# Supplementary Material

## Quality control

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 [[1](#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 [[2](#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 [[3](#ref-chang)]. Finally, we merged participants with the 1000 Genomes Phase 3 (1000 Genomes) cohorts [[4](#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).

## NACC Exclusion Criteria

National Alzheimer’s Coordinating Center (NACC) participants were excluded if they had any of the following conditions noted in the NACC Neuropathology Data Set. Variable names and descriptions are taken from <https://files.alz.washington.edu/documentation/rdd-np.pdf>. Variable descriptions may be lightly edited. Participants were not excluding for missing data in any of these fields.

| Supplementary Table 1: NACC Exclusion Criteria | |
| --- | --- |
| NACC variable | description |
| NACCDOWN | Down syndrom |
| NPPDXB | Multiple system atrophy |
| NPPDXE | Malformation of cortical development |
| NPPDXD | Trinucleotide disease (Huntington disease, SCA, other) |
| NPPDXF | Metabolic/storage disorder of any type |
| NPPDXG | White matter disease, leukodystrophy |
| NPPDXH | White matter disease, multiple sclerosis or other demyelinating disease |
| NPPDXI | Contusion/traumatic brain injury of any type, acute |
| NPPDXJ | Contusion/traumatic brain injury of any type, chronic |
| NPPDXK | Neoplasm, primary |
| NPPDXL | Neoplasm, metastatic |
| NPPDXM | Infectious process of any type (encephalitis, abscess, etc.) |
| NPPDXN | Herniation, any site |
| NACCPRIO | Prion disease |
| NPPATH10 | CADASIL (hereditary stroke disorder) |
| NPALSMND | ALS/motor neuron disease (MND) |
| NPFTDTAU | FTLD with tau pathology (FTLD-tau) or other tauopathy |
| NPFTDTDP | FTLD with TDP- 43 pathology (FTLD-TDP) |
| NPOFTD | Other FTLD |
| NPPDXA | Pigment-spheroid degeneration/NBIA |

## Harmonization of NPE

We combined or harmonized 11 neuropathology endophenotypes across the four data sources used. 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”). Braak NFT Stage 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) [[5](#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 [[6](#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).

## Study participant demographics and distributions of endophenotypes

## Supplementary Figure 1: Dendrogram of Neuropathology Endophenotype Polychoric Correlations

## Supplementary Figures 2-12: Manhattan plots of Stage 3 NPE GWAS

## Sensitivity analyses for novel APOE-adjusted CAA risk locus

We performed several sensitivity analyses to further investigate the novel CAA-associated locus on Chromosome 19 that we identified when adjusting for *APOE* diplotype. First, for each of the NACC, ROSMAP, and ACT data sources, we re-analyzed the association between CAA and lead variant rs7247551 from the meta-analysis while stratifying by *APOE* diplotype and visually compared effect sizes across groups. Due to low sample sizes prevent model convergence, APOE carriers (diplotypes ) Then, for each data source, we performed an an association analysis between CAA and rs7247551 that included interaction effects between rs7247551 and *APOE* diplotype. We then performed ANOVA between these models and nested models which did not include interaction effects and performed Chi-Square tests to test whether any interaction terms were significant.

Results from stratified analyses are shown in Supplementary Figures 13-15 below. No interaction effects were significant in any of the data sources (NACC Chi-Square P = 0.84; ROSMAP Chi-Square P = 0.13; ACT Chi-Square P = 0.91). These results show that the association between rs7247551 and CAA does not significantly differ based on *APOE* diplotype in any of the three data sources used.

1. 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>

2. 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>

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

4. 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>

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

6. 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>