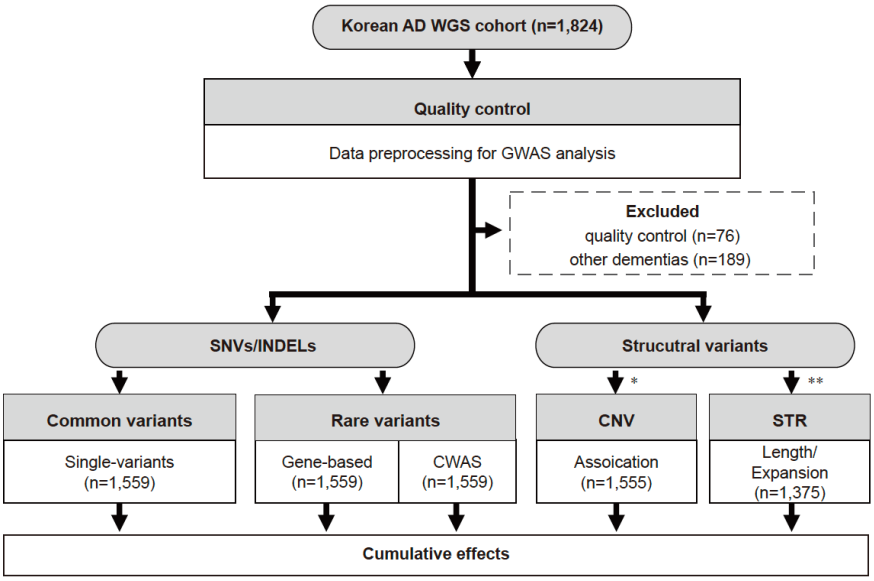
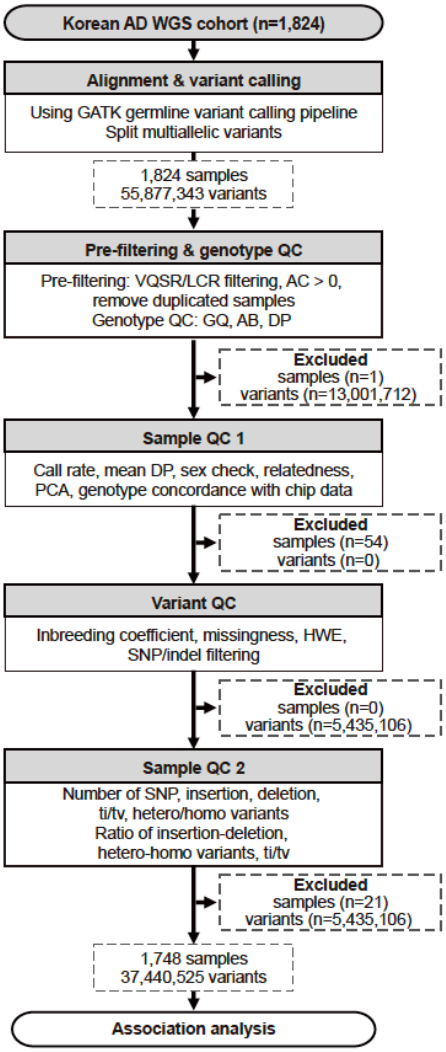
Supplementary Figures



Supplementary Fig. 1 Overview of the study design. A framework was developed to identify genetic factors for AD in the Korean population. Alignment, variants calling, and quality control were performed and total 1,559 individuals were included in study. For single variants, single-variant association analysis was conducted for common variants, along with gene-based association analysis and CWAS for rare variants. For structural variants, CNVs and STRs association analysis were conducted. Through these analyses, prioritized genes associated with AD or Aβ in the Korean population were identified, and their characteristics were elucidated using single-cell and functional analysis. * Additional exclusion of samples that failed to call CNV. **Additional quality control for STR, excluding sequencing batch 3 and two additional samples with depth discrepancies in STR calling results. n indicates the number of independent participants.

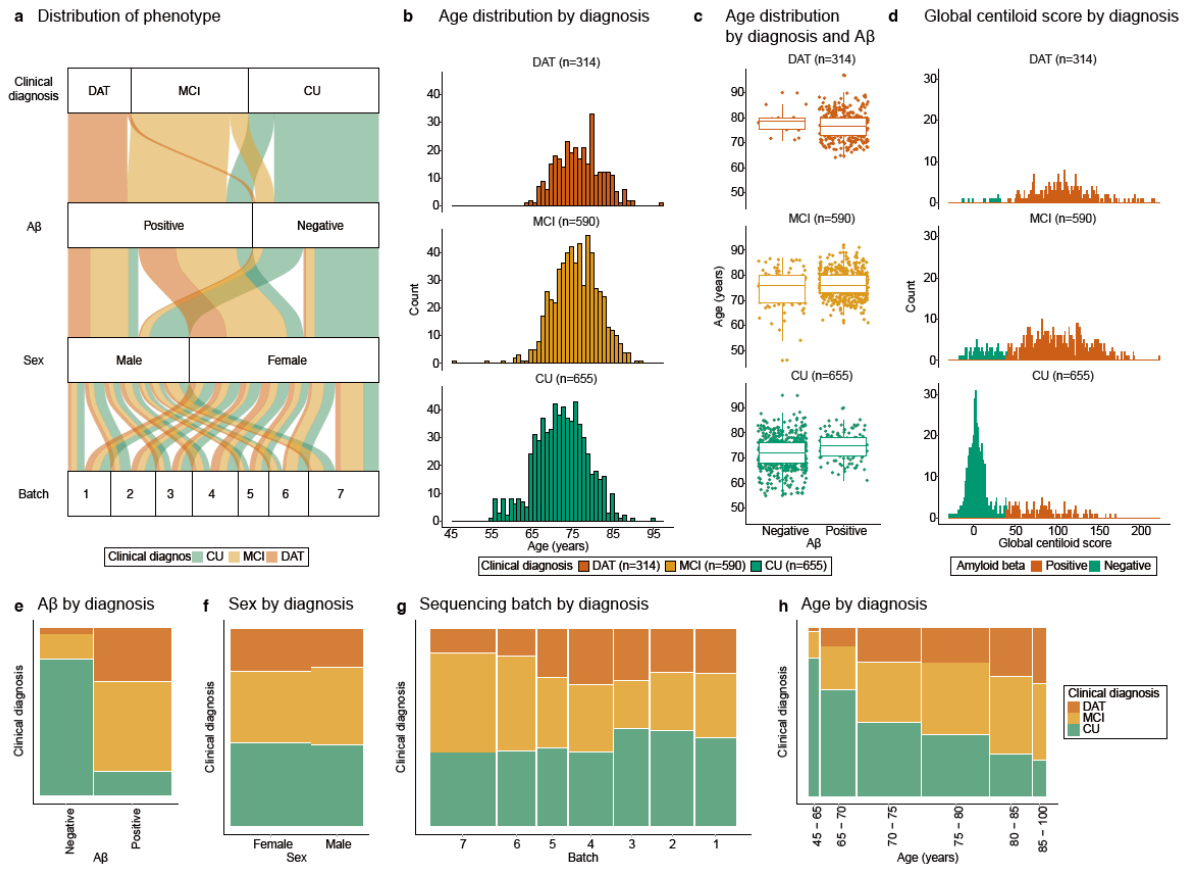
CWAS, category-wide association study; STR, short tandem repeats; CNV, copy number variation; AD, Alzheimer's disease; Aβ, amyloid beta.



Supplementary Fig. 2 Overview of the QC pipeline. A QC pipeline was established to eliminate low-quality samples and variants. Initially, alignment and variant calling were conducted using the GATK germline variant calling pipeline. Subsequently, pre-filtering and genotype QC, including VQSR and LCR region filtering, were performed prior to sample QC. Then, step 1 sample QC, variant QC, and step 2 sample QC were performed sequentially. After QC procedures, samples and variants that passed QC criteria were utilized for association analysis. n indicates the number of independent participants or variants.

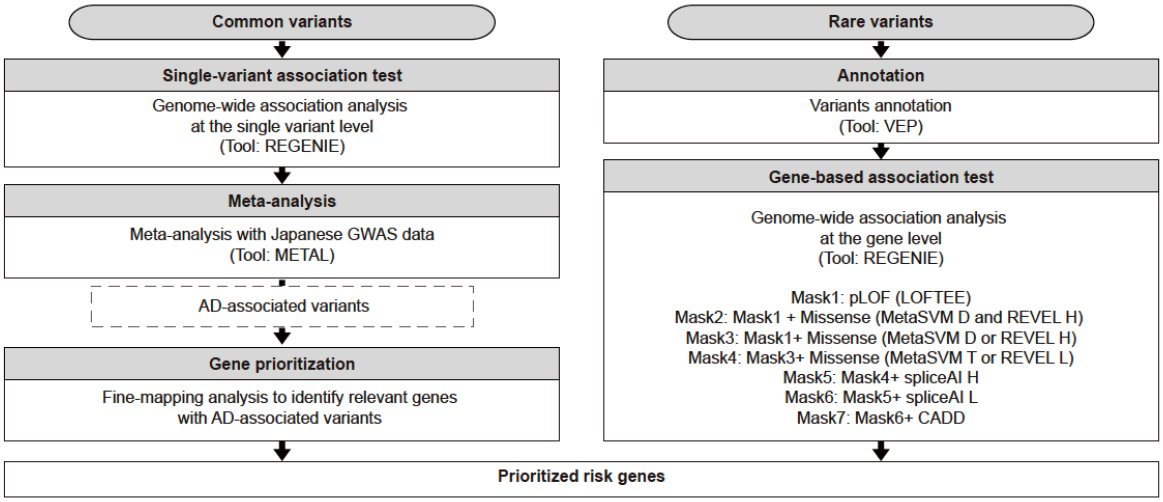
AD, Alzheimer's disease; WGS, whole genome sequencing; GATK, Genome Analysis Toolkit; QC, quality control; VQSR, Variant Quality Score Recalibration; LCR, low

complexity region; GQ, genotype quality; AB, allele balance; DP, depth; PCA, principal component analysis; HWE, Hardy-Weinberg equilibrium; Ti/Tv, transition/transversion.



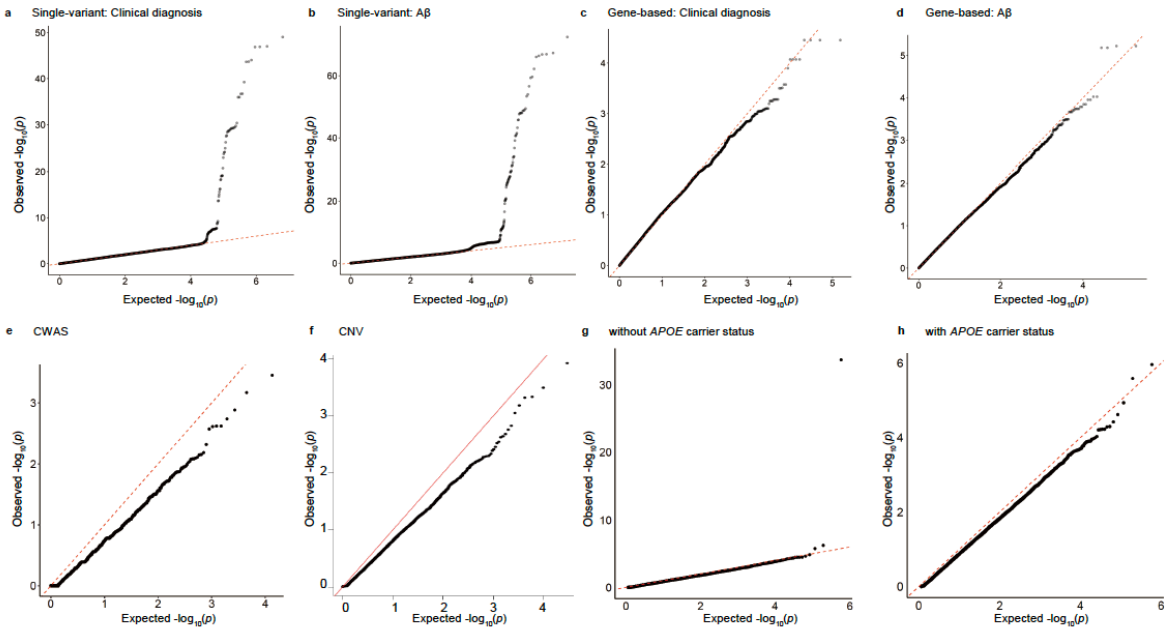
Supplementary Fig. 3 Phenotype distribution in the Korean AD cohort. a, Alluvial plot showing the distribution of phenotypes using clinical diagnosis, Aβ, sex, and sequencing batch. **b,** Age distribution by clinical diagnosis. **c,** Age distribution by clinical diagnosis and Aβ positivity. **d**, Distribution of global centiloid score categorized by diagnosis, color-coded based on Aβ positivity, with the threshold set at global centiloid of 40. **e-h,** Distribution of Aβ positivity (**e**), sex (**f**), sequencing batch (**g**), and age (**h**) by clinical diagnosis. The width of each bar corresponds to the sample size for each category. n indicates the number of independent participants.

DAT, dementia of Alzheimer's type; MCI, mild cognitive impairment; CU, cognitively unimpaired; Aβ, amyloid beta; AD, Alzheimer's disease.



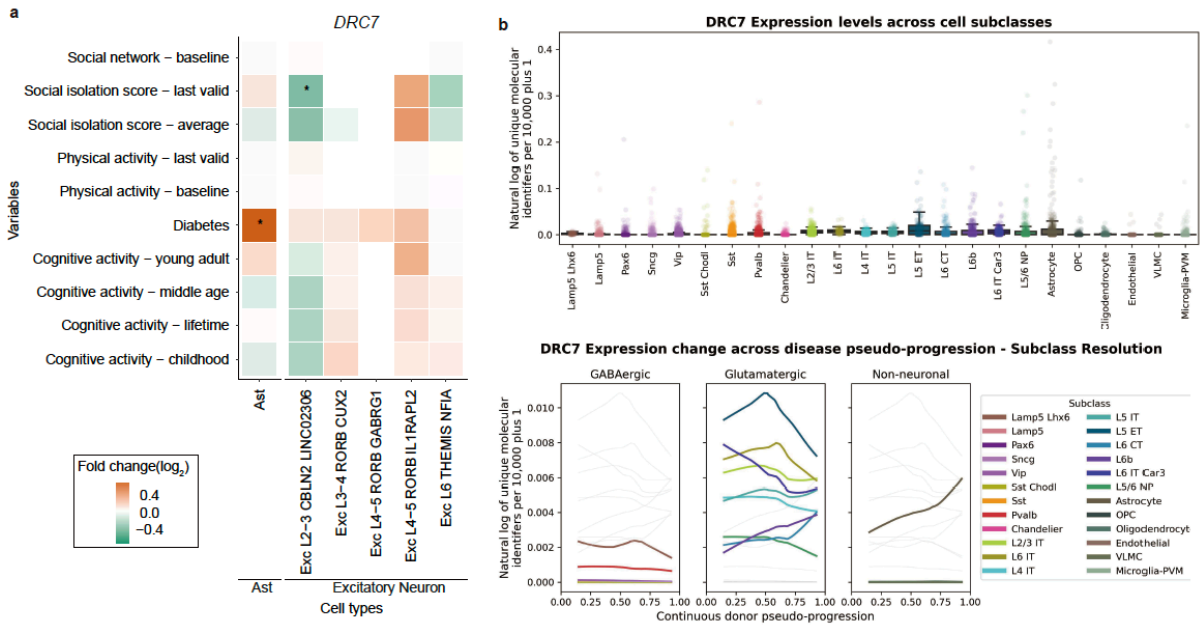
Supplementary Fig. 4 Overview of single-variant and gene-based association tests. A GWAS pipeline was conducted to identify variants and genes associated with AD and Aβ. For common variants, single-variant association tests were performed using REGENIE. Additionally, for clinical diagnosis, a meta-analysis with Japanese GWAS summary statistics was performed using METAL. Gene prioritization was conducted to identify relevant genes with significant loci. For rare variants, variants were annotated using VEP, and gene-based association tests were conducted using REGENIE with seven defined sets of variants (masks). Following these analyses, prioritized risk genes were identified.

GWAS, genome-wide association study; AD, Alzheimer's disease; Aβ, amyloid beta; VEP, Variant Effect Predictor.



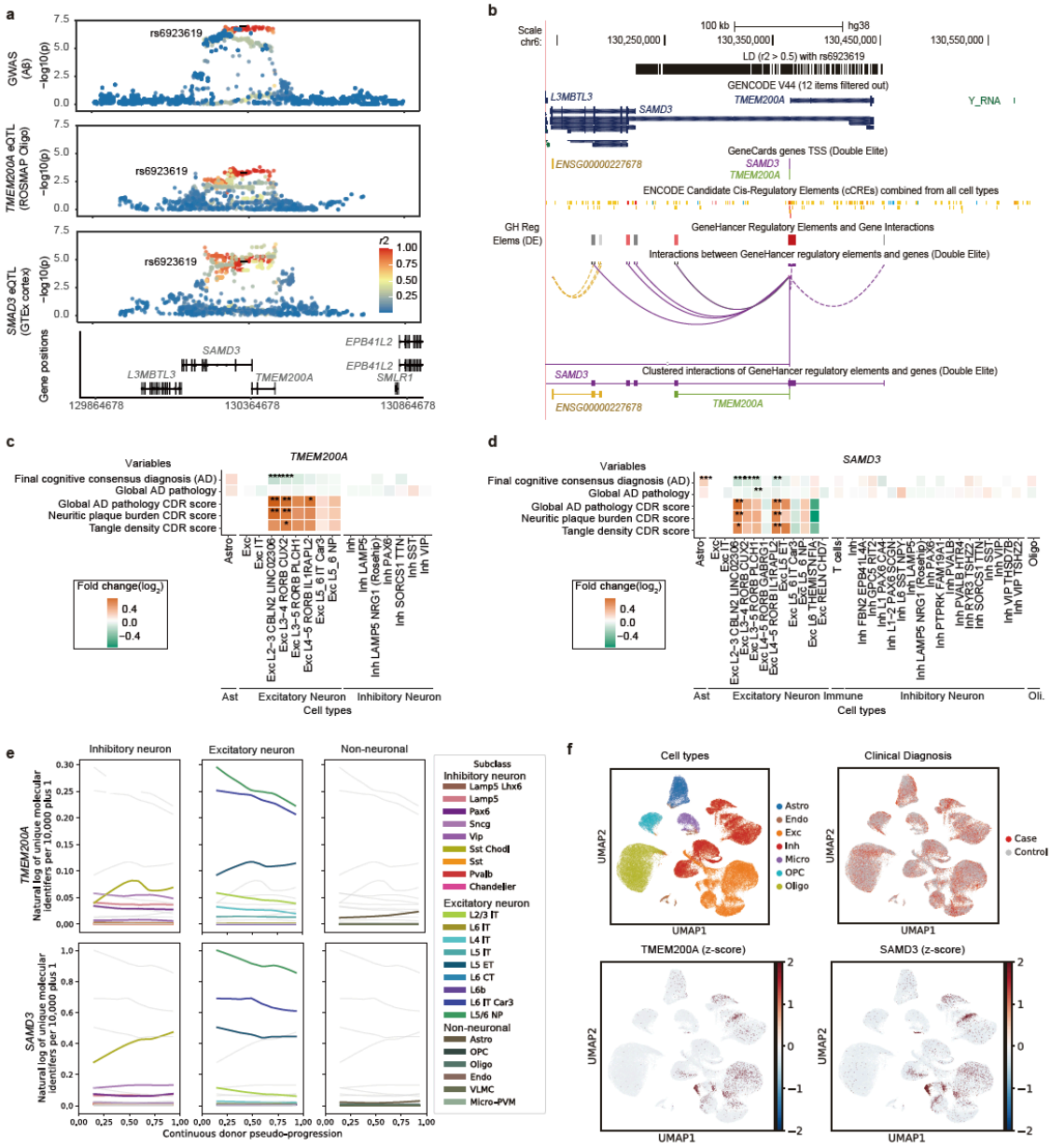
Supplementary Fig. 5 QQ plots for GWAS, CWAS, and STR. QQ plots of observed vs. expected $-\log_{10}(p)$ for GWAS, CWAS, and STR analysis model 1. **a-d,** QQ plots of single-variant association analyses for clinical diagnosis (**a**) and Aβ (**b**), and gene-based association analyses for clinical diagnosis (**c**) and Aβ (**d**). **e**, QQ plot for CWAS. **f**, QQ plot for CNV. **g**, STR analysis under Model 1 without APOE4 carrier status. **h**, QQ plot for STR model with *APOE* ε4 carrier status. Source data are provided as a Source Data file.

QQ plot, quantile-quantile plot; GWAS, Genome-wide association study; CWAS, category-wide association study; CNV, copy number variant; STR, short tandem repeats; Aβ, amyloid beta.



Supplementary Fig. 6 Functional analysis of *DRC7*. a, Heatmap of differential expression analysis for *DRC7* across brain sub-cell types from a previous study by Mathys et al. Significance levels are indicated as * for $p < 0.05$, ** for $p < 0.01$, and *** for $p < 0.001$. **b,** Expression levels of *DRC7* across brain sub-cell types and expression patterns across pseudo-progression from Seattle Alzheimer's Disease Brain Cell Atlas. Upper panel: gene expression across brain sub-cell types. Lower panel: expression change pattern with pseudo-progression. Source data are provided as a Source Data file.

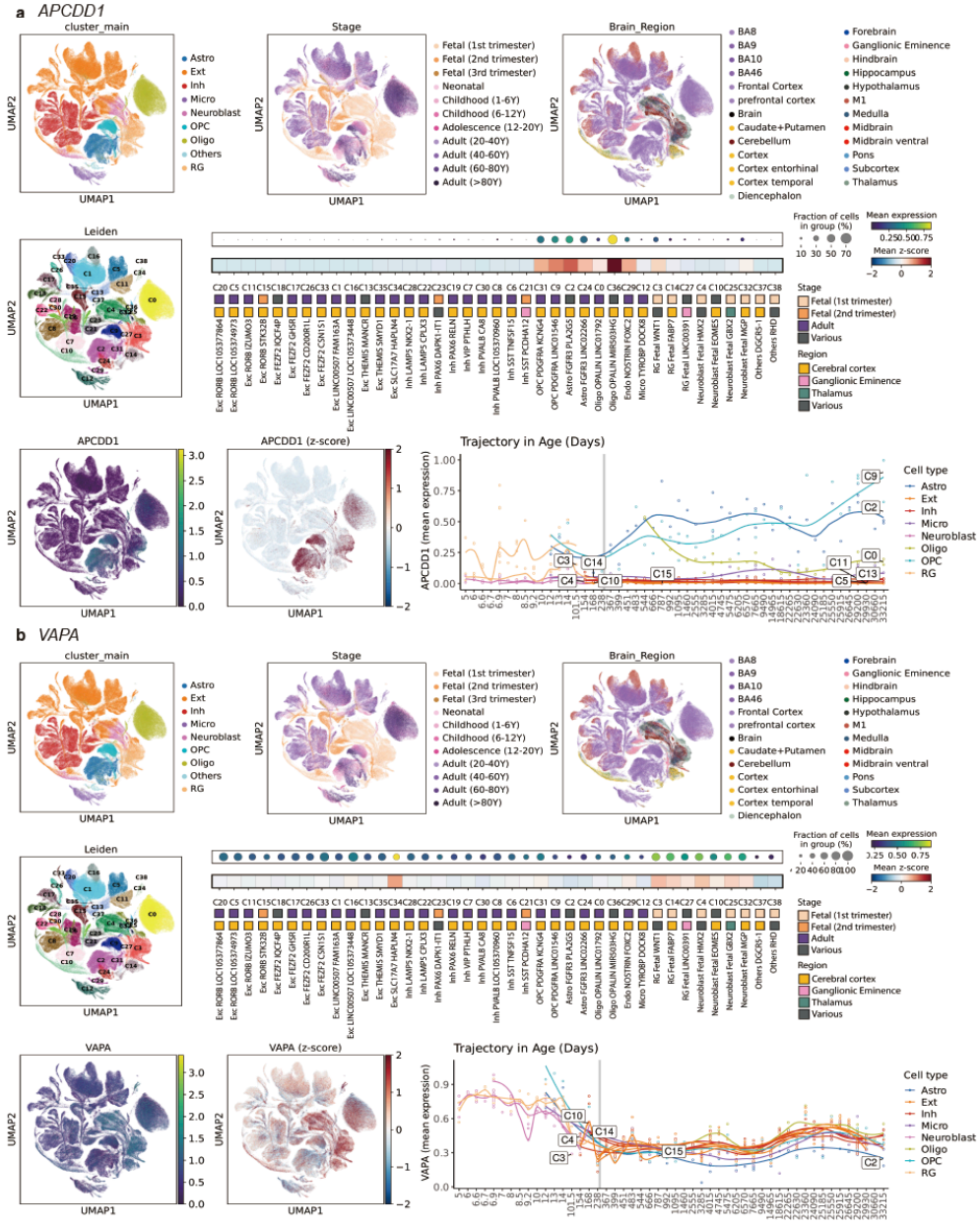
Ast, astrocyte; Ext, excitatory neuron; IT, inhibitory neuron; OPC, oligodendrocyte precursor cell; VLMC, ventral lateral motor cortex



Supplementary Fig. 7 Fine-mapping and statistical colocalizations analysis of the *SAMD3* locus. a, Regional plot of the Aβ GWAS, ROSMAP oligodendrocyte eQTLs for *TMEM200A*, and GTEx cortex eQTLs (v8) for *SAMD3* within a 500 kb window of the lead variant (rs6923619). **b**, Regional plot showing gene regulatory regions with the lead variant and variants in LD with the lead variant (LD $r^2 > 0.5$), gene regions from GENCODE, TSS from GeneCards, cCRE regions from ENCODE, and regulatory elements, interactions, and clusters from GeneHencer. **c,d,** Heatmap of DEGs for *TMEM200A* (**c**) and *SAMD3* (**d**) across brain cell subtypes. Significance

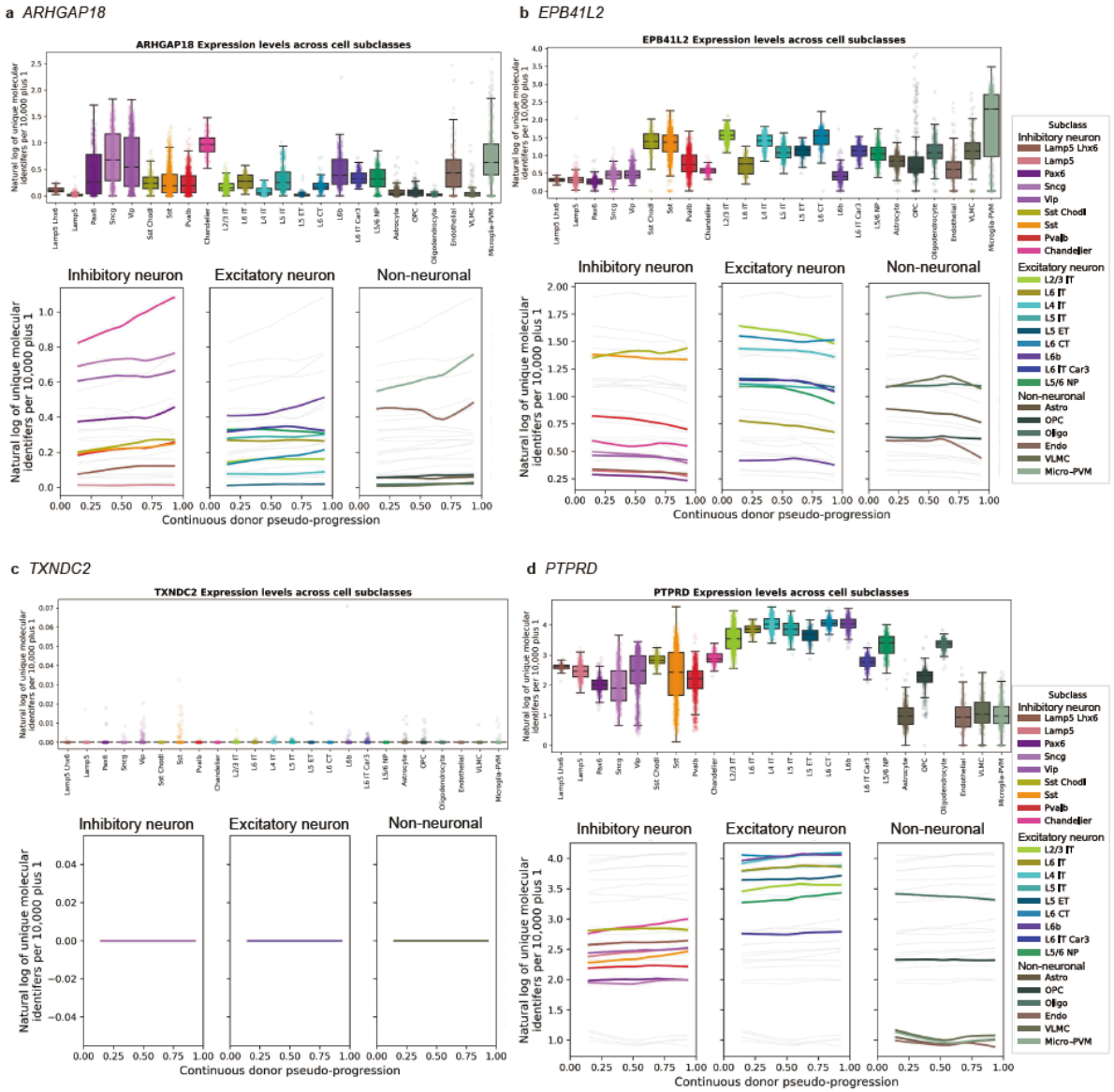
levels are indicated as * for $p < 0.05$, ** for $p < 0.01$, and *** for $p < 0.001$. **e**, *TMEM200A* and *SAMD3* expression across different cell types in response to pseudo-progression from Seattle Alzheimer's Disease Brain Cell Atlas. **f**, UMAP plots using Korean scRNA-seq data, colored based on cell type, clinical diagnosis, and *TMEM200A* and *SAMD3* expression. Source data are provided as a Source Data file.

Aβ, amyloid beta; GWAS, genome-wide association study; LD, linkage disequilibrium; eQTL, expression quantitative trait loci; Oligo, oligodendrocyte; TSS, transcription start site; cCRE, candidate cis-regulatory element; AD, Alzheimer's disease; CDR, cognitive decline resilience; Astro, astrocyte; Ext, excitatory neuron; Inh, inhibitory neuron; CAMs, cell adhesion molecules; Micro, microglia; OPC, oligodendrocyte precursor cell; Endo, endothelial cells; SMC, smooth muscle cell; DEG, differential expressed gene; UMAP, Uniform Manifold Approximation and Projection; scRNA, single-cell RNA sequencing.



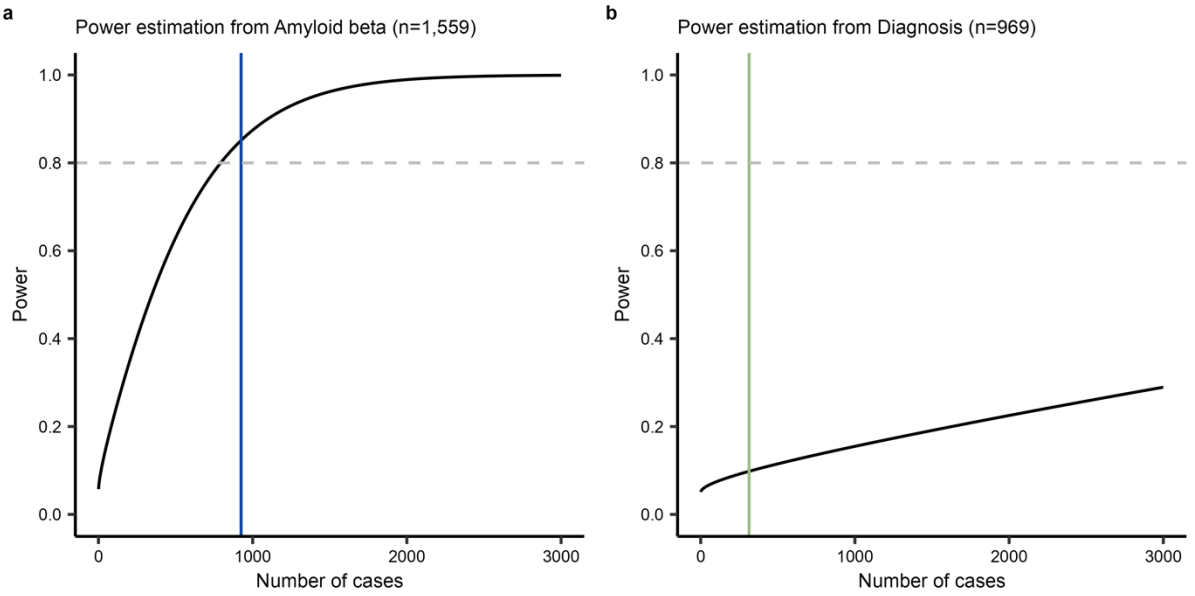
Supplementary Fig. 8 Cell type-specific expression of *APCDD1* and *VAPA* from the BTS Atlas. The expression patterns of *APCDD1* (**a**) and *VAPA* (**b**) the brain transcriptome at the single-cell level (BTS atlas). Upper panel: UMAP plots colored by cell type clustering (left), stage (center), and brain region (right). Middle panel: UMAP plot of Leiden clustering (left), heatmap of gene expression and fraction of cells across sub-cell types (right). Bottom panel: UMAP plots of gene expression colored by mean expression (left) and z-score (center), and expression changes based on trajectory (right).

scRNA, single-cell RNA; UMAP, uniform manifold approximation and projection; Astro, astrocyte; Ext, excitatory neuron; Inh, inhibitory neuron; Micro, microglia; OPC, oligodendrocyte precursor cell; Oligo, oligodendrocyte; RG, radial glia.

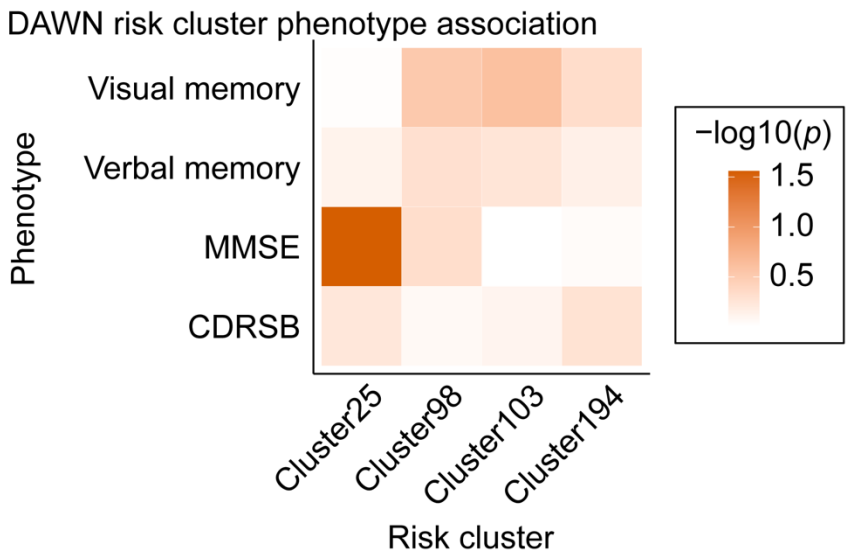


Supplementary Fig. 9 Expression of prioritized genes. Expression of *ARHGAP18* (**a**), *EPB41L2* (**b**), *TXNDC2* (**c**), and *PTPRD* (**d**) across brain sub-cell types and expression patterns across pseudo-progression from Seattle Alzheimer's Disease Brain Cell Atlas. Upper panel: gene expression across brain sub-cell types. Lower panel: expression change pattern with pseudo-progression.

OPC, oligodendrocyte precursor cell

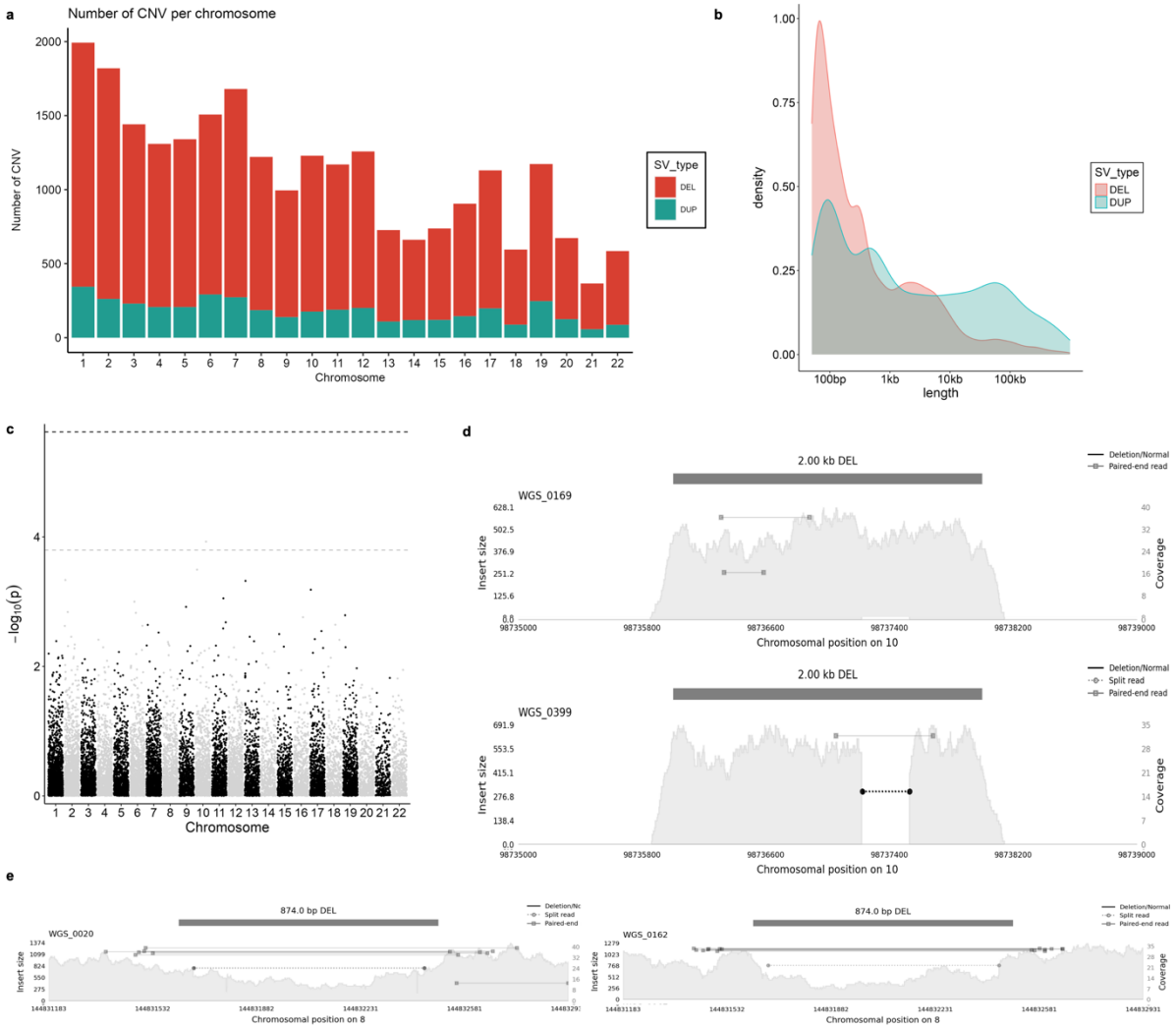


Supplementary Fig. 10 Power estimation from CWAS. **a**, Power estimation from Aβ-positivity test (Aβ-positive (n=925); Aβ-negative (n=634)) in CWAS. Estimated power = 0.85. **b**, Power estimation from diagnosis test (DAT (n=314); CU (n=655)) in CWAS. Estimated power = 0.098. n refers to the number of individual samples. CWAS, category-wide association study. Source data are provided as a Source Data file.

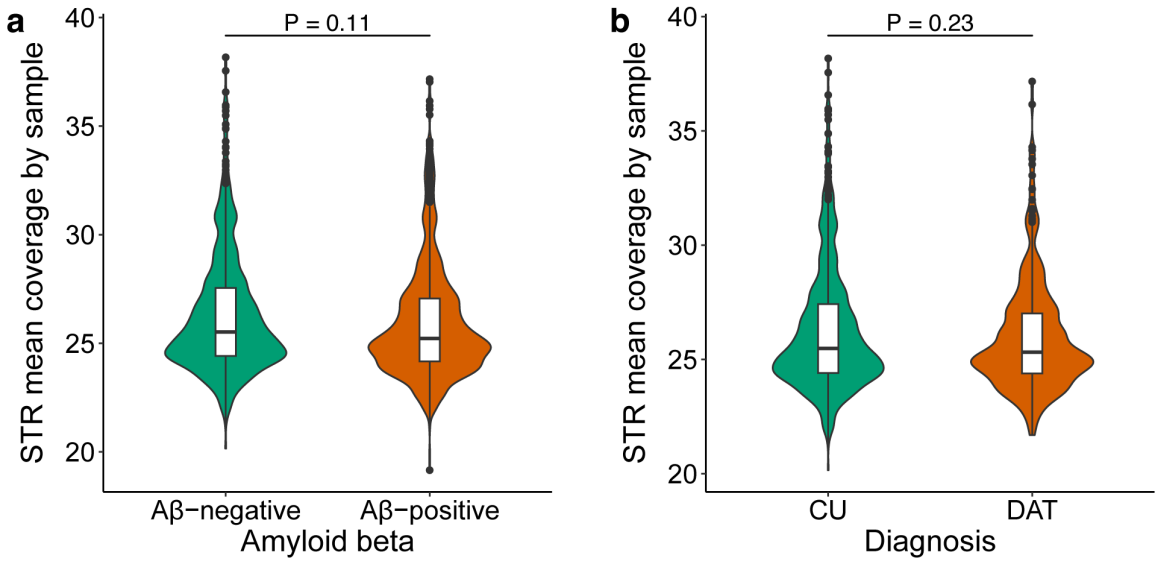


Supplementary Fig. 11 Association analysis of AD clinical phenotypes across risk clusters. Evaluation of cognitive function severity across four DAWN-identified risk clusters based on four AD clinical phenotypes (visual memory, verbal memory, K-MMSE, CDRSB) in Aβ positive samples. Two-sided linear regression was used with adjustment for sample covariates. The exact P values are in the source data.

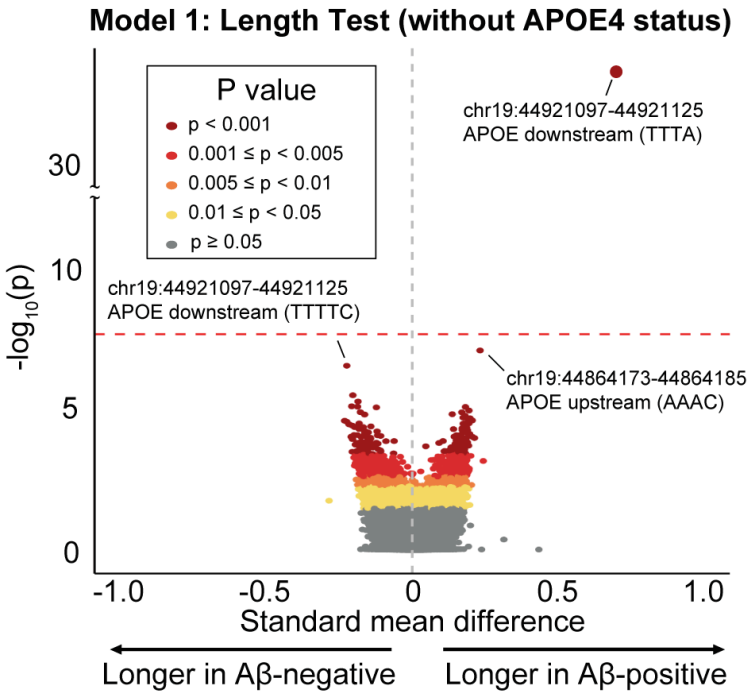
K-MMSE, Korean Mini-Mental State Examination; CDRSB, Clinical Dementia Rating Sum of Boxes; FDR, False Discovery Rate. Source data are provided as a Source Data file.



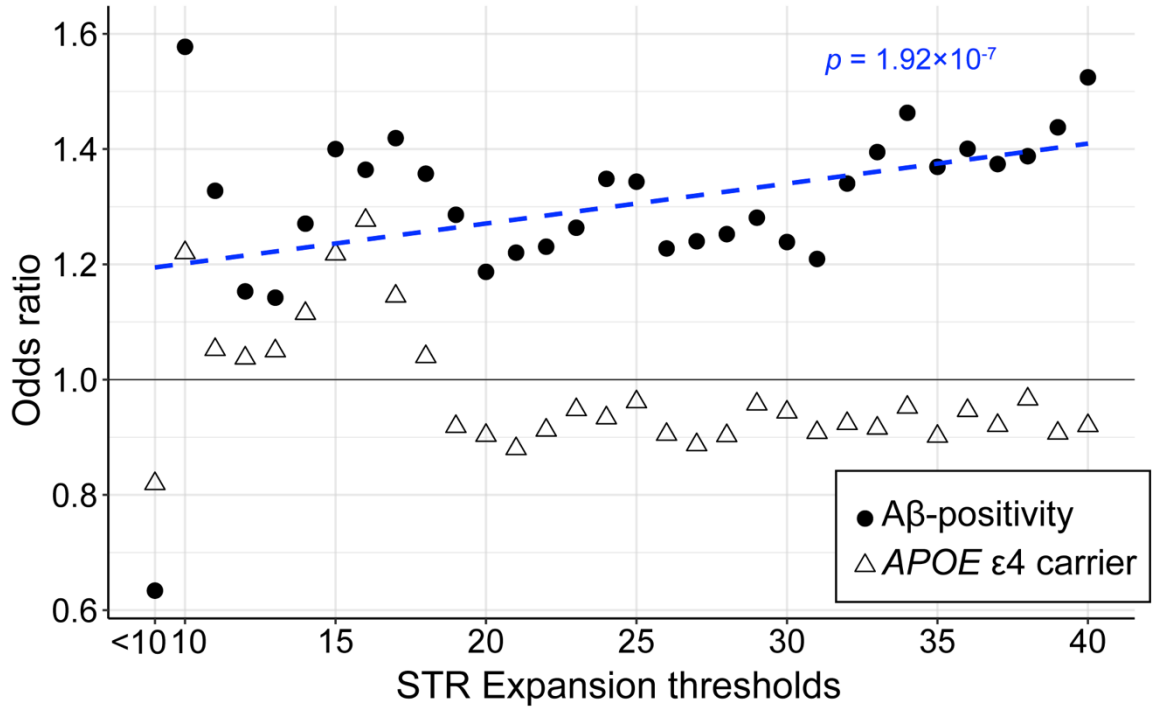
Supplementary Fig. 12 Characteristics of copy number variants. a, The types and quantities of copy number variants (CNVs) for each chromosome. **b**, The length distribution of deletion, and duplication. **c**, Manhattan plot for copy number variation association analysis of clinical status of the Korean cohort. Black and gray dashed lines indicate the Bonferroni ($p = 2.4\times10^{-6}$) and cytoband-based Bonferroni ($p = 1.6\times10^{-4}$) thresholds, respectively. Two-sided logistic regression was used with sample covariates. **d**, Differences between *HPSE2* deletion carriers and non-carriers confirmed through the samplot. **e**, Ambiguity between *ZNF7* false positive deletion carriers and positive deletion carriers confirmed through the samplot. Source data are provided as a Source Data file.



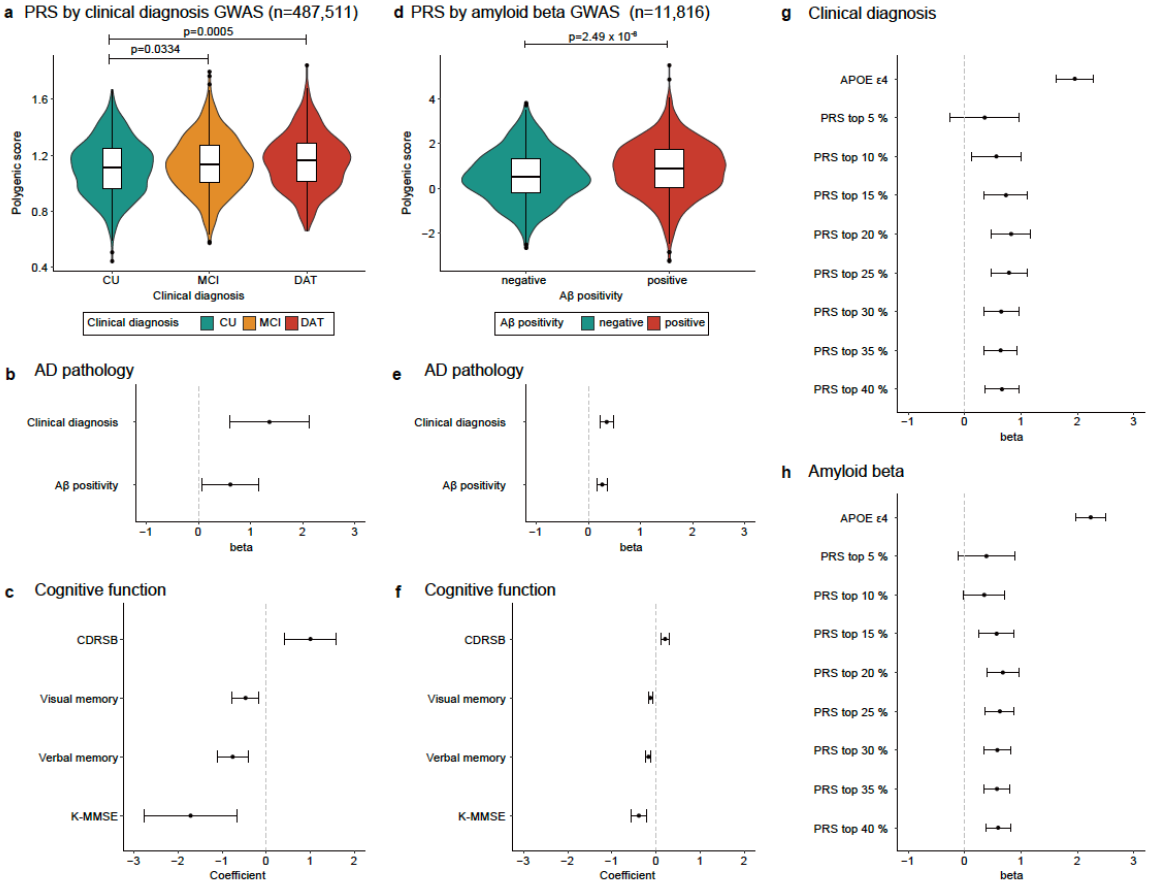
Supplementary Fig. 13 Comparison of sample mean STR coverage between groups. a, Sample mean coverage by Aβ-positivity (Aβ- negative (n = 634); Aβ-positive (n = 925)). **b**, Sample mean coverage by diagnosis (CU (n = 655); DAT (n = 314)). Two-sided linear regression was used with adjustment for technical covariates. n refers to the number of individual samples. Aβ, amyloid beta; CU, cognitively unimpaired; DAT, dementia of Alzheimer's type; STR, short tandem repeat. Source data are provided as a Source Data file.



Supplementary Fig. 14 The length of individual STRs was evaluated using a logistic regression model. The volcano plot shows the association between individual STRs (x-axis: difference of the mean length between Aβ-positive and Aβ-negative samples divided by the SD of each STR; y-axis: p-value from logistic regression). For this analysis, the *APOE* ε4 carrier status was not included as a covariate because the goal was to test for replicating findings from the European study. The three most significant STRs were located near the *APOE* gene, and the STR that passed Bonferroni correction showed LD with the *APOE* allele (r^2 = 0.56). STR lengths were assessed using two-sided logistic regression with sample and technical covariates. Source data are provided as a Source Data file.



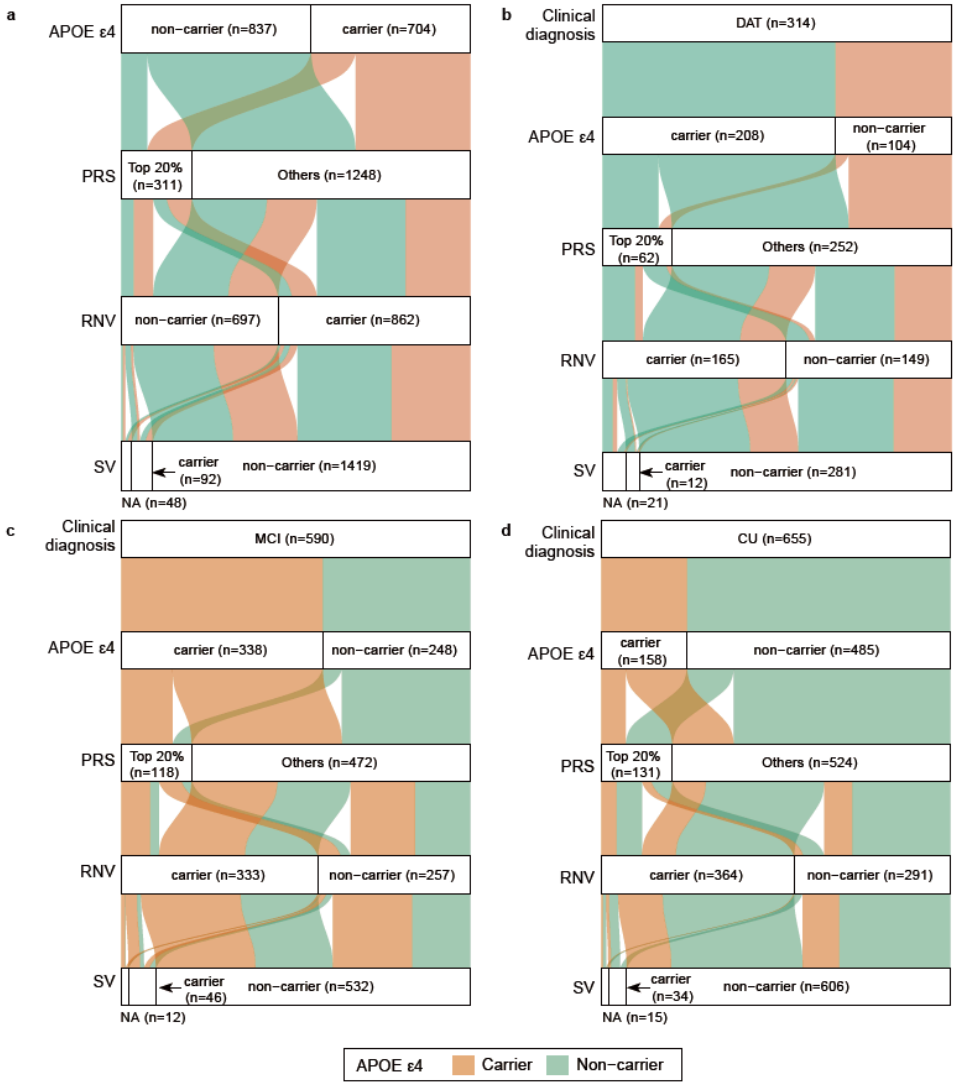
Supplementary Fig. 15 Correlation between STR outlier counts and AD odds ratio adjusted for *APOE* ε4 odds ratio. Linear regression model showing the correlation between the number of STR outliers and the odds ratio for Alzheimer's disease (AD), adjusted for the *APOE* ε4 odds ratio. The odds ratio (black circles) increases with the number of STR outliers, independent of *APOE* ε4 odds ratio (white triangles). The dashed blue line represents the regression fit (p = 9.21×10^{-7}). Two-sided linear regression was performed, adjusting for threshold and APOE4 odds ratio. Source data are provided as a Source Data file.



Supplementary Fig. 16 Comparison of performance between clinical diagnosis and Aβ GWAS-derived PRS. a-c, Comparison of PRS calculated from clinical diagnosis GWAS (n = 487,511) with clinical diagnosis (**a**), AD pathology (**b**), and cognitive function (**c**). **d-f**, Comparison of PRS calculated from Aβ GWAS (n = 11,816) with Aβ (**d**), AD pathology (**e**), and cognitive function (**f**). **g-h**, Comparison of PRS calculated using GWAS with Aβ (n = 11,816) with clinical diagnosis (**g**) and Aβ (**h**). Samples were stratified into percentiles ranging from 5% to 40%, and the significance between the high PRS group and other samples was compared. Significance was calculated using linear regression for AD pathology and logistic regression for cognitive function. Error bars represent the 95% confidence interval. n indicates the number of independent participants. Source data are provided as a Source Data file.

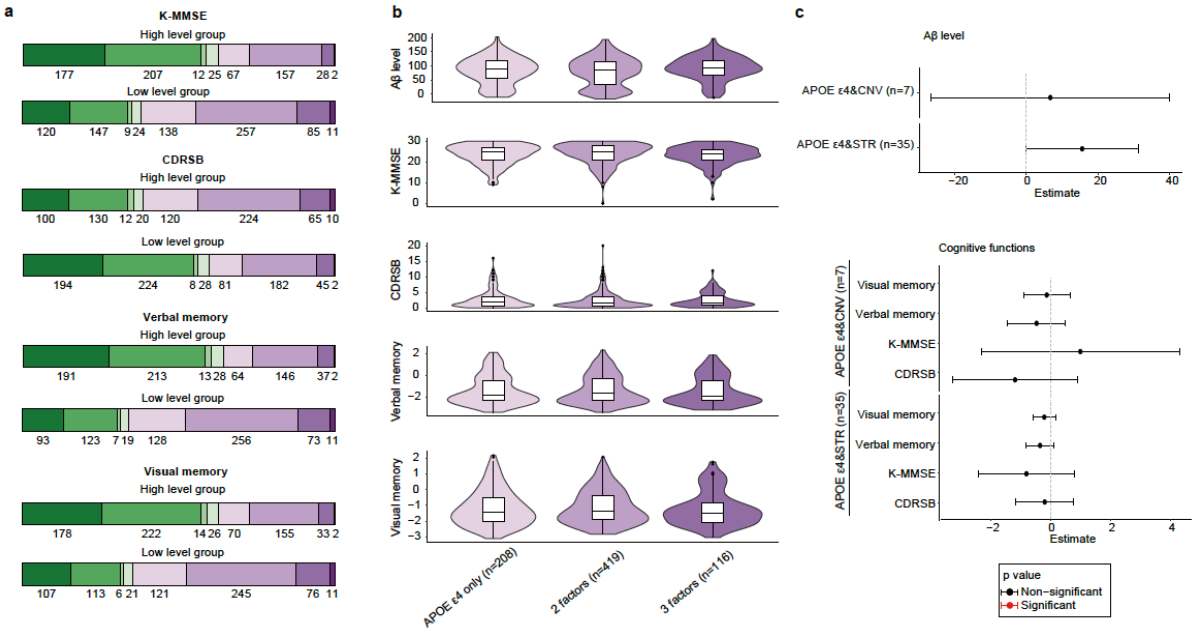
Aβ, amyloid beta; GWAS, genome-wide association study; PRS, polygenic risk score;

DAT, dementia of Alzheimer's type; MCI, mild cognitive impairment; CU, cognitively unimpaired; K-MMSE, Korean Mini-Mental State Examination; CDRSB, Clinical Dementia Rating Sum of Boxes.



Supplementary Fig. 17 Distribution of *APOE* ε4 carriers, individuals in the high PRS group, RNV, and SV carriers. a-d, Alluvial plot of *APOE* ε4, PRS, RNVs, and SVs carrier. Samples are grouped from left to right as follows: all samples (**a**), DAT (**b**), MCI (**c**), and CU (**d**) samples. n indicates the number of independent participants.

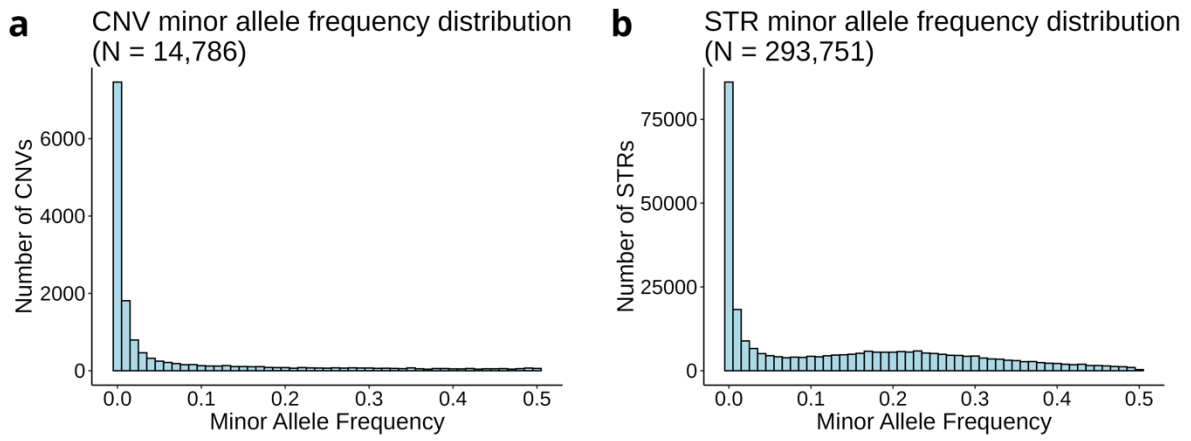
PRS, polygenic risk score; RNV, rare noncoding variants; SV, structural variants; DAT, dementia of Alzheimer's type; MCI, mild cognitive impairment; CU, cognitively unimpaired.



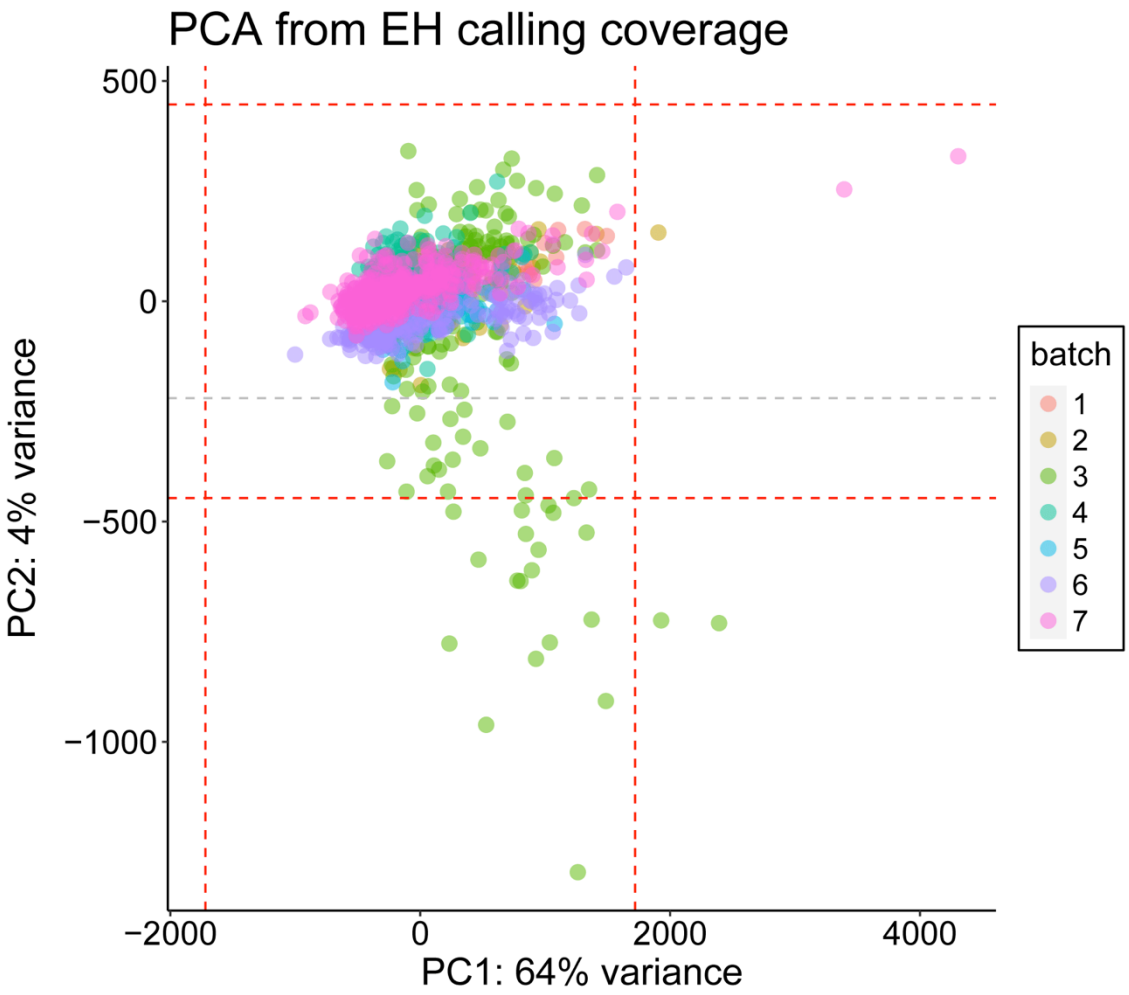
Supplementary Fig. 18 Distribution of genetic effects on cognitive function and comparison of the variant types. a, Stacked bar plot depicting the distribution of samples categorized by *APOE* ε4 carriers, PRS top 20% group, RNV or SV carriers per cognitive function markers. Individuals were divided into high-level and low-level groups based on whether their values were above or below the median, respectively. Cases having two or more genetic factors were categorized into groups according to the number of factors: two factors, three factors, and four factors groups. **b,** Violin plots illustrating the distribution of Aβ level, K-MMSE, CDR-SB, verbal memory, and visual memory scores across each group. Wilcoxon rank-sum tests were employed for statistical significance testing, and FDR correction was used to adjust for multiple comparisons. **c,** Forest plot comparing the difference in Aβ level and cognitive function markers between groups with only *APOE* ε4 (n = 633) and those with CNV or STR carriers along with *APOE* ε4. For statistical significance testing in forest plots, linear regression was employed, adjusting for age, sex, batch, education, and 10 principal components of genetic ancestry. n indicates the number of independent participants.

Source data are provided as a Source Data file.

RNV, rare non-coding variant; CNV, copy number variants; STR, short tandem repeat; K-MMSE, Korean Mini-Mental State Examination; CDR-SB, Clinical Dementia Rating Sum of Boxes.



Supplementary Fig. 19 Minor allele frequency (MAF) distribution of CNVs and STRs. **a**, The histogram illustrated the MAF distribution for CNVs across samples. Alleles observed in only a single sample were excluded to ensure a robust distribution profile. **b**, The histogram showed the MAF distribution for STRs. All detected alleles were included in the analysis to capture the full range of STR. *n* refers to the number of distinct variants of each type included in the analysis. STR, short tandem repeats; CNV, copy number variation. Source data are provided as a Source Data file.



Supplementary Fig. 20 PCA results for the coverage of all samples of called STRs. Samples showing more than four times the standard deviation from the mean in the distributions of PC1 and PC2 were excluded. Additionally, samples outside the main cluster were filtered out using a threshold of less than -220 on PC2. EH, ExpansionHunter; PC, principal component. Source data are provided as a Source Data file.