INDIANA UNIVERSITY
School of Medicine

AAIC 2026 · London · 12 to 15 July

High predicted regulatory impact distinguishes case-enriched rare variants in Alzheimer's disease

Taeho Jo, Li-San Wang, Li Shen, Paul M. Thompson, Yunlong Liu, Kwangsik Nho, Andrew J. Saykin, for the Alzheimer's Disease Sequencing Project

Indiana University School of Medicine, Indianapolis, IN, USA University of Pennsylvania University of Southern California tjo@iu.edu

Background

- GWAS have identified 75+ AD loci, mostly common variants of small effect that explain only a fraction of heritability. Rare variants (MAF < 1%) may carry larger, more functional effects.
- Beyond a few coding variants (TREM2, SORL1, ABCA7), the function of most rare variants is unknown, and association testing is underpowered at low allele frequency.
- Conservation tools such as CADD poorly predict regulatory disruption. AlphaGenome predicts regulatory effects directly from DNA sequence across RNA-seq, CAGE, ChIP-histone, and DNase.

Methods

Cohort

- ADSP WGS R4, 24,595 participants (6,296 AD, 18,299 CN), multi-ancestry (NHW, African American, Asian, Hispanic).

Variants and genes

- Rare variants (MAF < 1%, allele count ≥ 3) within 85 AD-associated genes from the ADSP Gene Verification Committee. APOE excluded (no qualifying rare variants). 9,943 variants annotated by AlphaGenome.

Regulatory prediction

- AlphaGenome scored each variant as the maximum absolute alt-vs-ref difference across modalities. Processed at scale with alphagenome-mcp, a custom Model Context Protocol server.

Analysis

- Per-variant case-control allele frequency ratio defines case-enriched (ratio > 1) vs control-enriched. Interaction ratio (IR) = percent case-enriched among high-effect (top 20%) over low-effect variants. Fisher exact and logistic regression. Independent replication in ADSP R5 (N = 11,545).

24,595
WGS participants (6,296 AD)

9,943
rare variants in 85 AD genes

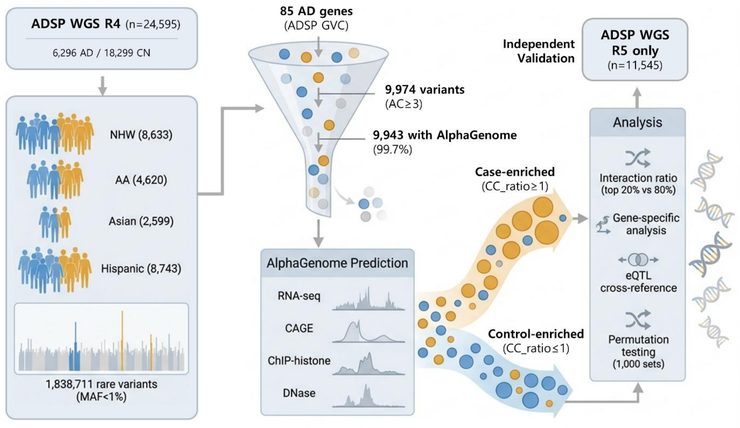

Figure 1. Study design. ADSP WGS R4 rare variants in 85 AD genes are scored by AlphaGenome across functional modalities, then compared between case-enriched and control-enriched variants.

Key results

Higher regulatory impact in AD-enriched variants

- Case-enriched variants scored higher for 3 of 4 modalities: RNA-seq +6.0% (P = 3.3e-20), CAGE +11.0% (P = 8.7e-9), ChIP-histone +5.1% (P = 9.7e-8). DNase showed an inverse pattern.

High-effect variants preferentially case-enriched

- 76.0% of top-20% RNA-seq variants were case-enriched vs 70.0% of the rest. RNA-seq OR = 1.37 (95% CI 1.22-1.53, P = 5.6e-8); CAGE OR = 1.22; ChIP-histone OR = 1.23.

Specific to AD genes

- Across 1.73M non-AD gene variants IR was near 1; effects were 10-30x larger in AD genes. CADD did not discriminate (AUC 0.509) versus AlphaGenome (AUC 0.554).

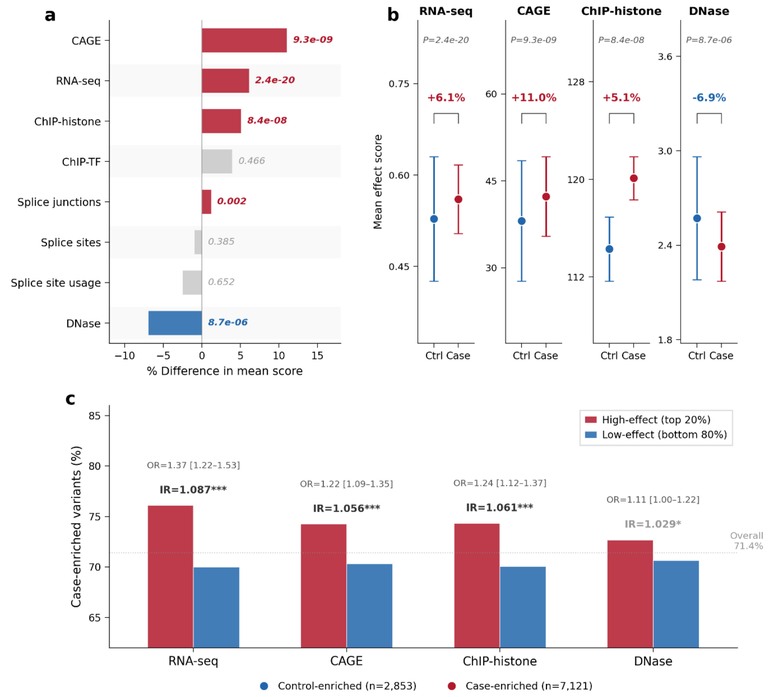

Figure 2. AlphaGenome score comparison. (a) Percent difference by modality. (b) Mean effect scores, case- vs control-enriched. (c) Case-enrichment among high- vs low-effect variants with odds ratios.

Cell type-specific dual architecture

- Genes were grouped by predominant brain cell type, revealing a quantity versus quality trade-off detailed at right.
- Microglia genes (14 genes, 1,489 variants): highest case-control ratio (2.34) but low per-variant effect (RNA-seq 0.30).
- Neuron genes (16 genes, 2,985 variants): lower case-control ratio (1.90) but highest per-variant effect (RNA-seq 0.95).
- Neuron-gene variants showed 2-fold higher occupancy in brain active regulatory elements than microglia-gene variants (6.4% vs 3.3%), a mechanistic basis for the higher per-variant effect.

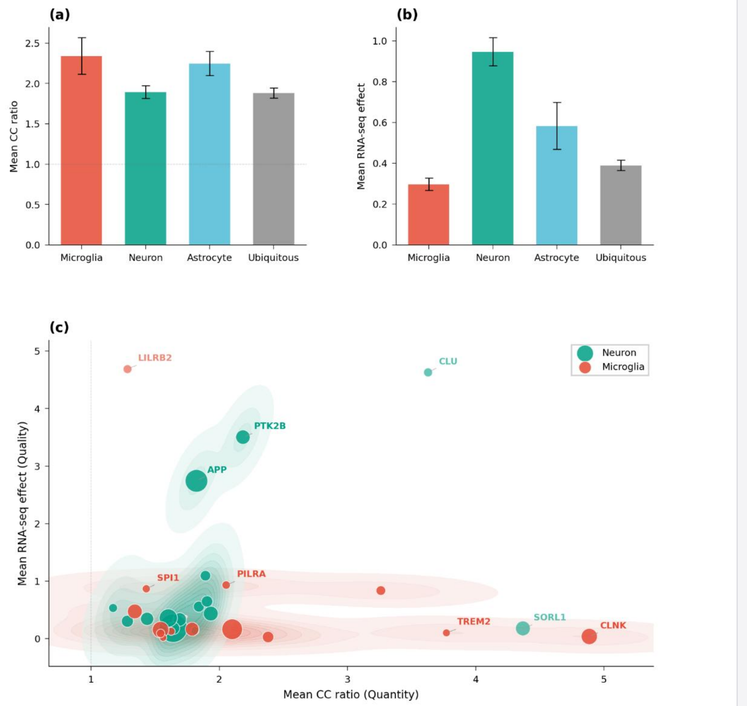

Figure 4. Cell type-specific architecture. (a) Mean case-control ratio by cell type. (b) Mean RNA-seq effect by cell type. (c) Quantity (case-control ratio) versus quality (per-variant effect) trade-off for microglial and neuronal genes.

Independent replication and concordance

- Replicated in an independent ADSP R5 cohort (N = 11,545, no R4 overlap; 84.1% of variants present). After 1:2.9 case/control matching, all four modalities had IR > 1; RNA-seq IR = 1.129 encompassed the R4 value.
- AlphaGenome DNase scores correlated with snATAC-seq accessibility (GSE174367) in microglia (r = 0.121) and astrocytes (r = 0.159), not neurons, recapitulating the glia-driven pattern.

Conclusions

- Rare variants with high predicted regulatory impact are systematically enriched in AD cases, across modalities, independently replicated, and specific to AD genes.
- A dual architecture distinguishes microglial genes (many modest-effect variants) from neuronal genes (fewer high-impact variants).
- Sequence-based regulatory prediction captures disease-relevant signal in non-coding rare variants that CADD misses, providing a framework for prioritization.

Resources and funding

- Data and catalog taehojo.github.io/rarevariants
- Funding Alzheimer's Association AARG-22-974053 and NIH U01 AG068057, R01 AG092591, U19 AG074879, U01 AG024904, P30 AG072976, U01 AG072177. ADSP data via dbGaP phs000572.