# 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

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).

## Participants

NACC participants are recruited nationally at over 30 National Institute of Aging-funded Alzheimer’s Disease Research Centers (ADRC). Different ADRC use different recruitment methods, but participant data at each ADRC is collected using a standard form and aggregated by NACC. Clinical and neuropathology data were taken from the December 2021 freeze of the NACC Universal Data Set (UDS) and Neuropathology Data Set (NDS). Participants were excluded if they did not have autopsy data available or if they were noted in the NDS 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&&&

## Genotyep data and quality Control

Quality control and inclusion/exclusion criteria closely follow that used in our previous brain arteriolosclerosis GWAS.[5](#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.[6](#ref-taliun2021)

1. Retrieve participants labeled as non-Hispanic white or similar (European, etc.) if available.
2. Perform imputation on TOPMed Imputation Server (ROSMAP and ADNI had already been imputed using same reference panel by collaborators, so this step was skipped for those cohorts).
3. Covert VCF files to PLINK file set, keeping variants will minor allele frequencies (MAF) of 0.1% or greater.
4. Merge sub-cohorts of same study (e.g. NACC cohorts 1-12 and ACT 1-3).
5. Remove duplicate participants in each study.
6. Remove participants without autopsy data available.
7. Iteratively remove variants and participants with excess missingness until no variant is missing in more than 5% of participants and no participant is missing more than 5% of variants.
8. Remove participants with unusually high/low heterozygosity (+/- 3 SD cutoff).
9. Merge with 1000 Genomes Phase 3 cohorts[7](#ref-the1000genomesprojectconsortium2012) and remove participants without primarily European ancestry as determined by principal component analysis.
10. Remove participants with rare neurological diseases that substantially raise the likelihood that present NPE could be due to unusual pathological processes (NACC only, as the other cohorts did not have this information available).
11. Remove duplicate participants between studies, preferentially keeping the observation in the larger cohort (*i.e.* NACC > ROSMAP > ACT > ADNI in preference).
12. Merge cohorts.
13. Harmonize NPE variables to analyze.
14. Create list of related ( 2nd-degree relatives) individuals using KING[8](#ref-manichaikul2010) to remove for each NPE variable, randomly keeping one participant in each related cluster.
15. Create genetic principal components using the PC-AiR method in the GENESIS package in R 4.1.[9](#ref-conomos2015)

## Defining 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. Dense genetic relationship matrices (GRM) 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[12](#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) > 80% 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 package for ordinal traits. Mediation analyses were performed using the mediation[13](#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 four cohorts used passed 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.

| Table 1: Participant Demographics | | | | | |
| --- | --- | --- | --- | --- | --- |
|  | NACC | ROSMAP | ACT | ADNI | Overall |
|  | (N=5625) | (N=1183) | (N=616) | (N=39) | (N=7463) |
| msex |  |  |  |  |  |
| 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\_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 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. | | | | | |

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. 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).

6. 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).

7. 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).

8. Manichaikul, A. *et al.* [Robust relationship inference in genome-wide association studies](https://doi.org/10.1093/bioinformatics/btq559). *Bioinformatics* **26**, 2867–2873 (2010).

9. 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).

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

13. 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).