|  |
| --- |
| A. |
| B. |
| Figure S1: A. Variance explained by the highest 35 PCs, compared to the null expectation of a broken stick model. B. Cumulative proportion of variance explained by all eigenvalues. |

|  |
| --- |
| A. |
| B. |
| C. |
| Figure S2: Results of fitting models of different number of archetypes, using the first 400 PC’s (80% of the variance in original data). A. Variance explained. B. Proportion added to explained variance by increasing number of archetypes. C. Variance around archetype position based on bootstrap |

|  |
| --- |
|  |
| Figure S3. Comparing archetype classification with k-means clustering. A and B are subtypes previously found using k-means clustering. Bar height is proportional to deviation from expectation under the null hypothesis of equal representation (i.e, no enrichment), tested using the chi2 statistic. |

|  |
| --- |
|  |
| Figure S4. Association of CERAD score with sex across the archetypes. Bar height is proportional to deviation from expectation under the null hypothesis of equal representation (i.e, no enrichment), tested using the chi2 statistic. Disease progression proceeds from right to left. |

|  |
| --- |
|  |
| Figure S5. Association of Apoe4 status with sex across the archetypes. Bar height is proportional to deviation from expectation under the null hypothesis of equal representation (i.e, no enrichment), tested using the chi2 statistic. Disease risk proceeds from left to right: 0-homozygotes or heterozygotes without the e4 allele, 1-heterozygotes with e4 and any other allele, 2-homozygotes for e4. |

|  |
| --- |
| A. |
| B. |
| C. |
| Figure S6. Association of pseudotime estimates with distance from archetypes. |

|  |
| --- |
| A. |
| B. |
| C. |
| D. |
| Figure S7. QQ plots of p-values from single-variant association mapping of our four new phenotypes: distance of each sample from archetype 1 (A), archetype 2 (B), and archetype 3 (C), and assignment of each sample to the nearest archetype (D). |

|  |
| --- |
| A. |
| B. |
| C. |
| D. |
| Figure S8. Manhattan plots of p-values from single-variant association mapping of our four new phenotypes: distance of each sample from archetype 1 (A), archetype 2 (B), and archetype 3 (C), and assignment of each sample to the nearest archetype (D). |

|  |
| --- |
|  |
| Figure S9. Network representation of GWAS results for different phenotypes illustrating overlap among phenotypes in their genetic architecture. Phenotypes include distance from each archetype, assignment to nearest archetype, previously-published subtypes (using k-means clustering) and submodule based on the same gene expression data, CERAD and Braak scores, and cognitive diagnosis. Loci are represented as circles. Edges connect genetic loci to phenotypes. Loci for the submodules are shown only if they overlap with at least one of the other phenotypes. |