. We also confirmed associations between known Alzheimer’s genes and multiple neuropathology endophenotypes, including *APOE* (neurofibrillary tangles, neuritic plaques, diffuse plaques, cerebral amyloid angiopathy, and TDP-43 pathology); *BIN1* (neurofibrillary tangles and neuritic plaques); and *TMEM106B* (TDP-43 pathology and hippocampal sclerosis). After adjusting for *APOE* genotype, we identified a locus near *APOC2* (lead variant rs4803778) associated with cerebral amyloid angiopathy that influences DNA methylation at nearby CpG sites in the cerebral cortex.

## Conclusions

rs2000660 is in strong linkage disequilibrium with a synonymous coding variant (rs650724) of *COL4A1*, providing a candidate functional variant. Two CpG sites affected by the cerebral amyloid angiopathy-associated *APOC2* locus were previously associated with dementia in an independent cohort, suggesting that the effect of this locus on disease may be mediated by DNA methylation. *BIN1* is associated with neurofibrillary tangles and neuritic plaques but not with amyloid pathology. *TMEM106B* is associated with hippocampal sclerosis and TDP-43 pathology but not the canonical Alzheimer’s disease pathologies. These findings provide insights into known Alzheimer’s disease risk loci by refining the pathways affected by these risk genes.

# Key words

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

# Background

Aged individuals often accumulate multiple brain pathologies that contribute individually and synergistically to cognitive decline and dementia. Alzheimer’s disease (AD) is the most commonly diagnosed form of dementia and poses an enormous burden on human health and well-being. In the United States alone, over six million individuals are living with AD, and the disease imposes financial costs of treatment and care of over $300 billion annually [[1](#ref-2021alz2021)].

The most common form of AD, late-onset AD (LOAD), presents in individuals aged 65 years and older and is highly heritable, with twin and family studies estimating its heritability at ~60% [[2](#ref-ridge2013)–[4](#ref-gatz2006)]. Genome-wide association studies (GWAS) have to date identified over 70 genetic risk loci for LOAD that are involved biological processes including amyloid precursor protein processing, immune response, and extracellular matrix regulation [[5](#ref-jansen2019)–[7](#ref-bellenguez2022)]. Most dementia GWAS to date have focused on clinically diagnosed LOAD or proxy outcomes based on family history of dementia [[5](#ref-jansen2019)–[8](#ref-kunkle2019)]; however, late-onset dementia diagnosed as 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 [[9](#ref-farfel2016)–[11](#ref-karanth2020)].

While neuritic plaques and neurofibrillary tangles (NFT) are present in the brains of the majority of people diagnosed with clinical AD, many people either do not have substantial AD neuropathology or have both AD and comorbid non-AD pathologies [[11](#ref-karanth2020), [12](#ref-nelson2009)]. For example, TDP-43 pathology contributes to an amnestic syndrome, limbic-predominant age-related TDP-43 encephalopathy (LATE), that presents similarly to LOAD [[13](#ref-nelson2019)]. However, LATE and LOAD exhibit differing patterns of pathogenesis in aged autopsy cohorts. LOAD pathology peaks in prevalence in individuals who die at age 95 years and decreases thereafter [[14](#ref-farfel2019)]. In contrast, prevalence of LATE neuropathologic change (LATE-NC) increases monotonically with age of death, with TDP-43 pathology often first appearing in the limbic region or entorhinal cortex and progressing to the neocortex [[13](#ref-nelson2019), [14](#ref-farfel2019)]. Hippocampal sclerosis (HS) of aging, a pathology characterized by uni- or bilateral 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 [[13](#ref-nelson2019), [15](#ref-brenowitz2014)].

In addition to proteinopathies, non-stroke vascular pathologies are common in elderly autopsied research participants and contribute to cognitive decline and dementia [[16](#ref-skrobot2016)]. Cerebral amyloid angiopathy (CAA) is one such vascular pathology characterized by amyloid-beta deposition in cerebral blood vessels [[17](#ref-weber2018)]. CAA often co-occurs with LOAD, but can independently contribute to cerebral injury by causing hemorrhages in the brain parenchyma [[16](#ref-skrobot2016), [17](#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 [[16](#ref-skrobot2016), [18](#ref-smith2012)]. Cerebral atherosclerosis and sclerosis of small cerebral blood vessels, called brain arteriolosclerosis, predispose individuals to infarcts and hippocampal sclerosis [[19](#ref-arvanitakis2017), [20](#ref-neltner2014)]; moreover, they contribute to cognitive decline even after adjusting for the presence of related pathologies [[21](#ref-arvanitakis2016), [22](#ref-ighodaro2017)]. Collectively, these factors reveal an increasingly complex and synergistic web of pathologies contributing to cognitive impairment and dementia, and demonstrate the need for continued study into genomic risk factors of dementia beyond those of clinically diagnosed LOAD.

The investigation of genetic risk of neuropathologies is an attractive approach for uncovering biological pathways involved in the development of LOAD and related dementias. Neuropathologies are commonly operationalized in aged autopsy cohorts as binary or semi-quantitative (or ordinal) neuropathology endophenotypes (NPE). Prior GWAS of neuropathology endophenotypes have confirmed known LOAD risk loci and have identified potential neuropathology risk loci [[23](#ref-beecham2014)–[26](#ref-farrell2022)]. However, continued recruitment in ongoing autopsy cohorts and methodological development for GWAS and downstream functional analysis provides additional opportunity for in-depth investigation into multiple neuropathologies studied across multiple cohorts.

In this study, we perform GWAS on eleven neuropathology endophenotypes collected in four high-quality studies with autopsy and genotype data available. Endophenotypes studied include AD-associated pathologies; LATE-NC; hippocampal sclerosis; vascular NPE including cerebral amyloid angiopathy, gross infarcts, microinfarcts, atherosclerosis, and arteriolosclerosis; and Lewy body pathology. We also perform downstream functional analyses to explore potential biological functional mechanisms of identified risk loci and provide insight into the shared genomic risk of neuropathologies.

# Methods

## Participants

We used genotype and neuropathology data from four high-quality data sources: the National Alzheimer’s Coordinating Center (NACC), the Religious Orders Study and the Rush Memory and Aging Project (ROSMAP), the Adult Changes in Thought (ACT) study, and the AD Neuroimaging Initiative (ADNI). 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 in order to maximize sample size of the initial analysis using only NACC participants (see Statistical Analyses subsection).

All study participants were deceased and the resulting data 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.”

The present study used NACC data from 35 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 [[27](#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 has been previously described and consists of harmonized data from two longitudinal cohort studies: The Religious Orders Study (ROS) and the Rush Memory and Aging Project (MAP) [[28](#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.

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 [[29](#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.

The 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.

## 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 [[30](#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.

Quality control on genotype data was performed separately for each data source. Quality control and inclusion/exclusion criteria closely followed that used in our previous brain arteriolosclerosis GWAS [[24](#ref-shade2022)]. Briefly, we included only participants labeled as non-Hispanic white or similar (European, etc.) if available. After imputation, we first identified and subsequently removed duplicate samples using the KING software “--duplicate” option [[31](#ref-manichaikul2010)]. We then removed participants without autopsy data available by merging genotype sample identification data with neuropathology data sets and retaining only samples present in both data sets. We then iteratively removed genetic variants and participants until no variants were missing in more than 5% of participants and no participants were missing more than 5% of variants (however, average genotype coverage was 99.7%, and no variants or participants were actually removed during this process). We then excluded participants with unusually high or low ( 3 standard deviations from mean) genetic heterogeneity, as measured by the PLINK 1.9 “--het” flag [[32](#ref-chang)]. Finally, we merged participants with the 1000 Genomes Phase 3 (1000 Genomes) cohorts [[33](#ref-the1000genomesprojectconsortium2012)] and performed principal components analysis (PCA) on a subset of independent variants (measured by pairwise ). We excluded participants with substantial non-European ancestry, as determined by distance from the 1000 Genomes EUR superpopulation centroid using the first two principal components (PCs).

## Defining and harmonizing neuropathology endophenotypes for analysis

In total, we combined 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. A detailed listing of variables harmonized across cohorts to construct neuropathology endophenotypes for analysis is available in **Supplementary Table S3**. We briefly describe our harmonization methods below.

Arteriolosclerosis, Braak NFT stage, CAA, atherosclerosis, Consortium to Establish a Registry for Alzheimer’s Disease (CERAD) score for neuritic plaques, microinfarcts, and gross infarcts had variables in each cohort with directly comparable coding definitions and were straightforwardly renamed and combined with minimal recoding. Analyzed arteriolosclerosis, CAA, atherosclerosis, and CERAD score variables each had four stages with the following labels: 0 (“none”), 1 (“mild”), 2 (“moderate”), and 3 (“severe”). Microinfarcts and gross infarcts were labeled either 0 (“absent”) or 1 (“present”). The Braak NFT Stage variable analyzed followed the staging criteria previously described in the literature and had seven levels, ranging from 0 (absent NFT) to six (diffuse NFT throughout cortex and large loss of neurons) [[34](#ref-braak1991)].

LATE-NC was recorded differently in several of the data sets and was harmonized to a four-level outcome variable following the simplified staging of TDP-43 pathology outlined in Figure 3B of the 2019 LATE working group report [[13](#ref-nelson2019)]. The following levels were used in the analyzed LATE-NC variable: 0, indicating lack of recorded TDP-43 proteinopathy; 1, indicating TDP-43 deposits in the amygdala only; 2, indicating deposits in the hippocampus or entorhinal cortex; and 3, indicating deposits in the neocortex. TDP-43 data was not available in ACT. In ROSMAP, TDP-43 pathology is recorded as a single variable following the same staging detailed above. In NACC and ADNI, the presence of TDP-43 pathology in each region of the brain is recorded as a separate binary indicator variable. To collapse TDP-43 pathology to a single ordinal variable in these data sets, we assigned a value based on the presence of the “highest” region where TDP-43 was present (*e.g.* a participant with TDP-43 pathology in both the hippocampus and the neocortex would be assigned a value of 3). Participants were labeled as 0 if they met two conditions: (1) they had recorded TDP-43 data available for at least one of the brain regions used for staging and (2) TDP-43 pathology was noted as absent in all the regions for which they had data available.

Diffuse amyloid plaque pathology was recorded as a four-stage Thal phase of amyloid deposition in NACC, ACT, and ADNI with the following levels: 0 (“none”), 1 (“mild”), 2 (“moderate”), and 3 (“severe”). In ROSMAP, diffuse plaques were examined and quantified in five regions (midfrontal cortex, entorhinal cortex, inferior parietal cortex, and hippocampus), then scaled by each region’s standard deviation and averaged. To discretize this continuous variable in ROSMAP participants to a four-level variable, as recorded in the other data sets, we assigned participants a value of 0 (“none”) if their averaged score was equal to 0, 1 (“mild”) if their score was higher than 0 but 0.5, 2 (“moderate”) if their score was between 0.5 and one, and 3 (“severe”) if their score was above 1. These labels roughly corresponded to score quartiles in ROSMAP.

Hippocampal sclerosis is recorded as a binary indicator of the presence or absence of pathology in the NACC NP data set form version 1-9, ROSMAP, and ACT. In versions 10 and 11 of the NACC NP form and in ADNI, HS pathology is recorded as being absent, unilateral, or bilateral. To harmonize, we dichotomized HS pathology as being present if either unilateral or bilateral pathology was indicated.

The Lewy body pathology variable we analyzed had four levels: 0, indicating absent Lewy body pathology in all regions examined or limited to the olfactory bulb; 1, indicating Lewy body pathology limited to the brainstem, including the substantia nigra; 2, indicating Lewy body pathology involving the limbic system or amygdala; and 3, indicating Lewy body involvement in the neocortex. In NACC and ROSMAP, Lewy body pathology was graded in ordinal variables with levels corresponding to the levels in the final harmonized outcome variable analyzed. In ADNI, a single variable was also used to record Lewy body pathology but included separate levels for (1) no pathology and olfactory-predominant and (2) limbic-predominant and amygdala-predominant pathology. These stages were collapsed into levels 0 and 2, respectively, for harmonization. In ACT, separate binary indicator variables were used to indicate presence of Lewy body pathology in each brain region checked. To harmonize, we created a new variable that coded Lewy body pathology stage according to the “highest” stage present in an individual (*e.g.* if pathology was present in both the amygdala and the neocortex, we assigned a value of 3).

## Statistical Analyses

### Single-variant GWAS

We analyzed ordinal endophenotypes using proportional odds logistic mixed-effects models implemented in the POLMM R package [[35](#ref-bi2021)]. We analyzed binary variables similarly with logistic mixed-effects models implemented in the SAIGE R package [[36](#ref-zhou2018)]. Covariates included age at death, sex, cohort, and the first 10 genetic PCs created using the PCA in Related Samples (PC-AiR) method in the GENESIS R package [[37](#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 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 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 p-value of 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 [[32](#ref-chang), [38](#ref-chang2015)].

We performed analyses both in individual cohorts separately and in a pooled data set with all participants. Single-variant analyses were first performed using NACC data. For all lead variants with a p-values , we attempted to validate the association in ROSMAP, ACT, and ADNI separately. Participants were then pooled into a single data set and genome-wide association analyses were performed again using all participants. As noted in the section on harmonization above, LATE-NC data was not available for ACT participants, so validation of LATE-NC-associated variants was not attempted in ACT.

### Re-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, and the and alleles are associated with lower and higher risk of LOAD, respectively, relative to the allele. 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 the hypothesis that genetic loci in the *APOE* region may affect neuropathology risk independently from *APOE* alleles, we re-analyzed variants in Chromosome 19 while adjusting for *APOE* genotype. We limited re-analysis to endophenotypes with at least one significant association signal within the *APOE* locus in the pooled single-variant association analysis described in the previous section. *APOE* genotypes were determined using the rs7412 and rs429358 variants according to the SNPedia online reference [[39](#ref-apoe-s)]. 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.

### Gene-based analyses

Analyses that test for the aggregate effect of all variants in a gene region may increase power for detection of risk genes. We thus performed gene-based analyses for all neuropathology endophenotypes studied. Gene start and end sites were determined using GRCh38 gene regions. Variants were mapped to a gene if they were located within 10 kbp of the gene’s start or end sites. Using an unrelated set of the pooled participant data set, gene-based analyses were performed using MAGMA v1.10 [[40](#ref-leeuw2015)]. The same fixed-effects covariates used in single-variant analyses were used, and 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 to be significantly associated with endophenotypes.

### Principal component outcome analyses

Many neuropathology endophenotypes are highly correlated with one another, and this phenomenon is not reflected in analyses using individual neuropathology endophenotypes as outcomes. Thus, we sought to perform genetic association analyses that could better account for the common co-occurrence of NPE. We first assessed the phenotype correlation of neuropathology endophenotypes used in the pooled GWAS using polychoric correlation and grouped endophenotypes by visually examining a dendrogram based on phenotype-phenotype correlations. Using the psych R package [[41](#ref-revelle2022)], for each group of endophenotypes, we then performed PCA on the matrix of neuropathology endophenotype outcomes in that group. The first PC (PC1) of each group was then used as the outcome variable for GWAS using SAIGE in a two-stage analysis. In the first stage, PC1 was regressed on fixed-effects covariates, with the GRM included as a random effect. Because the residuals of PC1 may not be normally distributed, the residuals of stage one were then rank inverse normalized and regressed on covariates and GRM again in stage two. Score tests were then performed on genetic variants and a saddle-point approximation used to estimate p-values.

### Colocalization analyses

Colocalization analysis is a Bayesian analytical method that seeks to answer whether a single genetic locus drives genotypic-phenotypic associations in two or more traits. We used multiple sources of publicly available summary statistics from external studies as data sources for 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 [[42](#ref-gtexconsortium2017)]. As an additional source of QTL data, we used gene expression and DNA methylation QTL (mQTL) analysis summary statistics from studies using tissue taken from the dorsolateral prefrontal cortex of ROSMAP participants [[43](#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 [[7](#ref-bellenguez2022)].

For each neuropathology endophenotype outcome in our study, we first compiled a list of genetic variants with p-values in the pooled single-variant genome-wide association analysis. 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 [[44](#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 pooled single-variant analysis.

# Results

| Table 1: Participant Demographics | | | | | |
| --- | --- | --- | --- | --- | --- |
|  | NACC | ROSMAP | ACT | ADNI | Overall |
|  | (N=5625) | (N=1183) | (N=616) | (N=39) | (N=7463) |
| Sex |  |  |  |  |  |
| Female | 2809 (49.9%) | 798 (67.5%) | 345 (56.0%) | 8 (20.5%) | 3960 (53.1%) |
| Male | 2816 (50.1%) | 385 (32.5%) | 271 (44.0%) | 31 (79.5%) | 3503 (46.9%) |
| Age of Death |  |  |  |  |  |
| Mean (SD) | 81.3 (9.71) | 89.6 (6.48) | 88.4 (6.68) | 83.2 (7.90) | 83.2 (9.66) |
| Median [Min, Max] | 82.0 [39.0, 111] | 90.1 [66.0, 108] | 89.0 [70.0, 106] | 84.0 [59.0, 97.0] | 84.0 [39.0, 111] |
| APOE e4 alleles |  |  |  |  |  |
| 0 | 2533 (45.0%) | 883 (74.6%) | 443 (71.9%) | 17 (43.6%) | 3876 (51.9%) |
| 1 | 2435 (43.3%) | 280 (23.7%) | 160 (26.0%) | 17 (43.6%) | 2892 (38.8%) |
| 2 | 654 (11.6%) | 20 (1.7%) | 13 (2.1%) | 5 (12.8%) | 692 (9.3%) |
| Missing | 3 (0.1%) | 0 (0%) | 0 (0%) | 0 (0%) | 3 (0.0%) |
| CERAD |  |  |  |  |  |
| None | 519 (9.2%) | 279 (23.6%) | 139 (22.6%) | 8 (20.5%) | 945 (12.7%) |
| Mild | 467 (8.3%) | 103 (8.7%) | 159 (25.8%) | 6 (15.4%) | 735 (9.8%) |
| Moderate | 1007 (17.9%) | 399 (33.7%) | 151 (24.5%) | 2 (5.1%) | 1559 (20.9%) |
| Severe | 3627 (64.5%) | 391 (33.1%) | 167 (27.1%) | 23 (59.0%) | 4208 (56.4%) |
| Missing | 5 (0.1%) | 11 (0.9%) | 0 (0%) | 0 (0%) | 16 (0.2%) |
| Selected NPE are shown in Table 1. All NPE summaries are shown in Supplementary Materials Table S2. | | | | | |
| Key: SD, standard deviation; Min, minimum; Max, maximum. Age of Death variable is integer in NACC, ADNI, and ACT but continuous in ROSMAP. | | | | | |

In total, 7,463 unique participants across all cohorts met our inclusion criteria and were included in our analyses; specifically, autopsy, and genotype data were available for these participants, and they passed our quality control measures for variant missingness, genetic heterogeneity, and ancestry. Included participants consisted of 5,625 participants from NACC; 1,183 from ROSMAP; 616 from ACT; and 39 from ADNI. Participant demographics are shown in **Table 1**. NACC participants died at a younger age (mean age at death 81 years) compared to ROSMAP (90 years) and ACT (88 years) participants, and had more balanced sex ratios, with 50% of NACC participants female vs. 68% and 58% in ROSMAP and ACT, respectively. NACC participants were also more likely to have an *APOE* allele (55%) vs. ROSMAP (25%) or ACT (28%). Across the 11 NPE harmonized across multiple cohorts, NACC and ADNI tended to show more advanced neuropathology than ROSMAP and ACT. Using CERAD neuritic plaque score as an example, 65% of NACC and 59% of ADNI participants were rated as “none” for CERAD score, whereas 33% of ROSMAP and 27% of ACT participants were rated as “severe.” This trend can be seen in most of the NPE harmonized across cohorts with the exceptions of CAA, LATE-NC, and microinfarcts (see **Supplementary Table S2** for summary demographics for all neuropathology endophenotypes).

## Single-variant analysis identifies novel neuropathology-associated mapped to *COL4A1* and *PIK3R5*

In the initial analysis using only NACC participants, variants within the *APOE* region were significantly associated with Braak NFT stage, CERAD score, CAA, and diffuse amyloid plaques. These associations were validated in all of the other data sets at a significance threshold of . No other loci were significantly associated with any neuropathology endophenotypes.

In total, 14 independent () loci had lead variants that met the genome-wide significance threshold of in the main pooled mega-analysis. Of these, four were duplicate signals in the *APOE* region, leaving 10 significant signals across 11 endophenotypes. Novel variant-endophenotype associations included (1) a variant 12 kbp upstream of *COL4A1,* rs2000660, associated with atherosclerosis in the circle of Willis (**Figure 1**) and (2) a *PIK3R5* intronic variant, rs72807981, associated with Braak NFT stage (**Figure 2**). Additionally, one locus with lead variant rs4803778 remained significantly associated with CAA in the *APOE* -adjusted analysis (**Figure 3**). Loci identified that have been previously associated with NPE or dementia included *APOE* with Braak stage, CAA, CERAD score, diffuse amyloid plaques, and LATE-NC; *BIN1* with Braak NFT stage; and *TMEM106B* with HS and LATE-NC (**Table 2**).

| Table 2: Significant NPE-Associated Loci in Mega-Analysis | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 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 | 1.14-1.29 | 3.8e-10 |
| LATE-NC | *TMEM106B* | rs6460900 | 7 | 12,213,462 | G/A | 0.70 | 0.62-0.79 | 7.8e-09 |
| HS | *TMEM106B* | rs4721058 | 7 | 12,227,630 | T/C | 0.67 | 0.59-0.76 | 1.7e-11 |
| Atherosclerosis | *COL4A1* | rs2000660 | 13 | 110,136,094 | A/G | 0.73 | 0.65-0.82 | 3.1e-08 |
| Braak NFT Stage | *PIK3R5* | rs72807981 | 17 | 8,939,344 | G/A | 0.71 | 0.62-0.81 | 3.9e-08 |
| Braak NFT Stage | *APOEe* | rs769449 | 19 | 44,906,745 | A/G | 1.89 | 1.76-2.03 | 2.3e-67 |
| LATE-NC | *APOEe* | rs769449 | 19 | 44,906,745 | A/G | 1.70 | 1.43-2.02 | 4.4e-10 |
| CERAD Score | *APOC1e* | rs12721051 | 19 | 44,918,903 | G/C | 2.12 | 1.96-2.29 | 5.3e-83 |
| CAA | *APOC1e* | rs4420638 | 19 | 44,919,689 | G/A | 2.24 | 2.09-2.4 | 1.3e-112 |
| Diffuse Amyloid Plaques | *APOC1e* | rs4420638 | 19 | 44,919,689 | G/A | 1.81 | 1.67-1.96 | 3.9e-47 |
| CAAd | *CLPTM1e* | rs4803778 | 19 | 44,952,989 | C/T | 1.24 | 1.16-1.32 | 5.8e-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 | | | | | | | | |
| eLocus in APOE region | | | | | | | | |

## Gene-based analysis confirms association between genes in *APOE* region and multiple neuropathology endophenotypes

Gene-level analyses further corroborated our single-variant analyses, 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. *CYP27C1*, located approximately 50 kbp downstream of *BIN1*, was significantly associated with NFT. *TMEM106B* was associated with both HS and LATE-NC.

## Analysis of neuropathology endophenotypic correlations identifies Alzheimer’s-associated and vascular clusters

Visual inspection of a dendrogram (**Supplementary Figure S1**) of neuropathology endophenotypes based on polychoric correlations reveals two primary clusters of NPE: (1) a cluster of AD- or amyloid-associated endophenotypes consisting of Braak NFT stage, CERAD score, diffuse amyloid, and CAA; and (2) a cluster of vascular endophenotypes consisting of atherosclerosis, arteriolosclerosis, gross infarcts, and microinfarcts. PC1 in the AD cluster accounted for 66% of variance in the cluster, while PC1 in the vascular cluster accounted for 50% of the variance in that cluster. Only the *APOE* region was significantly associated with the AD cluster PC1, while no loci were significantly associated with the vascular PC1.

## Colocalization analyses identifies colocalization between CAA and DNA methylation in *APOC2* region

Because CAA was the only NPE to have an associated locus in the *APOE* region with a signal independent of *APOE* diplotype, this region was excluded from endophenotype-endophenotype colocalization analysis. Pairs of endophenotypes 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 showed a high probability of colocalization, with both having .

Multiple suggestive NPE loci showed evidence of colocalization with eQTL in GTEx. A locus within *GRN* suggestively associated with HS () colocalized with *GRN* expression (), and the *TMEM106B* locus associated with HS and LATE-NC colocalized with *TMEM106B* eQTL. The CAA-associated *APOC2* locus independent of *APOE* diplotype also colocalized with *APOE* and *APOC2* eQTL in GTEx () and with four mQTL in ROSMAP ( for each of the four mQTL, see **Figure 3**).

# Discussion

We performed GWAS and downstream in silico functional analyses for eleven neuropathology endophenotypes across four high-quality neuropathology cohorts with a maximum sample size of 7,463 participants. Our work builds on previous genetic association studies of NPE, Alzheimer’s, and related dementias and provides another attempt to better understand the complex associations between different neuropathologies and genetic risk factors [[7](#ref-bellenguez2022), [10](#ref-katsumata2022), [23](#ref-beecham2014)–[25](#ref-vattathil2021), [45](#ref-dugan2021)]. Our study has several important limitations, including sample size and heterogeneity in study design. However, we discovered intriguing new loci mapped to *COL4A1* and *PIK3R5* associated with atherosclerosis in the circle of Willis and Braak stage for neurofibrillary tangles, respectively. We also investigated known loci in *BIN1*, *APOE*, and *TMEM106B* to provide additional context on their associations with multiple NPE.

## Limitations and conclusion

There are limitations to the approach we took in this study. The four data sources we used, while all of high quality, used different study designs and recruitment strategies. NACC participant data arises from recruitment at each ADRC and likely consists of a more clinical population, whereas MAP and ACT participants may be more representative of the broader aged communities from which they are recruited. ROS participants include Catholic brothers, sisters, and priests and may represent a population more highly educated than the United States as a whole. These differences are reflected in the relative frequencies and distributions of NPE between cohorts, with NACC and ADNI tending to show more severe pathology than ACT and ROSMAP, and their mean ages of death were markedly lower (**Tables 1 and S2**). Additionally, there is likely substantial heterogeneity in the way that neuropathology data were collected and graded both within and between studies. We review these potential limitations briefly in our previous GWAS of brain arteriolosclerosis [[24](#ref-shade2022)].

Additionally, while recruitment, autopsy, and genotyping of the cohorts used is ongoing, the sample sizes available for genome-wide investigation of NPE genetic risk is small relative to their clinical counterparts. A recent GWAS of AD included over 111,000 clinically diagnosed or “proxy” (based on family history) AD cases and nearly 700,000 controls. In contrast, our largest analyzed sample size for CERAD score totaled 7,447 participants, while LATE-NC totaled only 2,224. These sample reflect immense progress in autopsy cohorts from even eight years ago and will only continue to increase, but available sample sizes still pose limitations on power of GWAS and similar ’omics approaches to studying disease risk using neuropathology endophenotypes. A strength of our study is that data on multiple neuropathology is available for each participant, allowing for multivariate approaches to studying NPE. However, as our PCA-based multivariate analyses did not identify any risk loci not identified in single-outcome analyses, this approach alone is not sufficient overcome sample size limitations.

## *COL4A1* and atherosclerosis

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 (odds ratio , ). *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. 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 [[48](#ref-steffensen2018)]. The *COL4A1/COL4A2* locus has also been found to be associated with rare familial cerebrovascular diseases and lacunar ischemic stroke[[49](#ref-blevins2021), [50](#ref-rannikmae2017)]. In a recent GWAS, rs2000660 in particular was a risk variant for migraines [[51](#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 to vascular disease is possibly related to disruption of the extracellular matrix [[48](#ref-steffensen2018)].

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*. 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 [[25](#ref-vattathil2021)]. However, the sample size of the previous study was significantly smaller than the one used in the present study (1,325 vs 6,959). Indeed, in the present study, rs2000660 reached genome-wide significance only in the pooled mega-analysis, though it was nominally significant in the ROSMAP-only validation analysis ().

## *PIK3R5* and Braak NFT stage

The other identified 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 validated 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. There is previous research suggesting that *PIK3R5* may be involved in the development of NFT. One study compared differential mRNA expression in aged adults with Braak NFT stage VI vs. non-demented controls and found *PIK3R5* to be more highly expressed in cases than controls (false-discovery rate ) [[52](#ref-guennewig2021)]. Interestingly, *PIK3R5* is expressed preferentially in microglial cells in humans (see **Figure 2)**, suggesting that its association with neurofibrillary pathology may be immune-mediated [[53](#ref-zhang2016)].

## *APOE* and 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 [[23](#ref-beecham2014), [45](#ref-dugan2021)]. While multiple loci in this region were significantly associated with each of these four NPE, only one CAA-associated locus with lead variant rs4803778 remained significant (, ) when adjusting for *APOE* diplotype. rs4803778 was in high LD () with rs4803779, which was also significantly associated with CAA (, ), is a lead mQTL in ROSMAP for five methylation sites ( for top site cg09555818; see **Figure 3**), and colocalizes with four of these mQTL ( for all four sites). Furthermore, two of these associated methylation sites, including cg09555818, were significantly associated with dementia in the Generation Scotland: Scottish Family Health Study [[54](#ref-walker2020)]. These results indicate that this locus may be associated with CAA through epigenetic regulation. However, this locus is located approximately 30 kbp downstream of the *APOE* coding end site and is situated between the *APOC4*, *APOC2*, and *CLPTM1*. While *APOE* may be the most likely target gene of this locus due to its known association with CAA, this inference is not straightforward. In GTEx, this locus is a stronger eQTL of *APOC2* than *APOE*, although it is strongly significant and colocalizes with expression of both genes in multiple tissues. Future work should examine the regulatory role of the identified methylation sites and test them for association with CAA and other AD NPE.

## HS and LATE-NC

One intronic locus of *TMEM106B* was significantly associated with both HS (, ) and LATE-NC (, ). Additionally, a locus within *GRN* was suggestively associated with HS (, ). Both of these genes have been found previously to be associated with frontotemporal lobar degeneration [[55](#ref-rollinson2011), [56](#ref-ciani2019)], and in a recent GWAS by Bellenguez et al., both of these genes were identified as risk factors for LOAD [[57](#ref-bellenguez2022a)]. We downloaded available summary statistics from the GWAS Catalog and performed additional colocalization analyses at these loci to investigate if these loci were shared between clinical LOAD phenotypes and NPE [[58](#ref-buniello2019)]. We found that HS () and LATE-NC () colocalized with ADRD at both the *TMEM106B* (HS and LATE-NC both ) and *GRN* (HS , LATE-NC ) loci (**Figure 4**). HS and LATE-NC also colocalized with each other at these loci, as discussed above. These results indicate that HS, LATE-NC, and LOAD likely share causal loci at these genes.

## *BIN1*, neurofibrillary tangles, and neuritic plaques

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 (, ). In our colocalization analysis, we verified 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 [[7](#ref-bellenguez2022)]. However, 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. In one study, researchers identified that *BIN1* variant rs744373 is associated with tau-PET but not amyloid-PET levels [[59](#ref-franzmeier2019)]. In another study, authors found that *BIN1* protein expression is correlated with NFT pathology but not with diffuse or neuritic plaques [[60](#ref-holler2014)]. As neuritic plaques contain dying nerve cell processes with aberrant tau fibrils identical to those seen in neurofibrillary tangles [[12](#ref-nelson2009)], our findings are also consistent with the hypothesis that *BIN1* influences AD risk primarily through tau rather than amyloid pathogenic processes.

# Conclusions

In conclusion, we identified several promising novel loci associated with NPE and validated multiple known risk loci for AD using NPE. We also provided additional context and consideration for relationships between specific risk loci and different NPE. Overall, our study provides additional potential avenues of investigation into the relationship between genomics, Alzheimer’s disease, and neuropathology.

# List of abbreviations

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

# 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/>.

## 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.

The Adult Changes in Thought Study is funded through the National Institute on Aging [grant number U19AG066567].

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. D.A.B. and J.A.S. provided ROSMAP neuropathology dataand 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.

## Acknowledgements

The NACC database is funded by NIA/NIH Grant U01 AG016976. NACC data are contributed by the NIA-funded ADRCs: P30 AG019610 (PI Eric Reiman, MD), P30 AG013846 (PI Neil Kowall, MD), P50 AG008702 (PI Scott Small, MD), P50 AG025688 (PI Allan Levey, MD, PhD), P50 AG047266 (PI Todd Golde, MD, PhD), P30 AG010133 (PI Andrew Saykin, PsyD), P50 AG005146 (PI Marilyn Albert, PhD), P50 AG005134 (PI Bradley Hyman, MD, PhD), P50 AG016574 (PI Ronald Petersen, MD, PhD), P50 AG005138 (PI Mary Sano, PhD), P30 AG008051 (PI Thomas Wisniewski, MD), P30 AG013854 (PI Robert Vassar, PhD), P30 AG008017 (PI Jeffrey Kaye, MD), P30 AG010161 (PI David Bennett, MD), P50 AG047366 (PI Victor Henderson, MD, MS), P30 AG010129 (PI Charles DeCarli, MD), P50 AG016573 (PI Frank LaFerla, PhD), P50 AG005131 (PI James Brewer, MD, PhD), P50 AG023501 (PI Bruce Miller, MD), P30 AG035982 (PI Russell Swerdlow, MD), P30 AG028383 (PI Linda Van Eldik, PhD), P30 AG053760 (PI Henry Paulson, MD, PhD), P30 AG010124 (PI John Trojanowski, MD, PhD), P50 AG005133 (PI Oscar Lopez, MD), P50 AG005142 (PI Helena Chui, MD), P30 AG012300 (PI Roger Rosenberg, MD), P30 AG049638 (PI Suzanne Craft, PhD), P50 AG005136 (PI Thomas Grabowski, MD), P50 AG033514 (PI Sanjay Asthana, MD, FRCP), P50 AG005681 (PI John Morris, MD), P50 AG047270 (PI Stephen Strittmatter, MD, PhD).

# References

1. (2021) 2021 Alzheimer’s disease facts and figures. Alzheimer’s & Dementia: The Journal of the Alzheimer’s Association 17:327–406. <https://doi.org/10.1002/alz.12328>

2. Ridge PG, Mukherjee S, Crane PK, et al (2013) Alzheimer’s Disease: Analyzing the Missing Heritability. PLOS ONE 8:e79771. <https://doi.org/10.1371/journal.pone.0079771>

3. Gatz M, Pedersen NL, Berg S, et al (1997) Heritability for Alzheimer’s Disease: The Study of Dementia in Swedish Twins. The Journals of Gerontology: Series A 52A:M117–M125. <https://doi.org/10.1093/gerona/52A.2.M117>

4. Gatz M, Reynolds CA, Fratiglioni L, et al (2006) Role of genes and environments for explaining alzheimer disease. Archives of General Psychiatry 63:168–174. <https://doi.org/10.1001/archpsyc.63.2.168>

5. Jansen IE, Savage JE, Watanabe K, et al (2019) Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer’s disease risk. Nature Genetics 51:404–413. <https://doi.org/10.1038/s41588-018-0311-9>

6. Wightman DP, Jansen IE, Savage JE, et al (2021) A genome-wide association study with 1,126,563 individuals identifies new risk loci for Alzheimer’s disease. Nature Genetics 53:1276–1282. <https://doi.org/10.1038/s41588-021-00921-z>

7. Bellenguez C, Küçükali F, Jansen IE, et al (2022) New insights into the genetic etiology of Alzheimer’s disease and related dementias. Nature Genetics 1–25. <https://doi.org/10.1038/s41588-022-01024-z>

8. Kunkle BW, Grenier-Boley B, Sims R, et al (2019) Genetic meta-analysis of diagnosed Alzheimer’s disease identifies new risk loci and implicates A?, tau, immunity and lipid processing. Nature Genetics 51:414. <https://doi.org/10.1038/s41588-019-0358-2>

9. Farfel JM, Yu L, Buchman AS, et al (2016) Relation of genomic variants for alzheimer disease dementia to common neuropathologies. Neurology 87:489–496. <https://doi.org/10.1212/WNL.0000000000002909>

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

11. Karanth S, Nelson PT, Katsumata Y, et al (2020) Prevalence and Clinical Phenotype of Quadruple Misfolded Proteins in Older Adults. JAMA Neurology 77:1299. <https://doi.org/10.1001/jamaneurol.2020.1741>

12. Nelson PT, Braak H, Markesbery WR (2009) Neuropathology and cognitive impairment in alzheimer disease: A complex but coherent relationship. Journal of neuropathology and experimental neurology 68:1–14. <https://doi.org/10.1097/NEN.0b013e3181919a48>

13. Nelson PT, Dickson DW, Trojanowski JQ, et al (2019) Limbic-predominant age-related TDP-43 encephalopathy (LATE): consensus working group report. Brain 142:1503–1527. <https://doi.org/10.1093/brain/awz099>

14. Farfel JM, Yu L, Boyle PA, et al (2019) Alzheimer’s disease frequency peaks in the tenth decade and is lower afterwards. Acta Neuropathologica Communications 7:104. <https://doi.org/10.1186/s40478-019-0752-0>

15. Brenowitz WD, Monsell SE, Schmitt FA, et al (2014) 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. Journal of Alzheimer’s disease : JAD 39:691–702. <https://doi.org/10.3233/JAD-131880>

16. Skrobot OA, Attems J, Esiri M, et al (2016) Vascular cognitive impairment neuropathology guidelines (VCING): The contribution of cerebrovascular pathology to cognitive impairment. Brain 139:2957–2969. <https://doi.org/10.1093/brain/aww214>

17. Weber SA, Patel RK, Lutsep HL (2018) Cerebral amyloid angiopathy: diagnosis and potential therapies. Expert Review of Neurotherapeutics 18:503–513. <https://doi.org/10.1080/14737175.2018.1480938>

18. Smith EE, Schneider JA, Wardlaw JM, Greenberg SM (2012) Cerebral microinfarcts: The invisible lesions. Lancet Neurology 11:272–282. <https://doi.org/10.1016/S1474-4422(11)70307-6>

19. Arvanitakis Z, Capuano AW, Leurgans SE, et al (2017) The Relationship of Cerebral Vessel Pathology to Brain Microinfarcts. Brain Pathology (Zurich, Switzerland) 27:77–85. <https://doi.org/10.1111/bpa.12365>

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

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

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

23. Beecham GW, Hamilton K, Naj AC, et al (2014) Genome-Wide Association Meta-analysis of Neuropathologic Features of Alzheimer’s Disease and Related Dementias. PLOS Genetics 10:e1004606. <https://doi.org/10.1371/journal.pgen.1004606>

24. Shade LM, Katsumata Y, Hohman TJ, et al (2022) Genome-wide association study of brain arteriolosclerosis. Journal of Cerebral Blood Flow & Metabolism 0271678X211066299. <https://doi.org/10.1177/0271678X211066299>

25. Vattathil SM, Liu Y, Harerimana NV, et al (2021) A Genetic Study of Cerebral Atherosclerosis Reveals Novel Associations with NTNG1 and CNOT3. Genes 12:815. <https://doi.org/10.3390/genes12060815>

26. Farrell K, Kim S, Han N, et al (2022) Genome-wide association study and functional validation implicates JADE1 in tauopathy. Acta Neuropathologica 143:33–53. <https://doi.org/10.1007/s00401-021-02379-z>

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

28. Bennett DA, Buchman AS, Boyle PA, et al (2018) Religious orders study and rush memory and aging project. Journal of Alzheimer’s disease : JAD 64:S161–S189. <https://doi.org/10.3233/JAD-179939>

29. Kukull WA, Higdon R, Bowen JD, et al (2002) Dementia and Alzheimer Disease Incidence: A Prospective Cohort Study. Archives of Neurology 59:1737. <https://doi.org/10.1001/archneur.59.11.1737>

30. Taliun D, Harris DN, Kessler MD, et al (2021) Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. Nature 590:290–299. <https://doi.org/10.1038/s41586-021-03205-y>

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

32. Chang CC, Purcell S [Plink 1.9](https://www.cog-genomics.org/plink/1.9/)

33. The 1000 Genomes Project Consortium (2012) An integrated map of genetic variation from 1,092 human genomes. Nature 491:56–65. <https://doi.org/10.1038/nature11632>

34. Braak H, Braak E (1991) Neuropathological stageing of Alzheimer-related changes. Acta Neuropathologica 82:239–259. <https://doi.org/10.1007/BF00308809>

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

36. Zhou W, Nielsen JB, Fritsche LG, et al (2018) Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. Nature Genetics 50:1335–1341. <https://doi.org/10.1038/s41588-018-0184-y>

37. Conomos MP, Miller MB, Thornton TA (2015) Robust Inference of Population Structure for Ancestry Prediction and Correction of Stratification in the Presence of Relatedness. Genetic Epidemiology 39:276–293. <https://doi.org/10.1002/gepi.21896>

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

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

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

41. Revelle W (2022) [Psych: Procedures for psychological, psychometric, and personality research](https://CRAN.R-project.org/package=psych). Evanston, Illinois

42. GTEx Consortium, Laboratory, Data Analysis &Coordinating Center (LDACC)Analysis Working Group, Statistical Methods groupsAnalysis Working Group, et al (2017) Genetic effects on gene expression across human tissues. Nature 550:204–213. <https://doi.org/10.1038/nature24277>

43. Ng B, White CC, Klein H-U, et al (2017) An xQTL map integrates the genetic architecture of the human brain’s transcriptome and epigenome. Nature Neuroscience 20:1418–1426. <https://doi.org/10.1038/nn.4632>

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

45. Dugan AJ, Nelson PT, Katsumata Y, et al (2021) 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. Acta Neuropathologica Communications 9:152. <https://doi.org/10.1186/s40478-021-01250-2>

46. Klarin D, Lynch J, Aragam K, et al (2019) Genome-wide association study of peripheral artery disease in the million veteran program. Nature medicine 25:1274–1279. <https://doi.org/10.1038/s41591-019-0492-5>

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

48. Steffensen LB, Rasmussen LM (2018) A role for collagen type IV in cardiovascular disease? American Journal of Physiology-Heart and Circulatory Physiology 315:H610–H625. <https://doi.org/10.1152/ajpheart.00070.2018>

49. Blevins BL, Vinters HV, Love S, et al (2021) Brain arteriolosclerosis. Acta Neuropathologica 141:1–24. <https://doi.org/10.1007/s00401-020-02235-6>

50. Rannikmae K, Sivakumaran V, Millar H, et al (2017) COL4A2 is associated with lacunar ischemic stroke and deep ICH: Meta-analyses among 21,500 cases and 40,600 controls. Neurology 89:1829–1839. <https://doi.org/10.1212/WNL.0000000000004560>

51. Hautakangas H, Winsvold BS, Ruotsalainen SE, et al (2022) Genome-wide analysis of 102,084 migraine cases identifies 123 risk loci and subtype-specific risk alleles. Nature Genetics 54:152–160. <https://doi.org/10.1038/s41588-021-00990-0>

52. Guennewig B, Lim J, Marshall L, et al (2021) Defining early changes in Alzheimer’s disease from RNA sequencing of brain regions differentially affected by pathology. Scientific Reports 11:4865. <https://doi.org/10.1038/s41598-021-83872-z>

53. Zhang Y, Sloan SA, Clarke LE, et al (2016) Purification and Characterization of Progenitor and Mature Human Astrocytes Reveals Transcriptional and Functional Differences with Mouse. Neuron 89:37–53. <https://doi.org/10.1016/j.neuron.2015.11.013>

54. Walker RM, Bermingham ML, Vaher K, et al (2020) Epigenome-wide analyses identify DNA methylation signatures of dementia risk. Alzheimer’s & Dementia: Diagnosis, Assessment & Disease Monitoring 12:e12078. <https://doi.org/10.1002/dad2.12078>

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

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

57. Bellenguez C, Küçükali F, Jansen IE, et al (2022) New insights into the genetic etiology of Alzheimer’s disease and related dementias. Nature Genetics 1–25. <https://doi.org/10.1038/s41588-022-01024-z>

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

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

60. Holler CJ, Davis PR, Beckett TL, et al (2014) Bridging Integrator 1 (BIN1) Protein Expression Increases in the Alzheimer’s Disease Brain and Correlates with Neurofibrillary Tangle Pathology. Journal of Alzheimer’s Disease 42:1221–1227. <https://doi.org/10.3233/JAD-132450>

61. Rj P, Rp W, S S, et al (2010) LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics (Oxford, England) 26: <https://doi.org/10.1093/bioinformatics/btq419>

# Figure captions

**Figure 1: Locus near *COL4A1* associates with cerebral atherosclerosis.** **A)** Regional LocusZoom plot of associated locus [[61](#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 validated 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) [[53](#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.

**Figure 2: Intronic *PIK3R5* locus associates with Braak stage.** **A)** Regional LocusZoom plot of Braak-associated locus. Mb, megabase; uses hg19. **B)** 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 validated in ACT (, ). **C)** Human brain cell-type expression profile of *PIK3R5* in Zang et al. (2016) [[53](#ref-zhang2016)]. *PIK3R5* is primarily expressed in microglia. FPKM, Fragments Per Kilobase of transcript per Million mapped reads.

**Figure 3: Locus in *APOE* 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.

**Figure 4: HS, LATE-NC, and AD colocalize at *TMEM106B* and *GRN* loci.** Genome positions aligned to hg38. **A)** *TMEM106B* locus. Posterior probability of colocalization (PrC) equals 91% for each pair of phenotypes. **B)** *GRN* locus. PrC is greater than 99.9% between HS and AD. PrC equals 90% between LATE-NC and AD. PrC equals 89% between HS and LATE-NC. Key: MB, mega-basepairs; HS, hippocampal sclerosis; AD, clinical or proxy Alzheimer’s disease; LATE-NC, limbic-predominant age-related TDP-43 encephalopathy neuropathologic change.