# Introduction

Genome-wide association studies (GWAS) have to date identified over 70 genetic risk loci for Alzheimer’s Disease (AD).[1](#ref-jansen2019)–[3](#ref-bellenguez2022) Most AD GWAS have focused on clinical AD; however, clinical AD does not always correspond with AD neuropathology and vice-versa.[4](#ref-nelson2012) We hypothesized that investigating the genetic architecture of multiple neuropathological endophenotypes (NPE) would:

1. identify novel genetic risk loci for individual NPE
2. identify common sources of genetic risk (pleiotropy) between NPE
3. provide evidence that the effects of clinical AD genetic risk loci may be mediated by specific NPE.

# Methods

## Participants

We used genotype and NPE data from four aged autopsy cohorts: National Alzheimer’s Coordinating Center (NACC), the Religious Orders Study and the Memory and Aging Project (ROSMAP), Adult Changes in Thought (ACT), and the AD Neuroimaging Initiative (ADNI). All study participants were de-identified and deceased, and archival samples were exclusively used. 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 availablr, or/or recorded such that subjects cannot be identified.

This analysis used NACC data from 35 National Institute of Aging-funded Alzheimer’s Disease Research Centers (ADRC). Individual ADRC use different recruitment methods and perform autopsies on-site, but participant 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. The NACC Neuropathology (NP) data set based on the first version of this form was originally implemented in 2001[5](#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 invalidate results. These conditions include brain tumors, severe head trauma, and fronto-temporal dementia (see **Supplementary Materials** for full list of variables used).

ROSMAP has been previously described and consists of harmonized data from two longitudinal cohorts: The Religious Orders Study (ROS) and the Memory and Aging Project (MAP).[6](#ref-bennett2018) ROS began in 1994 and has recruited over 1,100 Catholic priests, nuns, and brothers across the United States. MAP started in 1997 and enrolls community members in northeastern Illinois. The ROSMAP NP data used in this study was received from Rush University researchers 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.21 The goal of the study has expanded to include three cohorts and continuous enrollment using the same enrollment criteria and has a current total of 4,960 participants across all three cohorts. The ACT NP data used in this study were obtained from Kaiser Permanente in May 2020.

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 (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early AD. A subset of ADNI participants undergo autopsy and receive neuropathological phenotyping. The ADNI NP data used in this study was downloaded from the ADNI website in October 2021.

## Genotype data and quality control

Quality control and inclusion/exclusion criteria closely follow that used in our previous brain arteriolosclerosis GWAS.[7](#ref-shade2022) Genotype data for all cohorts have undergone imputation using the Trans-’Omics for Precision Medicine (TOPMed) Imputation Server and the TOPMed reference panel.[8](#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 cohort. Firstly, only participants labeled as non-Hispanic white or similar (European, etc.) were used if available. After imputation, duplicate samples and samples without autopsy data available were removed. Genetic variants and participants were iteratively removed under 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). Then, participants with unusually high or low (+/- 3 standard deviations from mean) were removed. Finally, participants were merged with the 1000 Genomes Phase 3 cohorts[9](#ref-the1000genomesprojectconsortium2012), and participants with substantial non-European ancestry, as determined by distance from EUR superpopulation centroid in principal component analysis using the first two principal components, were removed. Some participants had NP and genotype data available in more than one cohort. In these cases, records in the larger cohort were preferentially kept and discarded in other cohorts (*i.e.* NACC > ROSMAP > ACT > ADNI in preference).

## Defining and harmonizing NPE

We note that there are substantial differences in the way that NPE are collected in different cohorts, 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. All code used for harmonization is available at <https://www.github.com/lincoln-shade/np_phewas>, and an overview of the harmonization strategy for each NPE is included in the **Supplemental Materials**.

## Statistical Analyses

### Single-variant GWAS

Ordinal variables were analyzed using proportional odds logistic mixed-effects models using the POLMM R package,[10](#ref-bi2021) while binary variables were analyzed with logistic mixed-effects models using the SAIGE R package.[11](#ref-zhou2018) Covariates included age at death, sex, cohort, and the first 10 genetic principle components created using the PC-AiR method in the GENESIS R package.[12](#ref-conomos2015). Dense genetic relationship matrices (GRM) created using a pruned set of independent () were used to account for relatedness between participants. An additive mode of inheritance was assumed in all analyses. Analyses proceeded in two stages: in stage one, null models with fixed covariates and GRM were fitted using either POLMM or SAIGE. In stage 2, score tests were performed on each variant with saddle-point approximation used to calculate p-values. Variants achieving a p-value of were considered “genome-wide” significant. To identify independent risk loci, variants were clumped using PLINK 1.9[13](#ref-chang) with a linkage-disequilibrium threshold of .

For NPE available in multiple cohorts, single-variant analyses were first performed in NACC participants. Lead variants with p-values exceeding a suggestive threshold of were then tested in ROSMAP, ACT, and/or ADNI as available for attempted validation. Participants were then pooled for genome-wide mega-analysis.

To examine whether variants in the APOE region independent of haplotypes influence NPE risk, we re-analyzed Chromosome 19 using APOE diplotype as an additional covariate for NPE with significant association signals within the APOE locus. APOE diplotypes were determined using the rs7412 and rs429358 variants.

### Multivariate analyses

Because many NPE have high phenotype correlation that single-outcome association analysis cannot take into account, we sought to perform genetic association analyses that could better account for the common co-occurrence of of NPE. We first assessed the phenotype correlation of NPE used in the pooled GWAS using polychoric correlation and grouped NPE visually using dendrograms.&&&

### Colocalization analyses

All variants in each logistic GWAS with p-values were checked for significant quantitative trait locus (QTL) activity in the Genotype-Tissue Expression Project (GTEx) v8 European ancestry data set in 48 tissues and in ROSMAP bulk RNA-seq data in the dorsolateral prefrontal cortex. Colocalization analysis using the coloc R package (cite) was then performed on GWAS loci with significant QTLs. For ordinal variables, cut points widely used in previous studies were chosen to determine case-control proportions. A prior probability of colocalization of was used, and a posterior probability of colocalization (PPH4) was used as a threshold for evidence of colocalization.

To investigate whether shared GWAS signals drive association between multiple NPE, colocalization analysis was also performed on loci with variants exceeding the suggestive threshold of for more than one NPE in the pooled mega-analytic GWAS.

### Conditional and mediation analyses

For genetic loci associated with multiple NPE, we considered mediation hypotheses to test whether the association of the lead variant with one NPE was mediated by another NPE. Regression analyses were run in R using the glm function for binary outcome traits or polr function in the MASS[14](#ref-venables2002) package for ordinal traits. Mediation analyses were performed using the mediation[15](#ref-tingley2014) R package with standard errors estimated using boot strap re-sampling with 1000 simulations. To assess the possibility of reverse causation, mediation analyses were performed with both NPE in each pair as outcomes.

# Results

In total, 7,463 participants across all cohorts underwent autopsy, had genotype data available, and passed quality control measures. This included 5,625 NACC, 1,183 ROSMAP, 616 ACT, and 39 ADNI participants. Participant demographics are shown in **Table 1**. NACC participants were younger (mean age-of-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 12 NPE harmonized across all four cohorts, NACC and ADNI tended to show more advanced neuropathology. Using CERAD neuritic plaque score as an example, 9% of NACC participants were rated as “none” for CERAD score, whereas 24% and 23% of ROSMAP and ACT participants were rated as “none,” respectively. At the other extreme, 65% of NACC participants were rated as having “severe” neuritic plaque pathology, whereas only 33% and 27% of ROSMAP and ACT participants, respectively, were rated as such. This trend can be seen across **XX** of the NPE harmonized across cohorts and is shown in **Supplementary Table S2**.

## Single-variant analysis

In total, fourteen independent () loci had lead variants that exceeded the genome-wide significance threshold of .

# Discussion

# References

1. 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).

2. 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).

3. 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).

4. Nelson, P. T. *et al.* [Correlation of Alzheimer disease neuropathologic changes with cognitive status: a review of the literature](https://doi.org/10.1097/NEN.0b013e31825018f7). *Journal of Neuropathology and Experimental Neurology* **71**, 362–381 (2012).

5. 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).

6. 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).

7. 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).

8. 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).

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

10. 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).

11. 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).

12. 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).

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

14. Venables, W. N. & Ripley, B. D. [*Modern applied statistics with s*](http://www.stats.ox.ac.uk/pub/MASS4). (Springer, 2002).

15. Tingley, D., Yamamoto, T., Hirose, K., Keele, L. & Imai, K. [mediation: R Package for Causal Mediation Analysis](https://doi.org/10.18637/jss.v059.i05). *Journal of Statistical Software* **59**, 1–38 (2014).

# Tables

| 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 for NACC, ADNI, and ACT but continuous in ROSMAP. | | | | | |

| Table 2: Significant NPE-Associated Loci in Mega-Analysis | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Phenotype | Variant | Chromosome | Positiona | Geneb | Minor/major allele | ORc | 95% CI | P |
| Braak Stage | rs6733839 | 2 | 127,135,234 | BIN1 | T/C | 1.21 | 1.14-1.29 | 3.8e-10 |
| LATE-NC | rs6460900 | 7 | 12,213,462 | TMEM106B | G/A | 0.70 | 0.62-0.79 | 7.8e-09 |
| HS | rs4721058 | 7 | 12,227,630 | TMEM106B | T/C | 1.50 | 1.33-1.69 | 1.7e-11 |
| Atherosclerosis | rs2000660 | 13 | 110,136,094 | COL4A1 | A/G | 0.73 | 0.65-0.82 | 3.1e-08 |
| Braak Stage | rs72807981 | 17 | 8,939,344 | PIK3R5 | G/A | 0.71 | 0.62-0.81 | 3.9e-08 |
| Braak Stage | rs769449 | 19 | 44,906,745 | APOEe | A/G | 1.89 | 1.76-2.03 | 2.3e-67 |
| LATE-NC | rs769449 | 19 | 44,906,745 | APOEe | A/G | 1.70 | 1.43-2.02 | 4.4e-10 |
| CERAD Score | rs12721051 | 19 | 44,918,903 | APOC1e | G/C | 2.12 | 1.96-2.29 | 5.3e-83 |
| CAA | rs4420638 | 19 | 44,919,689 | APOC1e | G/A | 2.24 | 2.09-2.4 | 1.3e-112 |
| Diffuse Amyloid Plaques | rs4420638 | 19 | 44,919,689 | APOC1e | G/A | 1.81 | 1.67-1.96 | 3.9e-47 |
| CAAd | rs4803778 | 19 | 44,952,989 | CLPTM1e | C/T | 1.24 | 1.16-1.32 | 5.8e-12 |
| aGenome positions are based on build HG38. | | | | | | | | |
| bClosest protein-coding gene according to GENCODE release 40. | | | | | | | | |
| cORs are with respect to minor allele. | | | | | | | | |
| dResult from APOE diplotype-adjusted analysis | | | | | | | | |
| eLocus in APOE region | | | | | | | | |