# Multi-cohort cross-trait polygenic endophenotype prediction in Alzheimer's disease reveals distinct multi-level phenotypic and biomarker subtypes

William F. Li[1,2,3,†], David A. Bennett[4,5], Manolis Kellis[1,2,*], Yosuke Tanigawa[1,2,6,*]

[1]Computer Science and Artificial Intelligence Laboratory, MIT; [2]Broad Institute of MIT and Harvard; [3]Harvard/MIT MD-PhD Program, Harvard Medical School;
[4]Rush Alzheimer's Disease Center, Rush University Medical Center; [5]Department of Neurological Sciences, Rush University Medical Center; [6]Department of Bioengineering, UCLA;
[†]william_li@hms.harvard.edu; *manoli@mit.edu, tanigawa@ucla.edu

## Overview of the study

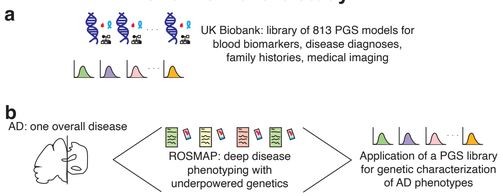


**Figure 1. Study overview. (a)** PGS models previously trained on the UK Biobank. **(b)** AD phenotypic heterogeneity revealed by ROSMAP.

- We integrated 713 PGS models from the UK Biobank with 37 AD clinicopathological variables derived across 1678 individuals in the ROSMAP aging cohort to study the genetics of AD phenotypic heterogeneity.
- We identified 272 FDR<0.1 associations between 37 AD phenotypes and 12 PGS. We used patterns of association to group AD phenotypes into 3 clusters for cognition, amyloid pathology, and tangles pathology.
- We show improvement of multi-PGS models over AD PGS or ApoE alone for prediction of cognition and amyloid pathology values.

## Alzheimer's disease presents with rich interindividual phenotypic heterogeneity

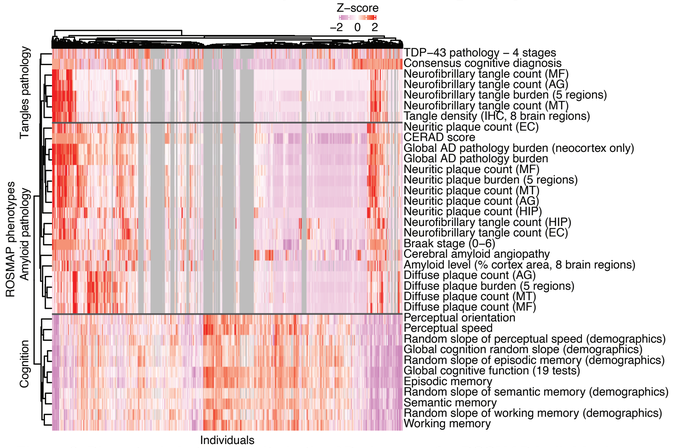


**Figure 2. Phenotypic heterogeneity in the AD phenome for ROSMAP individuals.** 1678 individuals, 36 AD-related clinical and pathological phenotypes, with three overarching phenotype clusters.

TDP-43, transactive response DNA binding protein 43; MF, midfrontal cortex; AG, inferior parietal cortex; MT, midtemporal cortex; IHC, immunohistochemistry; EC, entorhinal cortex; CERAD, Consortium to Establish a Registry for Alzheimer's Disease12; HIP, hippocampus; "(demographics)", variable is controlled for demographics.

## Two-stage polygenic association identifies genetic predictors of AD phenotypes

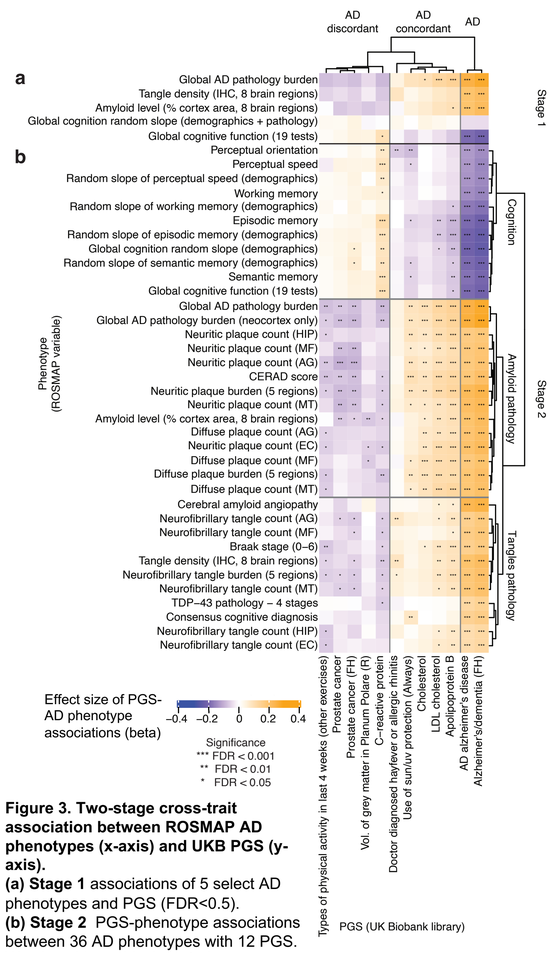


**Figure 3. Two-stage cross-trait association between ROSMAP AD phenotypes (x-axis) and UKB PGS (y-axis).**
**(a) Stage 1** associations of 5 select AD phenotypes and PGS (FDR<0.5).
**(b) Stage 2** PGS-phenotype associations between 36 AD phenotypes with 12 PGS. We show hierarchical clustering as dendrograms.

## Multiple-PGS model enhances phenotype predictions

| | Covariates only (no ApoE) | ApoE + covariates | AD PGS | All PGS | PGS PCs |
|---|---|---|---|---|---|
| Global cognitive function (19 tests) | 0.059 | 0.138 | 0.106 | **0.190** | 0.184 |
| Global AD pathology burden | 0.117 | 0.242 | 0.205 | **0.260** | 0.224 |
| Neuritic plaque burden (5 regions) | 0.078 | 0.241 | 0.211 | 0.257 | **0.291** |
| Amyloid level (% cortex area, 8 brain regions) | 0.235 | 0.275 | 0.219 | 0.265 | **0.305** |
| Diffuse plaque burden (5 regions) | 0.069 | 0.149 | 0.168 | 0.149 | **0.194** |
| Tangle density (IHC, 8 brain regions) | 0.061 | **0.159** | 0.063 | 0.027 | 0.109 |

**Table 1. Predictive performance of five gradient-boosted prediction models:**

1. Prediction with only covariates for age, sex, population genetic principal components, and genotyping site.
2. Prediction with covariates and quantified apolipoprotein E genotype.
3. Prediction with covariates and the PGS for AD family history.
4. Prediction with covariates and all 12 select AD PGS from stage 1 of the cross-trait association.
5. Prediction with covariates and the top 8 principal components (PCs) of the AD PGS.

## Multiple-PGS prediction informs clustering of individuals

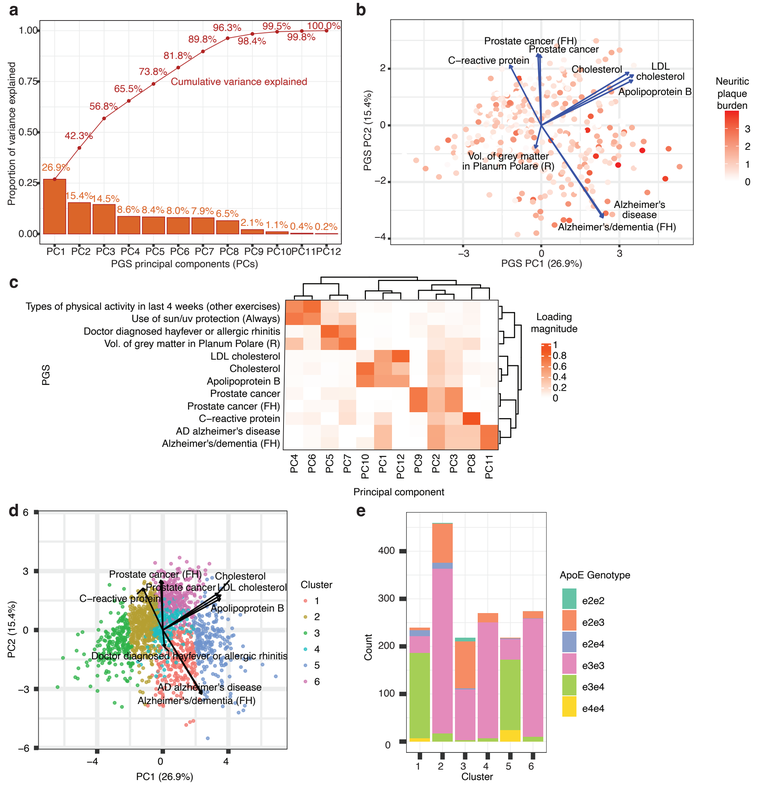


**Figure 4. Individual-level clustering using principal components (PCs) derived from identified PGS.**
**(a)** Variance explained by each PC in the principal component analysis of PGS values for ROSMAP individuals (training set).
**(b)** Biplot of PCs 1 and 2 on the testing set of the predictive performance analysis, colored by neuritic plaque values.
**(c)** The magnitudes of the PCA loadings (color) for each pair of PC (x-axis) and PGS (y-axis).
**(d)** We cluster individuals by PGS PCs**.** We show a biplot of PCs 1 and 2 for all ROSMAP individuals, colored by cluster assignments.
**(e)** We show the prevalence of each ApoE genotype within each cluster.

## References and Acknowledgements

1. Tanigawa, Y. et al. Significant sparse polygenic risk scores across 813 traits in UK Biobank. *PLoS Genetics*. 18, e1010105 (2022).
2. Bellenguez, C. et al. New insights into the genetic etiology of Alzheimer's disease and related dementias. *Nature Genetics*. 54, 412–436 (2022).

*Thank you to the NIH for sponsoring this work.*