Received: 10 May 2024 | Revised: 30 September 2024 | Accepted: 11 October 2024

DOI: 10.1002/alz.14379

Alzheimer's & Dementia®
THE JOURNAL OF THE ALZHEIMER'S ASSOCIATION

RESEARCH ARTICLE

Genetic variants associated with age-related episodic memory decline implicate distinct memory pathologies

Amanat Ali[1] | Sofiya Milman[1,2] | Erica F. Weiss[3] | Tina Gao[1] | Valerio Napolioni[4] | Nir Barzilai[1,2] | Zhengdong D. Zhang[2] | Jhih-Rong Lin[2]

[1]Department of Medicine, Albert Einstein College of Medicine, Bronx, New York, USA

[2]Department of Genetics, Albert Einstein College of Medicine, Bronx, New York, USA

[3]Department of Neurology, Albert Einstein College of Medicine, Bronx, New York, USA

[4]School of Biosciences and Veterinary Medicine, University of Camerino, Camerino, Italy

Correspondence
Sofiya Milman, Zhengdong Zhang, and Jhih-Rong Lin, Department of Genetics, Albert Einstein College of Medicine, Bronx, NY 10461, USA.
Email: sofiya.milman@einsteinmed.edu; zhengdong.zhang@einsteinmed.edu; jhih-rong.lin@einsteinmed.edu

Funding information
Foundation for the National Institutes of Health, Grant/Award Numbers: R01AG061155, U19AG056278, P01AG017242, P01AG047200

Abstract

BACKGROUND: Approximately 40% of people aged ≥ 65 experience memory loss, particularly in episodic memory. Identifying the genetic basis of episodic memory decline is crucial for uncovering its underlying causes.

METHODS: We investigated common and rare genetic variants associated with episodic memory decline in 742 (632 for rare variants) Ashkenazi Jewish individuals (mean age 75) from the LonGenity study. All-atom molecular dynamics simulations were performed to uncover mechanistic insights underlying rare variants associated with episodic memory decline.

RESULTS: In addition to the common polygenic risk of Alzheimer's disease, we identified and replicated rare variant associations in *ITSN1* and *CRHR2*. Structural analyses revealed distinct memory pathologies mediated by interfacial rare coding variants such as impaired receptor activation of corticotropin releasing hormone and dysregulated L-serine synthesis.

DISCUSSION: Our study uncovers novel risk loci for episodic memory decline. The identified underlying mechanisms point toward heterogenous memory pathologies mediated by rare coding variants.

KEYWORDS
Alzheimer's disease, episodic memory decline, protein modeling, rare variants

Highlights

- We demonstrated the contribution of the common polygenic risk of Alzheimer's disease to episodic memory decline.
- We discovered and replicated two risk genes associated with episodic memory decline implicated by rare variants, were discovered and replicated.
- We demonstrated molecular mechanisms and potential novel memory pathologies underlying interfacial rare coding variants.

Sofiya Milman, Zhengdong D. Zhang, and Jhih-Rong Lin contributed equally to this study.

- Molecular dynamics simulations were performed to understand the downstream effects of risk rare coding variants.

1 | BACKGROUND

Cognitive aging is a natural process characterized by functional impairments across different cognitive domains including memory.[1–3] Episodic memory is a fundamental cognitive function that is particularly vulnerable to decline with aging.[4,5] Age-related episodic memory loss can manifest in various ways, including difficulties in learning and retaining new information. While some cognitive changes are considered manifestations of normal aging, progressive cognitive decline, especially in episodic memory, is a key clinical feature of Alzheimer's disease (AD).[6,7] Previous family and twin studies have demonstrated a moderate to high genetic influence on episodic memory in both cognitively healthy and cognitively impaired older adults.[8,9] However, further understanding of the genetic and molecular basis of episodic memory decline in the older population is needed to aid in the identification of novel drug targets to support cognitive reserve and, ultimately, to prevent and treat AD and related dementias.

In general, episodic memory reaches its peak in adulthood and may be subject to decline after the age of 60 years.[5,10] While memory function can be influenced by environmental and other factors that change with age,[11] the rate of memory loss appears to have a genetic component.[12,13] Previous genome-wide association studies (GWAS) of age-related cognitive decline,[14–16] as well as episodic memory performance among older adults,[17,18] have identified AD risk loci with apolipoprotein E (*APOE*) being the major one; thus, prior findings strongly suggested a connection between AD pathology and episodic memory impairment in the general population. However, the discovery of other gene loci and their relationship to episodic memory decline has been relatively limited, despite experimental and clinical evidence that other factors, such as dysregulation of hormone and cerebrovascular supply,[19–22] may contribute to memory decline. Furthermore, the association of rare coding genetic variants with episodic memory decline remains largely unexplored. Uncovering these rare risk variants and understanding their underlying mechanisms may not only bridge existing research gaps but may also be of high scientific and clinical values, given the heterogeneity in pathological episodic memory decline in the population.

Using a longitudinal study design, we tracked and analyzed the decline of episodic memory in 923 Ashkenazi Jewish participants from the LonGenity study, all age ≥ 60 years. We investigated both common and rare coding variants underlying individual differences in episodic memory decline using single nucleotide polymorphism (SNP) array and whole exome sequencing (WES) data, respectively. This study cohort is especially advantageous for identifying associations with causal rare variants at higher frequencies compared to the general population as it originates from a founder population of Ashkenazi Jews (AJs).[23] Furthermore, AJs are relatively homogenous in their education and socioeconomic status, which allows for better control of environmental factors that may impact memory decline.[24] Through genetic association analyses, we identified putative risk genes and pinpointed variants significantly associated with episodic memory decline. Subsequent post-GWAS and protein structural analyses were performed to gain insight into the biological mechanisms underpinning age-related episodic memory loss.

2 | METHODS

2.1 | Recruitment of study participants

LonGenity is a longitudinal study initiated in 2008 with the goal of identifying genotypes and phenotypes that protect against age-associated diseases. The LonGenity cohort is composed of older AJ individuals based on self-reported ethnicity, with the criterion that all four grandparents must be AJ. The ancestry of our cohort was later confirmed with genetic ancestry modeling (Figure S1 in supporting information). Approximately half of LonGenity participants have a history of parental longevity, defined as having at least one parent who lived to age ≥ 95 years. Other prerequisites for eligibility include the absence of dementia at enrollment and age ≥ 60 (vast majority ≥ 65; mean age 75.4; Figure S2 in supporting information). At annual visits, study participants are thoroughly characterized with evaluations that include medical history and neurocognitive assessments. Cognitive assessments were available for 1112 (56.9% female) study individuals, but only 923 participants had at least two cognitive assessments. The Albert Einstein College of Medicine Institutional Review Board approved the LonGenity study.

2.2 | Episodic memory assessment

Episodic memory was assessed at annual visits as part of a comprehensive cognitive battery. Standardized neuropsychological memory tasks (Weschler Memory Scale-R Logical Memory I and II;[25] Free and Cued Selective Reminding Test,[26] Free Recall and Total Recall, Recall of the Repeatable Battery for the Assessment of Neuropsychological Status figure)[27] were administered and scored by the study research assistants under the supervision of the study neuropsychologist. As the tests were on different scales, we calculated *Z* scores standardized to baseline for the individual tests by subtracting the cohort mean score at baseline from the participant's test score at each assessment and dividing by the standard deviation of the baseline scores. A composite memory score was generated by adding four standardized memory scores.

2.3 | Measures of episodic memory decline

Because memory assessments at each study visit, as represented by standardized memory scores, were skewed, we transformed the standardized memory score phenotype using the log-based transformation (Figure S3 in supporting information):

$$\textit{Transformed memory score} = \begin{cases} \log(1 + \textit{Memory score}), & \textit{if Memory score} > 0 \\ 0, & \textit{if Memory score} = 0 \\ -\log(1 - \textit{Memory score}), & \textit{if Memory score} < 0 \end{cases}$$

Nine hundred twenty-three subjects had memory assessments for at least two waves, with each wave defined as a study visit. We measured the episodic memory decline of these 923 subjects by fitting a linear mixed model to the transformed memory scores of subjects at each wave. Baseline age (age at enrollment), sex, education years, and comorbidity score were included as predictors with fixed effects, while the number of years of follow-up after enrollment at each cognitive assessment was included as a predictor with mixed effects. The comorbidity score was defined as the sum of binary variables for the history of diagnoses of diabetes, high blood pressure, and stroke at the time of each cognitive assessment. The random effects of years of follow-up on the transformed memory score for each subject were collected as the measurement of the corresponding episodic memory decline (i.e., residual memory slope).

RESEARCH IN CONTEXT

1. **Systematic review**: The literature has been reviewed using conventional sources (e.g., PubMed and Google Scholar). There are several genome-wide association studies relevant to age-related episodic memory impairment. No studies investigating both common and rare variant genetic risk for age-related episodic memory impairment were found.
2. **Interpretation**: Identified rare variants in risk genes, along with elucidated molecular mechanisms, and the presence of common polygenic risk of Alzheimer's disease (AD), supports the notion that memory pathologies encompass heterogenous genetic causes. These causes are linked to prominent AD and other risk loci, in line with previous experimental and clinical evidence.
3. **Future directions**: This research lays a foundation for future studies aimed at comprehending the association of likely functional rare coding variants with episodic memory decline. Our findings shed light on interventions targeting episodic memory decline, warranting a potential strategy to improve resilience against AD. Further longitudinal studies using diverse populations can deepen our understanding of identified novel targets.

2.4 | SNP array genotyping

Among 923 study participants with memory measurements for at least two waves, only 791 had SNP genotype data. SNP genotyping was performed on two different genetic platforms. Regeneron and Center for Inherited Disease Research (CIDR) carried out genotyping using Illumina's Infinium Global Screening Array-24 v.1.0 BeadChip with 642,824 markers and Illumina's HumanOmniExpress-12 v.1 Array with 730,525 markers, respectively. A total of 629 study participants were genotyped on the Infinium Global Screening Array-24 v.1.0 BeadChip and 162 individuals were genotyped on the HumanOmniExpress-12 v.1 Array. Each SNP array genotyped dataset was "lifted over" to the human genome assembly GRCh38 separately. PLINK software (v.1.9) was used to perform quality control (QC) on each array-based genotyped data set separately. SNP array data was used for our common variant association analysis; SNPs with minor allele frequency (MAF) < 1% were thus removed. SNP and sample missing rate was checked in two steps using > 20% and > 2% genotype calls threshold. SNPs and samples that passed these thresholds were retained. Study participants whose self-reported sex differed from that predicted by sex chromosome heterozygosity were excluded. SNPs were excluded if their genotype frequencies deviated from Hardy–Weinberg equilibrium by a χ^2-test P < 1E-06. After that, samples were excluded if their heterozygosity varied from the mean by more than three standard deviations. A total of 150 CIDR-genotyped and 602 Regeneron-genotyped samples passed all QC steps.

2.5 | Exome sequencing and genotyping

The WES on samples from the LonGenity cohort was conducted by the Regeneron Genetics Center (RGC) using the OQFE protocol.[28] After removing subjects with low sequencing coverage (< 80% bases with coverage ≥ 20x), discordant sex, genotype call rate < 0.9, WES data were generated for subjects in the LonGenity cohort with at least two memory measurements. Genotypes called on human genome assembly GRCh38. We performed additional sample QC to remove subjects with a first-degree relationship (based on the identify-by-descent analysis) with another subject in the final study cohort that consisted of 632 subjects. For variant QC, the genotype of a variant was not called if the read depth (DP) < 5 or genotype quality (GQ) < 10. We also removed variants from our analyses if their genotype frequencies deviated from Hardy–Weinberg equilibrium (χ2-test P < 1E-15). Only variants with missing rates < 1% in the study cohort were analyzed. All samples in the study cohort have genotype call rate > 99% for the analyzed variants.

2.6 | Genotype imputation

Infinium Global Screening Array-24 v.1.0 (n = 602) and HumanOmniExpress-12 v.1 Array (n = 150) genotyped datasets were independently imputed for genotype using the Michigan Imputation Server (Minimac4). For imputation, the Haplotype Reference Consortium (HRC, r1.1 2016) was used as the reference panel, the European (EUR) population for QC, and the Eagle v.2.3 for phasing. After imputation, variants that passed imputation reliability (R2 > 0.3) were retained for further analysis. Overall, 13,189K and 11,656K variants of genotyped datasets from the Infinium Global Screening Array-24 v.1.0 and HumanOmniExpress-12 v.1 Array, respectively, passed the imputation reliability (R2 > 0.3). A total of 11,084K variants were found to be common between the two datasets and were merged. A standard GWAS QC mentioned in section 2.4 was again performed on the merged dataset. A total of 742 samples and 8027K variants passed all QC steps and were used for downstream analysis. The imputation accuracy is high: in 66 subjects of which whole genome sequencing (WGS) data are available, ≈ 98% of 1100 randomly selected variants were accurately imputed (Figure S4 in supporting information).

2.7 | Common variant association analysis

A test of association between residual memory slope and autosomal variants was performed using linear regression as implemented in PLINK v1.90. For this, only SNPs that passed MAF ≥ 1% were analyzed. We also filtered out duplicated variants and individuals with first-degree kinship. Baseline age, sex, years of education, the number of comorbidities, and the top 10 principal components from a principal component analysis (PCA) were included as covariates in the model. The number of comorbidities was defined as the sum of diagnoses history (0 to 3; 1 or 0 for each) for diabetes, hypertension, and stroke at or before the memory assessment. Risk loci were defined around genome-wide significant variants ($< 5 \times 10^{-8}$); a region of ± 250 kb was set around each variant. The PLINK clumping procedure was used to identify independent hits in each region. All variants were subjected to an iterative clumping procedure, beginning with the variant that has the lowest P value (known as the index variant). The index variant's clump included variants with a P value $< 5 \times 10^{-8}$, located within 250 kb of the index variant, and in linkage disequilibrium (LD) with the index variant (r2 > 0.1). After that, the clumping procedure was used to ensure that all the variants were clumped.

2.8 | Rare variant association analysis

In our rare variant analyses, we only analyzed autosomal rare non-synonymous variants. Rare variants are defined as variants with alternative allele frequency (AAF) < 1% in our study cohort and genome Aggregation Database Ashkenazi Jewish. Coding variants were identified based on Combined Annotation Dependent Depletion (CADD) annotation[29] (v.1.6, "CodingTranscript"), in which non-synonymous variants are those not annotated as "synonymous" variants based on Ensembl Variant Effect Predictor (VEP). For single rare variant association analyses, we applied linear regression to test the association between individual variant genotypes and residual memory slope. We tested association with every single rare non-synonymous coding variant and applied multiple test correction (false discovery rate [FDR] < 0.05). However, we further investigated only those that occurred at least five times in the study cohort and CADD scores ≥ 20. The reason for applying this additional selection is to ensure the false positive rate of our reported significant finding is < 5% based on our permutation test to address the issue of skewed phenotype (Supplementary Note and Figure S5 in supporting information). For gene-level rare variant association analyses, we first collected putative functional rare non-synonymous variants on each gene. Putative functional rare variants for a gene were defined as rare variants with VEP impact annotation either MODERATE or HIGH and with a CADD score greater or equal to the corresponding gene-specific cutoff. The gene-specific CADD cutoffs were derived based on a combined approach of mutation significance cutoff[30] and suggested range of CADD cutoff by the authors (Supplementary Note). We ran SKAT-O[31] to test association of putative functional rare non-synonymous variants (CADD scores ≥ gene-specific cutoff) on all coding genes. However, we further investigated only those that met the following criteria: (1) genes that harbor at least two putative functional rare non-synonymous variants to avoid redundant tests from single rare variant association analyses; and (2) genes with the summation of at least five occurrences of putative functional rare non-synonymous variants observed in the study cohort from which statistical tests can provide relatively reliable estimates (see Supplementary Note). For the pathway-level rare variant association analysis, we collected 7608 gene sets derived from the Gene Ontology (GO) biological process ontology from molecular signature database (MSigDB).[32] For each gene set, we ran SKAT-O to test the association of putative functional rare non-synonymous variants (as defined for gene-level association analysis) in all the annotated genes. All single-variant, gene-level, and pathway-level association tests included the following covariates: baseline age, sex, years of education, the number of comorbidities, and top 10 principal components that account for population structure. The principal components were obtained using PLINK (v.1.9) based on common variants in the WES data. For both single and gene-level rare variant association tests, FDR of 0.05 and 0.01, respectively, was used to account for multiple test correction. For the pathway-level association test, we applied a more stringent Bonferroni correction.

2.9 | The Alzheimer's Disease Sequencing Project study

We used WES data from the Alzheimer's Disease Sequencing Project (ADSP) study to examine whether carriers of rare coding variants had a higher chance of developing AD. The data consisted of 5492 cases and 4484 controls of non-Hispanic White ancestry based on self-reported ethnicity (Figure S6 in supporting information). We used Firth logistic

regression to access the association between a rare variant and AD status; sex and top 10 principal components derived from common variants in the WES data were included in the regression as covariates.

2.10 | Replication study cohort

We prepared our replication study using people from the Religious Orders Study/Memory and Aging Project (ROSMAP)[33] with both longitudinal episodic memory measurements ($n = 3771$) and WGS data ($n = 1196$)[34] available. The episodic memory measurements at different visits for an individual were obtained from the Research Resource Sharing Hub (Rush Alzheimer's Disease Center) (https://www.radc.rush.edu/home.htm), and the WGS data was obtained from the AD Knowledge Portal (http://adknowledgeportal.org). We only included subjects without a diagnosis of AD and aged ≥ 65 at the first visit in the replication study to mimic the characteristics of our LonGenity study cohort (Figure S2). Furthermore, to achieve high-quality measurements of episodic memory decline and thus reliable replication results, we limited subjects to those having at least four annual episodic memory measurements, with an acceptable compromise of the sample size. Therefore, we recalculated residual memory slopes for subjects in our replication study—using the same method as in our discovery cohort—instead of using the existing measurements of episodic memory decline that have been generated from previous studies.[14] For the WGS data, only variants passing all filtering criteria were considered. The top 10 principal components accounting for the subpopulation structure were calculated based on common variants in chromosome 22.

2.11 | Protein modeling

The top rare coding missense variants associated with episodic memory decline at the variant and gene levels were further evaluated for their impact on protein structure and function. A set of protein engineering approaches was used to obtain molecular insights into how rare coding missense variants affect protein structure and function. First, homology modeling was performed to determine the three-dimensional (3D) structure of proteins (wild types and mutants) whose human 3D structure was not available in the Protein Data Bank (PDB). Models were produced using Schrödinger Prime 2022-2, or AlphaFold,[35] where available. Loops refinement and structure validation of all 3D protein structures were carried out in Schrödinger Prime 2022-2. Schrödinger's protein preparation wizard was used to exclude unnecessary water molecules and it also relaxed multimeric complexes, allocated bond orders accurately, produced disulfide bonds, corrected the orientation of misoriented groups, and adjusted ionization states. Standard protonation states at pH 7 were applied, and hydrogen atoms were added to the protein structures. Then, to produce geometrically stable structures, the preprocessed structures were optimized and minimized.

Second, the residue scanning tool in Schrödinger Prime 2022-2 was used in an implicit solvent model to determine the impact of mutations on protein stability. Third, the functional nature of variants was ascertained, particularly those that were directly involved in the disruption or formation of new critical intermolecular interactions identified via protein engineering approaches. Fourth, to confirm if a residue was present at an interface essential for protein–protein interaction (PPI), it was checked in computationally predicted interface residues of experimentally determined binary PPIs from Interactome INSIDER (v.2018.3, http://interactomeinsider.yulab.org). If crystal structures of proteins with known interacting partners were available in PDB, either as homodimers or heterodimers, they were used for further analysis. For others, protein–protein docking was performed using Schrödinger Protein-Protein Docking Suite 2022-2 to obtain the structure of a complex. For PPI crystal structures and generated docked complexes, any residue that was at the surface of a protein and whose distance to the interacting partner interfacial residue(s) was < 5 Å was regarded as being at the interface.

To gain mechanistic insights into interfacial variants at an atomic level, all-atom molecular dynamics (MD) simulations were performed using Desmond.[36] For this, the 3D coordinates of corticotropin releasing hormone receptor 2 (CRHR2) bound to G proteins and PSPH were retrieved from PDB (PDB ID: 6PB1 and 1L8L), and structures were prepared using Schrödinger's protein preparation wizard as described above. To obtain the phosphoserine bound complex of PSPH, phosphoserine was first docked to the well-defined active site of phosphoserine phosphatase (PSPH) using Schrödinger's Glide Standard Precision (SP) method with default parameters. The best-docked pose of phosphoserine with the lowest Glide Score (GScore) value was used for MD simulations and further analysis. Complex structure of PSPH was placed in an orthorhombic box of size 100 Å × 100 Å × 120 Å. Because CRHR2 is a G protein–coupled receptor (GPCR), therefore, the bound structure of CRHR2 was embedded into a pre-equilibrated DPPC membrane in an orthorhombic box, with a buffer distance of 10 Å. Both systems were solvated with single point charge (SPC) water molecules using the Desmond System Builder (Schrödinger, LLC). The prepared simulation systems were neutralized with counter ions, and a 0.15 M salt concentration of NaCl was maintained during the simulation runs. All simulations were performed using Desmond. All calculations were carried out using an Optimized Potentials for Liquid Simulations force field. Each prepared system was subjected to the default eight-stage relaxation protocol of Desmond. After relaxation, production simulation runs were performed for 500 ns in triplicates using different seed velocities. The pressure at 1 atm and temperature at 300 K were maintained using an isotropic Martyna–Tobias–Klein barostat and the Nose–Hoover thermostat,[37,38] respectively, during the simulations. The smooth particle mesh Ewald (PME)[39] method was used for long-range columbic interactions whereas short-range cutoff was set as 9.0 Å. The molecular mechanics with generalized Born and surface area solvation (MM/GBSA) method was used to estimate the free energy of binding of the G protein subunit to CRHR2 and phosphoserine ligand to PSPH using frames extracted from MD simulation trajectories. Every 100 ns, frames from each simulated system were extracted, and MM/GBSA–based binding free energy was calculated using Schrödinger Prime, employing the VSGB 2.0 solvation model.[40] Simulation data were analyzed using packaged and in-house scripts.

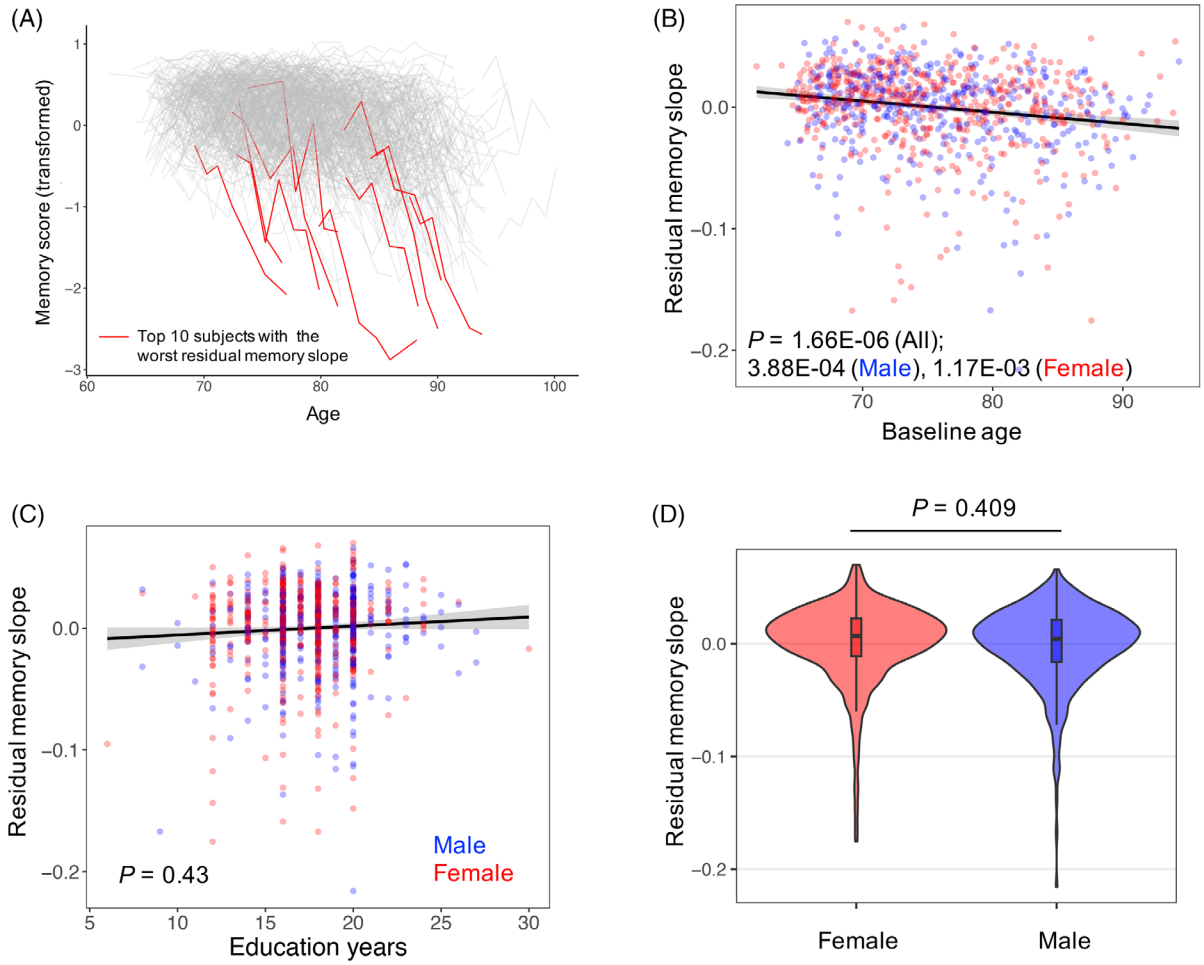

FIGURE 1 Distribution of memory measures. A, Memory trajectories of 923 subjects and 10 subjects with the worst residual memory slope (red) derived from the linear mixed model. B–D, The correlation between the residual memory slope and baseline age (B), education years (C), and sex (D).

3 | RESULTS

3.1 | Memory performance in the study cohort

The study cohort of episodic memory decline included 923 individuals, each having a minimum of two memory assessments. On average, each participant completed 5.5 cognitive assessments over 6.2 years (Figure S7 in supporting information). Linear mixed regression model analysis of 5034 memory assessments revealed a negative correlation between years of follow-up and memory scores (p = 3.41E-08), indicating that memory scores at later follow-up times tended to be lower. Older baseline age (age at enrollment) was also associated with lower memory scores (p < 2E-16). Females, on average, had higher memory scores than males (p = 5.86E-06). As expected, education years positively correlated with memory scores (p = 1.22E-10). There was no significant association between memory scores and the number of comorbidities in our cohort (p = 0.126). The measurement of episodic memory decline (residual memory slope) for each subject was derived as the random effect of years of follow-up in a linear mixed-effect regression model (see Methods section, Figure 1A, and Figure S8 in supporting information). The residual memory slope exhibited a significant negative correlation with baseline age (p = 1.66E-06; Figure 1B), suggesting that individuals who enrolled at a later age were more likely to experience episodic memory decline in subsequent years. Conversely, we did not observe a significant correlation between the residual memory slope and sex (p = 0.409) or education years (p = 0.43; Figure 1C,D). While participants varied in the number of cognitive assessments, the effect of practice/learning that results from exposure to the same tests multiple times was accounted for by adjusting for follow-up time, which was highly correlated with the number of cognitive assessments (correlation coefficient = 0.96; Figure S9 in supporting information). Furthermore, we confirmed that the residual memory slope was not correlated with the number of assessments completed by each participant (p = 0.826), indicating that the learning effect was accounted for in the linear mixed model. Approximately half of our participants were offspring of parents with exceptional longevity (OPEL), and the OPEL exhibited a borderline positive association with the residual memory slope (p = 0.062; Figure S10 in supporting information), suggesting that familial longevity may confer genetic protection against episodic memory decline. The derived residual memory slope from fitting linear

TABLE 1 Summary of GWAS-significant SNPs associated with episodic memory decline.

Rsid	chr:pos:ref:alt (hg38)	Gene	MAF	Beta	*p*	Reference
rs541421523	8:139211220:G:A		0.022	−0.034	1.98e-09	
rs35990795	21:34713536:T:C	*CLIC6*	0.024	−0.032	9.49e-09	Memory[71]
rs548640610	1:239816045:A:T	*CHRM3*	0.011	−0.046	1.54e-08	AD pathology[73]

Abbreviations: AD, Alzheimer's disease; GWAS, genome-wide association studies; MAF, minor allele frequency; SNP, single nucleotide polymorphism.

mixed-effect regression in 923 individuals was used as phenotype of subjects in the common and rare variant association analyses ($n = 742$ and 632, respectively; Figure S11 in supporting information). The slope derived from the entire cohort was very similar to the slope derived from the cohort subsets used for common or rare variant analyses (correlation coefficient > 0.99).

3.2 | Common variant association analysis

The common variant study cohort consisted of 742 subjects from the LonGenity cohort, for whom both SNP array and residual memory slope data were available (Table S1 in supporting information). We conducted GWAS on 8027K common variants (MAF ≥ 1%) that passed QC after imputation (see the Methods section) and confirmed no inflation of test statistics (the genomic inflation factor $\lambda = 1.01$; Figure S12 in supporting information). The association analysis identified three LD-independent genome-wide significant SNPs ($p < 5E-08$): rs548640610, rs35990795, and rs541421523 (Figure 2A and Table 1). These three SNPs were not tag SNPs but had the strongest association among imputed SNPs in their genomic regions. Because none of these SNPs were located at coding regions, we examined the epigenetic information for these positions (Figure 2B–G) using publicly available single cell Chip-seq and ATAC-seq data from different brain regions[41] to ascertain the biological relevance of these putative causal SNPs. rs548640610 resides within the neuron and astrocyte-specific regulatory elements (Figure 2E) and is in an intron of *CHRM3*, which is expressed in the cortex region of the brain. rs548640610 was predicted to occur in a region with unusually strong enrichment for the binding of transcriptional coactivators (super-enhancers). Approximately 19% of AD-associated SNPs have been shown to occur in genome regions encompassed by brain tissue super-enhancers.[42] rs541421523 resides in the neuron-, oligodendrocyte-, and astrocyte-specific regulatory elements (Figure 2F). rs35990795 resides in the microglia-, oligodendrocyte-, and astrocyte-specific regulatory elements (Figure 2G) and is in an intron of *CLIC6*. Interestingly, rs35990795 is expression quantitative trait loci (eQTL) for *CLIC6* in two brain regions based on the Genotype-Tissue Expression database[43] (putamen, $p = 7.0E-08$; cerebellar hemisphere, $p = 4.0E-05$).

3.3 | Analysis of AD common polygenetic risk

Because episodic memory loss is a prominent feature of AD, we hypothesized that some participants experienced episodic memory decline due to AD pathology. To test this hypothesis, we investigated whether the genetic risk for AD contributed to episodic memory decline in our cohort. First, we examined the *APOE* ε4 risk allele—the common risk variant with the strongest effect size in AD. However, we did not observe a significant association with the residual memory slope ($p = 0.09$), even though *APOE* ε4 carriers tended to exhibit a more negative residual memory slope. Next, we assessed in our cohort the common polygenic risk of AD, characterized by AD polygenic risk score (PRS) based on the summary statistics of AD GWAS.[44] Using the default setting of LD-clumping in PRSice-2, we identified a small but significant component of variance for residual memory slope explained by AD PRS ($R^2 = 1.1\%$, adjusted $p = 0.023$; Figure S13a in supporting information). Given that the summary statistics of AD were derived from subjects predominantly of European ancestry, and the subjects in the target cohort of the PRSice-2 analysis were AJs, the standard LD-clumping procedures might fail to select optimal representative SNPs in a genomic region accounting for AD risk due to the discrepancy in LD structures between the two populations. This discrepancy might result in power loss in PRS analyses. Indeed, when using extremely parsimonious clumping parameters to select top SNPs in genomic regions without considering LD structures (clump-kb = 500 kb, clump-$r2 = 0$), the result clearly showed that AD PRS contributed to episodic memory decline in the LonGenity discovery cohort ($R^2 = 2.6\%$, adjusted $p = 2.0E-04$; Figure S13b).

3.4 | Rare variant association analysis

The rare variant study cohort consisted of 632 subjects from the LonGenity cohort, for whom both WES and residual memory slope data were available (Table S1). We identified a total of 97,230 rare coding variants (95,071 SNPs and 2159 INDELs) with AAF < 1% across 16,892 genes in the study cohort. Among these, 35,010 were synonymous, and 62,220 were non-synonymous variants, which included 56,206 missense, 2264 loss-of-function (stop gained and frameshift), and other variants with multiple annotations. We first used linear regression analysis to assess the association of these rare coding variants with residual memory slope and confirmed that there was no systematic inflation (the genomic inflation factor $\lambda = 1.0$; Figure S14 in supporting information). Next, we focused on 7010 rare coding variants that had at least five occurrences in the cohort and had a CADD score ≥ 20, among which the derived significant associations were more reliable (Supplementary Note and Figure S5), and identified 16 putative risk rare coding variants (FDR < 0.05) associated with residual memory slope (Figure 3A and Table S2 in supporting

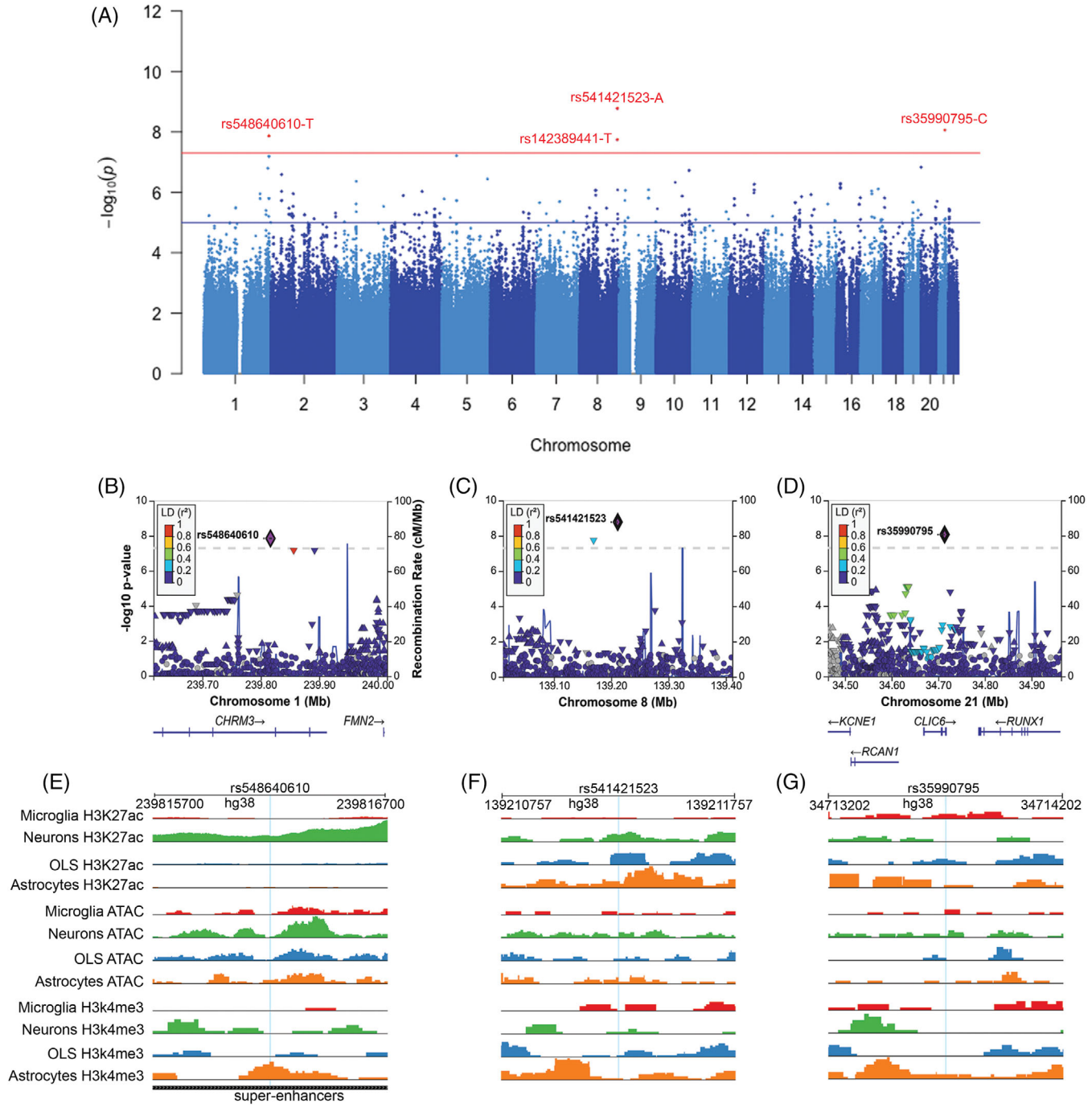

FIGURE 2 Common genetic variants associated with episodic memory decline. A, Manhattan plot demonstrating GWAS results for episodic memory decline. The red and blue lines represent thresholds for genome-wide significance ($p < 5 \times 10^{-8}$) and suggestive significance ($p < 1 \times 10^{-5}$), respectively. B–D, Detailed view of the GWAS SNPs demonstrating associations ± 400 kb from lead SNPs (labeled). The *x* axis represents the variant position on the chromosome and nearby gene positions. The right *y* axis indicates the GWAS *p* value, and the left *y* axis indicates the rate of recombination. Each plot point indicates a SNP in the dataset color-coded by (r2) value using the EUR population from the 1000G LD panel. LD plots were generated using the Locus Zoom plot (locuszoom.org). E–G, Cell type–specific regulatory architecture of the following GWAS-significant SNPs: rs548640610 (E), rs541421523 (F), and rs35990795 (G). A 1000 bp window flanking the SNPs is shown along with the genome tracks illustrating the aggregate accessibility of scChip-seq and scATAC-seq clusters at the locus. A blue line in E–G indicates the location of GWAS-significant SNPs. EUR, European; GWAS, genome-wide association study; LD, linkage disequilibrium; SNP, single nucleotide polymorphism.

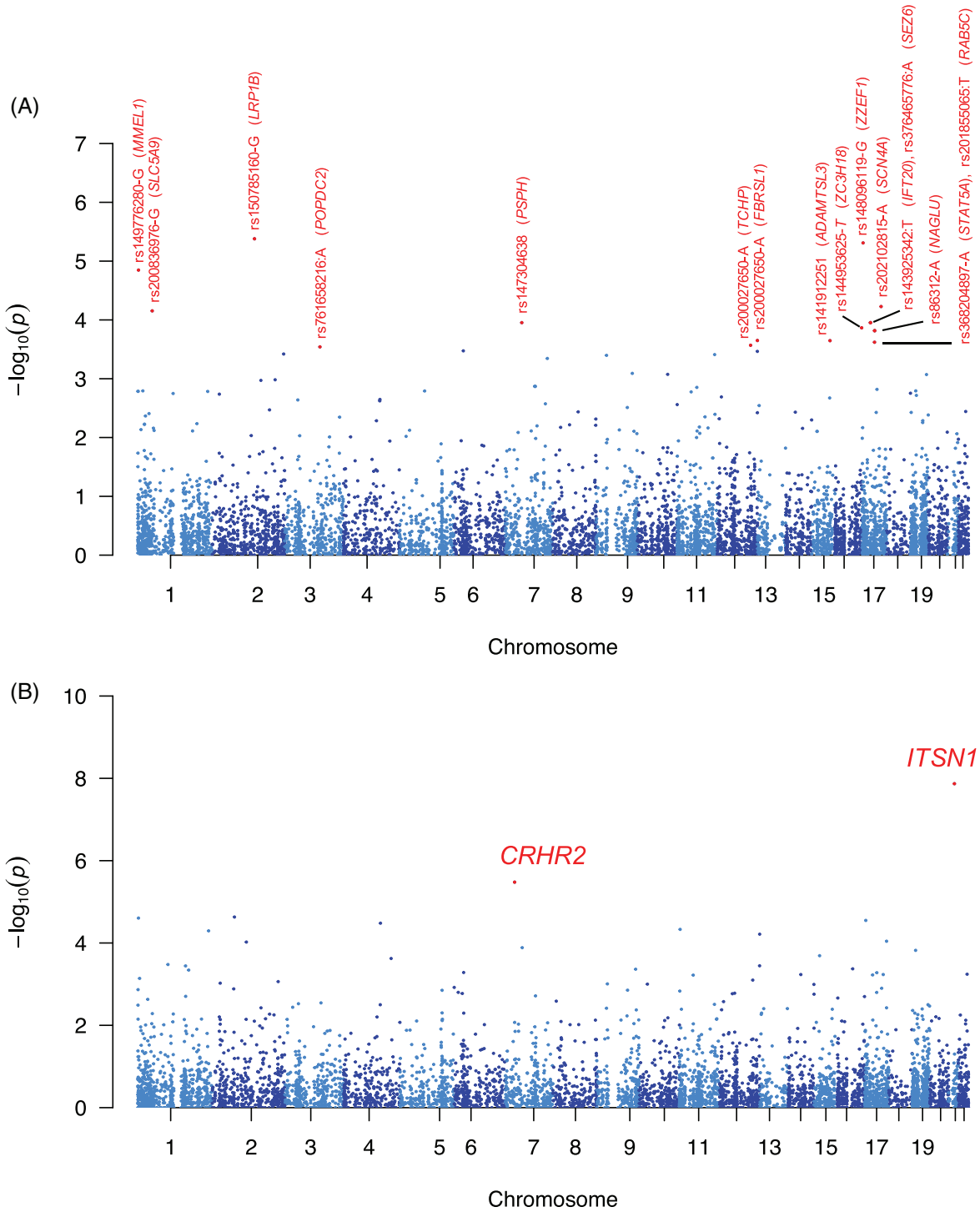

FIGURE 3 Rare variant association analyses. A, The Manhattan plot for 7010 prioritized rare coding variants across the exome. See Table S2 in supporting information for variant annotations. B, The Manhattan plot for the gene-level rare variant association analysis.

information) with an overall false positive rate < 0.05 based on permutation tests (Figure S5). All those putative risk rare coding variants increased the risk of episodic memory decline. For example, five out of six carriers of the top putative risk rare coding variant in the zinc finger ZZ-type and EF-hand domain containing 1 (*ZZEF1*; rs148096119; p = 1.12E-06) exhibited episodic memory decline (Figure S15 in supporting information).

Next, we conducted a gene-based rare variant association analysis to investigate the association between rare variants and residual memory slope at the gene level. In this analysis, we investigated 6282 genes containing multiple rare non-synonymous variants predicted to be functional (CADD score ≥ gene-specific cutoff and VEP impact = MODERATE or HIGH), each with a total occurrence of at least five in our study cohort (see the Methods section). For each of these

genes, we aggregated its rare non-synonymous variants and performed a variant-set association test to assess the association of variant sets with residual memory slope. We opted to use SKATO as the variant-set association test as it combines the strengths of the burden test and SKAT. Our analysis revealed significant rare variant associations with two genes at an FDR < 0.01: intersectin 1 (*ITSN1*; p = 1.35E-08) and *CRHR2* (p = 3.33E-06; Figure 3B and Table S3 in supporting information). We further examined rare coding variants in these two genes using the burden test and confirmed a significant burden of rare coding variants in these two genes that may contribute to episodic memory decline (p = 1.44E-06 and p = 1.21E-05, respectively). At pathway level, we examined rare variant association in 7546 gene sets corresponding to a specific biological process GO term. Pathways "synaptic vesicle recycling via endosome" and "synaptic vesicle endosomal processing" exhibited significant rare variant association (p = 2.3E-08 and p = 3.27E-07, respectively; Table S4 in supporting information).

Last, we performed SKAT_CommonRare to assess the combined effect of common and rare variants on episodic memory decline at the gene and pathway levels. At the gene level, we identified significant (FDR < 0.01) combined association from the same two genes implicated by rare variants alone (*ITSN1* and *CRHR2*). The enhanced association from *CRHR2* (p = 6.8E-07 vs. p = 3.3E-06) suggests the involvement of common risk loci in this gene too. Further testing of all common variants in the transcription region of *CRHR2* using SNP array data confirmed the role of common variants (p = 0.03). At the pathway level, we identified a new significant association with the "response to corticotropin releasing hormone" pathway (p = 5.5E-06).

Overall, there were 18 risk genes implicated by either single or gene-level rare variant associations (Table 2). Gene expression analyses revealed that most of these risk genes were widely expressed in brain tissues (Figure S16 in supporting information). Because AD progression is one possible reason underlying episodic memory decline, we examined AD genes previously implicated by rare coding variants (*APP*, *PSEN1*, *PSEN2*, and *TREM2*)[45] but did not observe significant association signals. However, two genes playing protective roles in AD pathology via regulation of amyloid beta level (*LRP1B* and *MMEL1*)[46,47] were implicated by single rare variant association (rs150785160-G and rs149776280-G, respectively). We examined these two rare coding variants in the ADSP study cohort. Although carriers of rs149776280-G were not found in the ADSP cohort, we confirmed that the carriers of rs150785160-G have a higher risk of developing AD (odds ratio = 5.7; p = 0.035).

TABLE 2 Eighteen risk genes for episodic memory decline implicated by rare variants.

Gene	Type of genetic evidence	Possible mechanisms
LRP1B	Variant-level, Gene-level	AD[46]
ZZEF1	V	Hippocampal plasticity[59]
MMEL1	V	AD[47]
SCN4A	V	–
SLC5A9	V	–
SEZ6	V	Synaptic connectivity[60]
IFT20	V	–
PSPH	V	**Reduced L-serine**
ZC3H18	V	–
NAGLU	V	Hyperphosphorylated tau[61]
STAT5A	V	Brain STAT5A signaling[62]
FBRSL1	V	–
ADAMTSL3	V	–
RAB5C	V	–
TCHP	V	–
POPDC2	V	–
ITSN1	G	Hippocampal plasticity[65]
CRHR2	G	**Corticotropin signaling hormone**

Note: Text in bold denotes novel mechanisms suggested from analysis of interfacial variants.
Abbreviations: AD, Alzheimer's disease; *CRHR2*, corticotropin releasing hormone receptor 2; *ITSN1*, intersectin 1; *PSPH*, phosphoserine phosphatase.

3.5 | Replication study

We used 1680 subjects of EUR ancestry without an AD diagnosis from the ROSMAP cohort who had their first visit at or after 65 years old and a minimum of four annual episodic memory measurements. Using the same approach as in the LonGenity cohort, we calculated the residual memory slope. The result from the linear mixed model analysis indicated a significant negative correlation between years of follow-up and memory scores (p < 2E-16). Older age at the first visit, fewer years of education, and male sex were associated with lower memory scores (p < 2E-16, p = 1.07E-14, and p = 4.41E-14, respectively). The ROSMAP cohort included 526 participants with both the residual memory slope and WGS data available.

For the two genes implicated by gene-based rare variant association, we first examined the same class of functional rare coding variants as defined in our discovery study with MAF < 1% in each gene. SKATO identified significant rare variant association in *ITSN1* (p = 0.035) but not *CRHR2* (p = 0.168). We hypothesized that AJs might have a higher frequency of certain rare causal variants due to the founder effect; certain rare causal variants may occur at lower frequency in the more heterogenous EUR population. Shifting the focus to functional singleton rare coding variants in ROSMAP WGS data covered by the initial criteria, we found a significant negative association between the variant burden and residual memory slope in both *ITSN1* (p = 0.02) and *CRHR2* (p = 0.046). At the pathway level, we did not replicate the finding of rare variant association directly from the two small gene sets (5 genes and 9 genes from "synaptic vesicle recycling via endosome" and "synaptic vesicle endosomal processing," respectively). We hypothesized that the EUR population has a more heterogenetic background than AJ, and the enrichment of risk variants could be observed in a

broader pathway. Examining the gene set "synaptic vesicle recycling," the direct parent term of "synaptic vesicle recycling via endosome," we identified significant rare variant association ($p = 0.04$; $p = 0.028$ in our discovery cohort). At the single variant level, we did not replicate our findings, as most of the putative risk single variants identified in our study cohort either did not appear in the replication cohort or were extremely rare, making our replication cohort likely underpowered to detect their effects. Out of 16 rare coding variants, only two appeared in the replication cohort. Among the three common variants in our cohort, two were rare (MAF < 1%) in the replication cohort. Conversely, by considering AD PRS that characterized aggregate effects of common variants, we replicated and confirmed our finding that AD common polygenic risk significantly contributed to episodic memory decline (adjusted $p = 2.0E-04$), even among individuals without a diagnosis of AD (Figure S17 in supporting information).

3.6 | Impact of rare variants on protein structure and function

To establish additional support for the functional impact of the 18 missense putative risk coding variants implicated by single and gene-based association (Tables S2 and S3), we studied their contribution to protein stability. The effect of each variant was evaluated separately in an implicit solvent model. Results indicated that most missense variants made the protein (mutant protein) unstable compared to the wild type (Figure S18 in supporting information). Among all, ZZEF1:p.Leu1959Pro and CRHR2:p.Arg148Trp produced the highest instability with a Gibbs free energy change value (ΔG) of 36.79 and 31.94 kcal/mol, respectively (Table S5 in supporting information). Several of the variants were noted to increase protein stability.

To ascertain the role of missense variants in perturbing the proteins' overall architecture as well as PPI network, we examined the impact of rare missense variants significantly associated with residual memory slope at amino acid resolution using different protein engineering approaches (Table S6 in supporting information). We constructed a pool of 3D structures of proteins that harbor missense variants using either the experimentally validated crystal structures from PDB or models generated by AlphaFold.[35] Homology modeling was carried out for low-quality structures predicted by AlphaFold using Schrödinger Prime 2022-2.

We first focused on variants that altered the secondary structures of the respective proteins. The majority of disease-causing mutations—≈ 80% of them—have been found to reside in secondary structures.[48] Most of them disturbed the overall configuration of respective proteins, compared to the wild type (Figures S19–S27 in supporting information). For instance, seizure protein 6 (SEZ6) variant p.Arg454Trp was in a conserved CUB 1 domain. The wild type Arg454 was found to make hydrogen bonds with Ser500, Phe527, and Asp579. However, the Trp454 variant was able to form hydrogen bonds with Ser500, Gly502, and Phe527 (Figure S24). Thus, this variant changed the overall configuration of SEZ6. The extra length of sidechains as well as polar functional groups allows the wild type Arg454 to make a stable interaction with Asp579, which was abolished when a non-polar residue Trp454 was introduced. Therefore, the Trp454 variant likely destabilized this region as evidenced by change in protein configuration and confirmed by the stability analysis (Figure S18), possibly contributing to episodic memory decline.

3.7 | PPI-perturbing alleles in memory impairment

Variants located at the protein interfaces produce significant perturbations and have been shown to be associated with disease phenotypes.[49] Interestingly, we identified two rare coding variants that were located at the protein interface where a binding partner (another protein) known to be involved in the pathology of memory-related disorders binds (Table S6). This indicates that variants found in risk genes in our study likely play a role in memory pathogenesis by interacting with another partner via protein–protein crosstalk.

First, we focused on an interfacial variant (Arg148Trp) found in CRHR2. It is a GPCR and has a role in multiple intracellular signaling pathways including adenylate cyclase-cAMP-protein kinase A (PKA), MAPK, and protein kinase C (PKC).[50] A stable binding of GPCR with its G protein is vital to mediate its signaling. To observe the effect of the Arg148Trp variant, we used the high-resolution 3D structure of CRHR2 bound with G proteins (PDB ID: 6PB1), obtained from PDB.[51] Arg148Trp variant is located at the interface where the G_α subunit binds, thus altering the interactions between CRHR2 and G_α (Figure 4 and Figure S28 in supporting information). Arg148 formed a hydrogen bond with G_α:Gln390. Interestingly, p.Arg148Trp disrupted this interaction and further amended surface topography due to the change in size, charge, and polarity between Arg and Trp (Figure 4B,C). To investigate the binding dynamics, stability, and energetics of wild type and mutant CRHR2–G protein complexes, all-atom MD simulations were performed. All simulations of the CRHR2–G protein complex maintained their overall structural integrity and a Cα root mean square deviation (RMSD) from the initial structure was under 5 Å (Figure S29a,b in supporting information). To identify and compare backbone stability and fluctuations of the two complexes, root mean square fluctuation (RMSF) of backbone Cα atoms was analyzed and plotted (Figure S29c,d). Overall, the backbone of both wild type and mutant CRHR2 complexes showed lower fluctuations. Interestingly, residues 27–35 (G_α:36-44), 748–751 (CRHR2:217-220), and 824–833 (CRHR2:293-306) fluctuated more in mutant simulations, compared to wild type ones (Figure S29c,d). CRHR2:Arg148–G_α protein subunit formed a sustained interaction in all simulations. Such an interaction (CRHR2:Trp148-G_α), however, was found to be very weak in the mutant simulations (Figure 4D). Importantly, wild type Arg148 stabilized the conformation of *CRHR2* in such a way that the N-terminal residues Arg38 and His 41 of G_α formed sustained interactions with CRHR2 residues Tyr217 and Glu220 (Figures S28 and S30 in supporting information). However, G_α N-terminal bound less stably with mutant CRHR2. To measure energetic contributions, the free energy of binding (ΔG_{bind}) of the G_α subunit to CRHR2 was estimated from frames of MD simulations every 100 ns using the

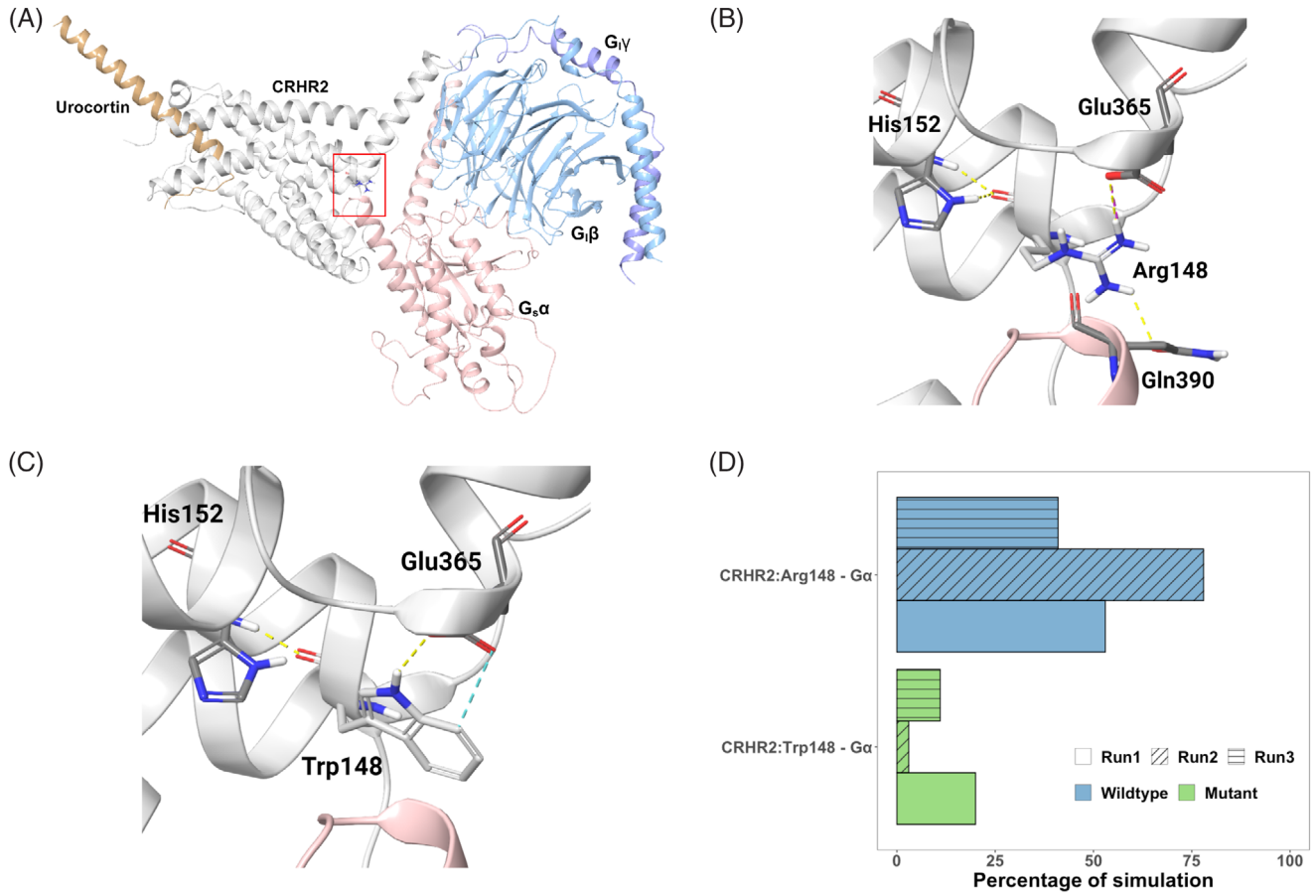

FIGURE 4 3D crystal structure of CRHR2 bound to G proteins. Different subunits are colored differently and labeled. Amino acids are represented with sticks. The red-boxed region in (A) is magnified in the successive images (B and C). A, 3D structure of CRHR2 complex; (B) wild type Arg148; (C) mutant Trp148; (D) the percentage of simulation time during which intermolecular contact was retained between CRHR2:Arg148/Trp148 and Gα protein. 3D, three dimensional; CRHR2, corticotropin releasing hormone receptor 2.

MM/GBSA method. All three wild type CRHR2–G_α bound simulations showed ΔG_{bind} values (−97.3 ± 7.4 kcal/mol, −78.6 ± 8.8 kcal/mol, and −106.2 ± 13.6 kcal/mol) that were substantially higher than mutant CRHR2–G_α bound simulations (−65.29 ± 17.2 kcal/mol, −70.8 ± 21.1 kcal/mol, and −82.9 ± 17.62 kcal/mol), indicative of stronger binding. This indicates Arg148Trp variant will likely alter CRHR2 downstream signaling due to impaired binding of the G_α subunit to the CRHR2 receptor. These results support previous findings in which diminished activity of CRHR2 has been associated with learning, memory, and AD.[52–54]

Next, we focused on another interfacial variant (p.Thr152Ile) observed in PSPH. Human PSPH is a critical enzyme in the phosphorylated pathway of L-serine biosynthesis. It is formed through the dephosphorylation of phosphoserine via PSPH. The impact of this variant on protein structure and function was evaluated by retrieving 3D coordinates of PSPH bound with phosphoserine substrate from PDB (PDB ID: 1L8L).[55] Thr152Ile variant is located near the vicinity of the binding site of PSPH where its natural substrate, phosphoserine, binds (Figure 5A). Thr152 formed two hydrogen bonds with Phe58 and Ser109. However, p.Thr152Ile was only able to form a hydrogen bond with Phe58 (Figure 5B,C). To investigate the influence of this variant on binding mode, interaction stability, and binding affinity of phosphoserine for PSPH, we ran all-atoms MD simulations in triplicates. All simulations remained stabilized throughout the simulation runs and a Cα RMSD from the initial structure was under 5 Å (Figure S31a,b in supporting information). Mutant PSPH simulations showed more fluctuations, especially residues 193–206, compared to wild type runs (Figure S31c,d). Importantly, residues (Asp20, Asp22, Ser23, Glu29, and Arg202) that were part of the PSPH active binding site also fluctuated more in mutant simulations compared to wild type. Interestingly, phosphoserine maintained more sustained interactions with essential binding residues of wild type PSPH, compared to mutant PSPH (Figure 5D and Figure S32 in supporting information). Previous mutagenesis studies focused on active binding site residues of PSPH showed that point mutations of most of these residues significantly diminished enzymatic activity, likely via altered substrate binding.[55] We also computed the binding affinity (ΔG_{bind}) of phosphoserine to PSPH in each simulation run after every 100 ns. Wild type simulations demonstrated ΔG_{bind} values (−31.7 ± 3.8 kcal/mol, −29.6 ± 2.9 kcal/mol, and −36.3 ± 3.6 kcal/mol) that were higher than mutant simulations (−24.7 ± 2.8 kcal/mol, −17.4 ± 10.5 kcal/mol, and −29.1 ± 3.5 kcal/mol) This indicates that Thr152Ile likely

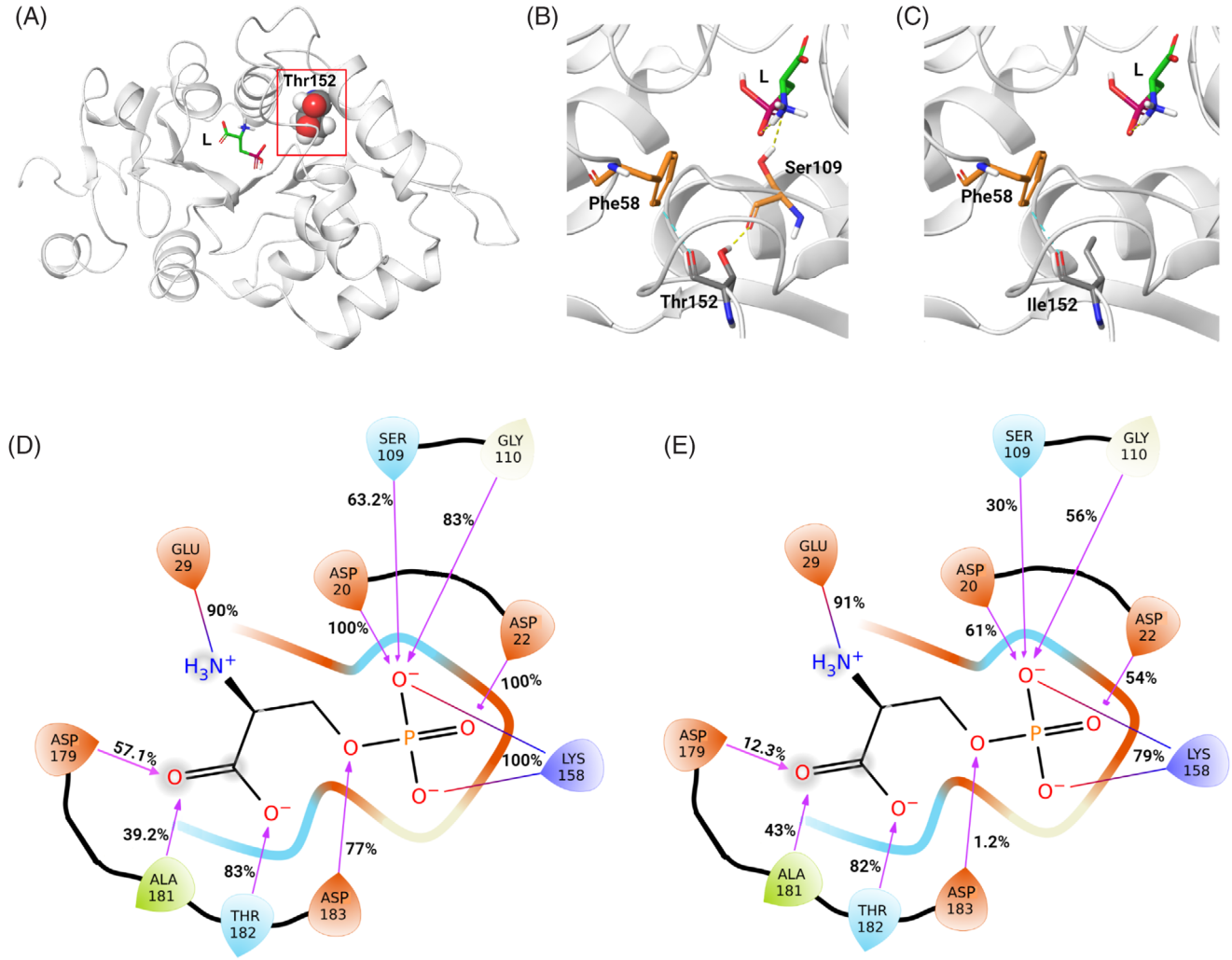

FIGURE 5 3D structure of PSPH-bound phosphoserine. Amino acids are represented with sticks. The red-boxed region in (A) is magnified in the successive images (B and C). (A) Crystal structure of PSPH complex; (B) wild type Thr152; (C) mutant Ile152; (D) average percentage of simulation time a wild type PSPH residue maintains contact with phosphoserine ligand (L); (E) average percentage of simulation time a mutant PSPH residue maintains contact with phosphoserine. Charged and polar amino acids are shown in orange and blue color, respectively. 3D, three dimensional; PSPH, phosphoserine phosphatase.

diminished PSPH enzymatic activity via altering phosphoserine binding to PSPH. Our findings are in agreement with previous results that L-serine protects hippocampal neurons from oxidative stress–mediated mitochondrial damage and apoptotic cell death in a mouse model.[56] Additionally, reduced production of L-serine via PSPH as well as diminished activities of phosphoprotein phosphatase has been associated with cognitive decline in AD.[57,58]

4 | DISCUSSION

In this study, we conducted a comprehensive genetic analysis of episodic memory in older adults from a longitudinal cohort, LonGenity, to elucidate the contributions of both common and rare coding variants. Our investigation revealed robust aggregate effects of both common and rare variants on episodic memory decline in the LonGenity cohort that were subsequently replicated in the ROSMAP cohort. PRS analysis and gene-based rare variant association analysis underscored the significant contributions of common polygenic risk of AD and rare coding variants in *ITSN1* and *CRHR2*, respectively, to episodic memory decline. Single-variant association analysis identified both putative risk common and rare coding variants in genes previously associated with learning and memory. Notably, the putative risk rare coding variants exhibited ≈ 1.5 times larger effect sizes than the identified common variants, suggesting an important role of rare variants in the pathological decline of episodic memory. To establish the functional impact of these identified risk rare coding variants, we applied protein modeling and all-atom MD simulations, revealing two novel genetic connections, CRHR2 and PSPH, to memory pathologies.

We demonstrated that the common polygenetic risk of AD contributed to episodic memory decline, a finding that may not be surprising in a cohort that included participants with AD. However,

even after excluding a small number of LonGenity participants who either had or were at high risk for developing dementia (subjects with memory impairment and cognitive or functional decline during their follow-up period), the association between AD PRS and episodic memory decline remained significant. This observation persisted in the ROSMAP replication cohort that excluded participants with AD diagnosis. We propose two potential explanations for these findings. First, some participants in both cohorts may have had early preclinical stages of AD pathology. Second, there may be a genetic overlap between AD and episodic memory decline contributing to the observed association. Despite this, AD common polygenetic risk only accounted for 2% to 3% of the variance in residual memory slope in both discovery and replication cohorts. This contribution may be expected to be more substantial in a cohort in which AD progression is the major factor underlying episodic memory decline. Nevertheless, acknowledging the contribution of AD PRS to age-related episodic memory decline, irrespective of AD diagnoses, may aid in risk prediction and identification of underlying biological causes.

We identified a total of 18 rare variant associations for episodic memory decline, 16 at the variant and 2 at the gene level. Replicating single rare variant associations was often unfeasible due to their low occurrence, given the limited sample size of the ROSMAP replication study cohort ($n = 526$). However, using the ADSP cohort with a substantially larger sample size ($n = 9976$), we confirmed the association of AD risk with the putative risk rare coding variant in *LRP1B*, which has been previously shown to protect against AD pathology.[46] At the gene level, we were able to replicate rare variants association in two genes (*ITSN1* and *CRHR2*) in an independent cohort of EUR ancestry discovered in our AJ cohort. Overall, we identified three robust rare-variant associations, supported by replications in independent cohorts, that implicated distinct pathogenic pathways in episodic memory decline—AD (*LRP1B*), hippocampal plasticity (*ITSN1*), and corticotropin signaling hormone (*CRHR2*; Table 2). Examining other putative risk genes implicated by rare variants revealed further evidence of hippocampal pathologies. Similar to *ITSN1*, *ZZEF1* also contributed to the modulation of hippocampal plasticity, and their disruption caused memory impairment in mouse knockout models.[59] *SEZ6* is known to play a role in enhancing synaptic connectivity.[60] Additionally, there are other putative risk genes for episodic memory impairment supported by mouse knock-out models through different mechanisms, such as hyperphosphorylated tau (*NAGLU*),[61] and brain *STAT5A* signaling.[62]

We have replicated the finding of risk genes/pathways implicated by rare variants uncovered in our AJs in an independent cohort of EUR ancestry. While we did not replicate individual memory-associated rare variants that could be ethnicity specific, the implicated genes and pathways can be used as targets to develop drugs with the potential to be effective for all ethnicities as human biology is universe.

The hippocampus is crucial for learning and memory; studies in mammals and humans have consistently shown a decline in hippocampal neurogenesis during physiological aging.[63,64] However, the exact molecular mechanisms underlying this decline remain unknown. Interestingly, we identified rare variants in genes (*ZZEF1*, *ITSN1*, *SEZ6*, *CRHR2*, *PSPH*) that are known to implicate hippocampal functions,[54,56,59,65] indicating their involvement in episodic memory decline in our aging cohort.

We were unable to replicate the findings from prior GWAS of episodic memory, cognitive function, AD-and-related phenotypes.[15,66] This may result from differences in study populations, variations in cognitive tests included in phenotype construction, and small sample sizes. In particular, our study cohort was enriched with offspring (49.1%) of centenarians. It is known that centenarians harbor beneficial variants against aging-associated diseases.[67,68] Having inherited protective genotypes that result in resilience to episodic memory decline in our study cohort may compromise the power of identifying pathogenic risk variants. Nevertheless, earlier studies have found other variants within the three risk loci regions in association with related phenotypes, such as pre-frontal cortex development, age-related cognitive decline, cognitive decline in AD, and education attainment,[15,66,69,70] indicating association of these loci to memory formation and general cognitive ability. Our epigenome fine-mapping results suggested that the three GWAS SNPs were in transcriptionally active regions (Figure 2). rs35990795 was an eQTL of *CLIC6*, which has been implicated in learning,[71] in the putamen that is important in learning, motor control, language, reward, and other cognitive functioning.[72] rs548640610 was located within the neuron and astrocyte-specific regulatory elements (Figure 2E) and implicated *CHRM3*. *CHRM3* is a member of the larger family of GPCRs, which influence both central and peripheral nervous system processes through interaction with acetylcholine. Mutations in this gene previously have been associated AD.[73]

In recent years, the importance of protein interfaces as hubs for disease-associated variants has become more widely recognized.[74,75] An interface mutation can destabilize the interface, prevent the partner from binding, alter the partner's binding affinity, or stabilize the overall binding complex.[76] Such variants produce significant perturbations and have been shown to be associated with disease phenotypes.[49] In this study, we identified two variants located at the binding interface of CRHR2 and PSPH. As a proof of concept, CRHR2 and PSPH interfacial variants were systematically studied to uncover their molecular insights using protein modeling and all-atom MD simulations. MD simulations have been successfully used in similar previously published studies and have shown consistency with experimental validation results.[77–80] Our results showed that both CRHR2 and PSPH interfacial variants destabilized the binding interface and substantially altered the binding affinity of their respective natural ligands (Figures 4 and 5). Here, our CRHR2 findings are supported by previous studies which showed that CRHR2 null mice exhibited a phenotype of mild cognitive impairment as well as pathological changes indicative of early AD.[52] Impaired corticotrophin hormone (CRH) signaling via CRHR1 and CRHR2 has also been linked to learning and memory.[54,81] Impaired CRH signaling has been shown to alter hippocampal neurogenesis, spatial memory, and the activity of hippocampal neural stem cells (hiNSCs) under physiological conditions.[54] These compelling results support our finding of a rare interfacial variant CRHR2:p.Arg148Trp that is associated with episodic memory decline. Similarly, PSPH:p.Thr152Ile simulation results demonstrated that phosphoserine bound more stably to wild type PSPH than to

mutant PSPH (Figure 5). PSPH:p.Thr152Ile also altered the binding stability of phosphoserine with critical residues of PSPH (Figure S32). In a mouse model, the treatment of L-serine demonstrated a protective effect on hippocampal neurons, guarding against oxidative stress–mediated mitochondrial damage and apoptotic cell death.[56] Altogether, our findings indicate that Thr152Ile is likely to diminish PSPH enzymatic activity via altering phosphoserine binding to PSPH.

While we provided mechanistic insights into two interfacial variants, it should not be inferred that other coding variants do not play a role in memory pathology. We limited the detailed characterization of variants to a subset of interfacial variants to avoid over-interpretation of MD simulation results for which we lacked a convincing protein-relevant mechanism. Due to the approximations and assumptions used in MD simulations parameters, values obtained here should be interpreted qualitatively rather than quantitatively.

We have demonstrated that both common and rare genetic variants play important roles in contributing to episodic memory decline; however, future molecular studies would be required to confirm their underlying biological mechanisms. Rare coding risk variants, which may have larger effect sizes than common variants, are particularly valuable for drug discovery. Integrating these rare variants into genetic risk modeling with common variants could improve risk prediction of episodic memory decline and AD. However, the inclusion of rare variants in risk prediction is still in its early stages, and its potential depends on the statistical power of rare variant studies. Overall, our findings advance the current understanding of the biology of episodic memory decline and could lead to new drug targets for preventing or treating memory impairment.

ACKNOWLEDGMENTS

We are grateful to the individuals for their persistent participation in our longitudinal study of memory decline. We thank the Sequencing and Lab Operations team at Regeneron RGC for their execution and oversight of the exome sequencing. We are also thankful for the support from Gregory Klein for his help in accessing RADC data. This work was supported by Foundation for the National Institutes of Health, R01AG061155 (SM), U19AG056278 (ZDZ), P01AG017242 (ZDZ), and P01AG047200 (ZDZ).

CONFLICT OF INTEREST STATEMENT

All authors declare no competing interests. Author disclosures are available in supporting information.

CONSENT STATEMENT

Written informed consent was obtained from all study participants.

REFERENCES

1. Ardila A. Normal aging increases cognitive heterogeneity: analysis of dispersion in WAIS-III scores across age. *Arch Clin Neuropsychol*. 2007;22:1003-1011
2. Harada CN, Love MCN, Triebel KL. Normal cognitive aging. *Clin Med Geriatr*. 2013;29:737-752
3. Lindenberger U. Human cognitive aging: corriger la fortune?. *Science*. 2014;346:572-578
4. Nyberg L, Maitland SB, Rönnlund M, et al. Selective adult age differences in an age-invariant multifactor model of declarative memory. *Psychol Aging*. 2003;18:149-160
5. Rönnlund M, Nyberg L, Backman L, Nilsson LG. Stability, growth, and decline in adult life span development of declarative memory: cross-sectional and longitudinal data from a population-based study. *Psychol Aging*. 2005;20:3-18
6. Backman L, Jones S, Berger AK, Laukka EJ, Small BJ. Multiple cognitive deficits during the transition to Alzheimer's disease. *J Intern Med*. 2004;256:195-204
7. Dubois B, Feldman HH, Jacova C, et al. Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria. *Lancet Neurol*. 2007;6:734-746
8. Panizzon MS, Lyons MJ, Jacobson KC, et al. Genetic architecture of learning and delayed recall: a twin study of episodic memory. *Neuropsychology*. 2011;25:488-498
9. Wilson RS, Barral S, Lee JH, et al. Heritability of different forms of memory in the late onset Alzheimer's disease family study. *J Alzheimers Dis*. 2011;23:249-255
10. Nyberg L. Functional brain imaging of episodic memory decline in ageing. *J Intern Med*. 2017;281:65-74
11. Lee T, Henry JD, Trollor JN, Sachdev PS. Genetic influences on cognitive functions in the elderly: a selective review of twin studies. *Brain Res Rev*. 2010;64:1-13
12. McGue M, Christensen K. The heritability of level and rate-of-change in cognitive functioning in Danish twins aged 70 years and older. *Exp Aging Res*. 2002;28:435-451
13. McGue M, Christensen K. Social activity and healthy aging: a study of aging Danish twins. *Twin Res Hum Genet*. 2007;10:255-265
14. De Jager PL, Shulman JM, Chibnik LB, et al. A genome-wide scan for common variants affecting the rate of age-related cognitive decline. *Neurobiol Aging*. 2012;33:1017
15. Kamboh MI, Fan KH, Yan Q, et al. Population-based genome-wide association study of cognitive decline in older adults free of dementia: identification of a novel locus for the attention domain. *Neurobiol Aging*. 2019;84:239.e15-239.e24
16. Zhang C, Pierce BL. Genetic susceptibility to accelerated cognitive decline in the US health and retirement study. *Neurobiol Aging*. 2014;35:1512.e11-8
17. Gao Y, Felsky D, Reyes-Dumeyer D, et al. Integration of GWAS and brain transcriptomic analyses in a multiethnic sample of 35,245 older adults identifies DCDC2 gene as predictor of episodic memory maintenance. *Alzheimers Dement*. 2022;18:1797-1811
18. Arpawong TE, Pendleton N, Mekli K, et al. Genetic variants specific to aging-related verbal memory: insights from GWASs in a population-based cohort. *PLoS One*. 2017;12:e0182448
19. Montaron MF, Drapeau E, Dupret D, et al. Lifelong corticosterone level determines age-related decline in neurogenesis and memory. *Neurobiol Aging*. 2006;27:645-654
20. de la Torre JC, Fortin T, Park GA, et al. Chronic cerebrovascular insufficiency induces dementia-like deficits in aged rats. *Brain Res*. 1992;582:186-195.
21. Frater J, Lie D, Bartlett P, McGrath JJ. Insulin-like growth factor 1(IGF-1) as a marker of cognitive decline in normal ageing: a review. *Ageing Res Rev*. 2018;42:14-27
22. Lupien SJ, de Leon M, de Santi S, et al. Cortisol levels during human aging predict hippocampal atrophy and memory deficits. *Nat Neurosci*. 1998;1:69-73
23. Slatkin M. A population-genetic test of founder effects and implications for Ashkenazi Jewish diseases. *Am J Hum Genet*. 2004;75:282-293
24. Avila-Rieger J, Turney IC, Vonk JMJ, et al. Socioeconomic status, biological aging, and memory in a diverse national sample of older US men and women. *Neurology*. 2022;99:e2114-e2124

25. Wechsler D. *WMS-R : Wechsler Memory Scale–Revised: Manual* Psychological Corp. Harcourt Brace Jovanovich; 1987
26. Grober E, Merling A, Heimlich T, Lipton RB. Free and cued selective reminding and selective reminding in the elderly. *J Clin Exp Neuropsychol*. 1997;19:643-654
27. Duff K, Leber WR, Patton DE, et al. Modified scoring criteria for the RBANS figures. *Appl Neuropsychol*. 2007;14:73-83
28. Krasheninina O, Hwang YC, Bai X, et al. Open-source mapping and variant calling for large-scale NGS data from original base-quality scores. *BioRxiv*. 2020:2020.12.15.356360
29. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet*. 2014;46:310-315
30. Itan Y, Shang L, Boisson B, et al. The mutation significance cutoff: gene-level thresholds for variant predictions. *Nat Methods*. 2016;13:109-110
31. Lee S, Emond MJ, Bamshad MJ, et al. Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. *Am J Hum Genet*. 2012;91:224-237
32. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. *Bioinformatics*. 2011;27:1739-1740
33. Bennett DA, Buchman AS, Boyle PA, Barnes LL, Schneider JA. Religious orders study and rush memory and aging project. *J Alzheimers Dis*. 2018;64:S161-S189
34. De Jager PL, Ma Y, McCabe C, et al. A multi-omic atlas of the human frontal cortex for aging and Alzheimer's disease research. *Sci Data*. 2018;5:180142
35. Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021;596:583-589
36. Bowers KJ, Chow E, Xu H, et al. Scalable algorithms for molecular dynamics simulations on commodity clusters. In: *Proceedings of the 2006 ACM/IEEE Conference on Supercomputing* 84-es. ACM; 2006
37. Martyna GJ, Klein ML, Tuckerman M. Nosé–Hoover chains: the canonical ensemble via continuous dynamics. *J Chem Phys*. 1992;97:2635-2643
38. Martyna GJ, Tobias DJ, Klein ML. Constant pressure molecular dynamics algorithms. *J Chem Phys*. 1994;101:4177-4189
39. Essmann U, Lalith Perera, Max L, et al. A smooth particle mesh Ewald method. *J Chem Phys*. 1995;103:8577-8593
40. Li J, Abel R, Zhu K, Cao Y, Zhao S, Friesner RA. The VSGB 2.0 model: a next generation energy model for high resolution protein structure modeling. *Proteins: Struct, Funct, Bioinf*. 2011;79:2794-2812
41. Nott A, Holtman IR, Coufal NG, et al. Brain cell type-specific enhancer-promoter interactome maps and disease-risk association. *Science*. 2019;366:1134-1139
42. Hnisz D, Abraham BJ, Lee TI, et al. Super-enhancers in the control of cell identity and disease. *Cell*. 2013;155:934-947
43. Consortium GT. The genotype-tissue expression (GTEx) project. *Nat Genet*. 2013;45:580-585
44. Bellenguez C, Küçükali F, Jansen IE, et al. New insights into the genetic etiology of Alzheimer's disease and related dementias. *Nat Genet*. 2022;54:412-436
45. Nicolas G, Charbonnier C, Campion D. From common to rare variants: the genetic component of Alzheimer disease. *Hum Hered*. 2016;81:129-141
46. Benoit ME, Hernandez MX, Dinh ML, Benavente F, Vasquez O, Tenner AJ. C1q-induced LRP1B and GPR6 proteins expressed early in Alzheimer disease mouse models, are essential for the C1q-mediated protection against amyloid-beta neurotoxicity. *J Biol Chem*. 2013;288:654-665
47. Huang JY, Hafez DM, James BD, Bennett DA, Marr RA. Altered NEP2 expression and activity in mild cognitive impairment and Alzheimer's disease. *J Alzheimers Dis*. 2012;28:433-441
48. Khan S, Vihinen M. Spectrum of disease-causing mutations in protein secondary structures. *BMC Struct Biol*. 2007;7:56
49. David A, Sternberg MJ. The contribution of missense mutations in core and rim residues of protein-protein interfaces to human disease. *J Mol Biol*. 2015;427:2886-2898
50. De Souza EB. Corticotropin-releasing factor receptors: physiology, pharmacology, biochemistry and role in central nervous system and immune disorders. *Psychoneuroendocrinology*. 1995;20:789-819
51. Ma S, Shen Q, Zhao LH, et al. Molecular basis for hormone recognition and activation of corticotropin-releasing factor receptors. *Mol Cell*. 2020;77:669-680
52. Rissman RA, Staup MA, Lee AR, et al. Corticotropin-releasing factor receptor-dependent effects of repeated stress on tau phosphorylation, solubility, and aggregation. *Proc Natl Acad Sci*. 2012;109:6277-6282
53. Deussing JM, Breu J, Kühne C, et al. Urocortin 3 modulates social discrimination abilities via corticotropin-releasing hormone receptor type 2. *J Neurosci*. 2010;30:9103-9116
54. Koutmani Y, Gampierakis IA, Polissidis A, et al. CRH promotes the neurogenic activity of neural stem cells in the adult hippocampus. *Cell Rep*. 2019;29:932-945
55. Kim HY, Heo YS, Kim JH, et al. Molecular basis for the local conformational rearrangement of human phosphoserine phosphatase. *J Biol Chem*. 2002;277:46651-46658
56. Kim KY, Hwang SK, Park SY, Kim MJ, Jun DY, Kim YH. l-Serine protects mouse hippocampal neuronal HT22 cells against oxidative stress-mediated mitochondrial damage and apoptotic cell death. *Free Radic Biol Med*. 2019;141:447-460
57. Le Douce J, Maugard M, Veran J, et al. Impairment of glycolysis-derived L-serine production in astrocytes contributes to cognitive deficits in Alzheimer's disease. *Cell Metab*. 2020;31:503-517
58. Gong CX, Singh TJ, Grundke-Iqbal I, Iqbal K. Phosphoprotein phosphatase activities in Alzheimer disease brain. *J Neurochem*. 1993;61:921-927
59. Segovia-Miranda F, Serrano F, Dyrda A, et al. Pathogenicity of lupus anti-ribosomal P antibodies: role of cross-reacting neuronal surface P antigen in glutamatergic transmission and plasticity in a mouse model. *Arthritis Rheumatol*. 2015;67:1598-1610
60. Gunnersen JM, Kim MH, Fuller SJ, et al. Sez-6 proteins affect dendritic arborization patterns and excitability of cortical pyramidal neurons. *Neuron*. 2007;56:621-639
61. Ohmi K, Kudo LC, Ryazantsev S, Zhao H-Z, Karsten SL, Neufeld EF. Sanfilippo syndrome type B, a lysosomal storage disease, is also a tauopathy. *Proc Natl Acad Sci*. 2009;106:8332-8337
62. Furigo IC, Melo HM, Lyra E Silva NM, et al. Brain STAT5 signaling modulates learning and memory formation. *Brain Struct Funct*. 2018;223:2229-2241
63. Imayoshi I, Sakamoto M, Ohtsuka T, et al. Roles of continuous neurogenesis in the structural and functional integrity of the adult forebrain. *Nat Neurosci*. 2008;11:1153-1161
64. Moreno-Jiménez EP, Flor-García M, Terreros-Roncal J, et al. Adult hippocampal neurogenesis is abundant in neurologically healthy subjects and drops sharply in patients with Alzheimer's disease. *Nat Med*. 2019;25:554-560
65. Jakob B, Kochlamazashvili G, Jäpel M, et al. Intersectin 1 is a component of the reelin pathway to regulate neuronal migration and synaptic plasticity in the hippocampus. *Proc Natl Acad Sci*. 2017;114:5533-5538
66. Adewuyi EO, O'Brien EK, Nyholt DR, Porter T, Laws SM. A large-scale genome-wide cross-trait analysis reveals shared genetic architecture between Alzheimer's disease and gastrointestinal tract disorders. *Commun Biol*. 2022;5:691
67. Milman S, Barzilai N. Dissecting the mechanisms underlying unusually successful human health span and life span. *Cold Spring Harb Perspect Med*. 2015;6:a025098

68. Lin J-R, Sin-Chan P, Napolioni V, et al. Rare genetic coding variants associated with human longevity and protection against age-related diseases. *Nat Aging*. 2021;1:783-794
69. Naqvi S, Sleyp Y, Hoskens H, et al. Shared heritability of human face and brain shape. *Nat Genet*. 2021;53:830-839
70. Sherva R, Tripodis Y, Bennett DA, et al. Genome-wide association study of the rate of cognitive decline in Alzheimer's disease. *Alzheimers Dement*. 2014;10:45-52
71. Ferreira E, Shaw DM, Oddo S. Identification of learning-induced changes in protein networks in the hippocampi of a mouse model of Alzheimer's disease. *Transl Psychiatry*. 2016;6:e849
72. Luo X, Mao Q, Shi J, Wang X, Li CR. Putamen gray matter volumes in neuropsychiatric and neurodegenerative disorders. *World J Psychiatry Ment Health Res*. 2019:3
73. Chee LY, Cumming A. Polymorphisms in the cholinergic receptors muscarinic (*CHRM2* and *CHRM3*) genes and Alzheimer's disease. *Avicenna J Med Biotechnol*. 2018;10:196-199
74. Jubb HC, Pandurangan AP, Turner MA, Ochoa-Montaño B, Blundell TL, Ascher DB. Mutations at protein-protein interfaces: small changes over big surfaces have large impacts on human health. *Prog Biophys Mol Biol*. 2017;128:3-13
75. Wang X, Wei X, Thijssen B, Das J, Lipkin SM, Yu H. Three-dimensional reconstruction of protein networks provides insight into human genetic disease. *Nat Biotechnol*. 2012;30(2):159-164
76. Li E, You M, Hristova K. FGFR3 dimer stabilization due to a single amino acid pathogenic mutation. *J Mol Biol*. 2006;356:600-612
77. Ratnapriya R, Acar İE, Geerlings MJ, et al. Family-based exome sequencing identifies rare coding variants in age-related macular degeneration. *Hum Mol Genet*. 2020;29:2022-2034
78. Cheng F, Zhao J, Wang Y, et al. Comprehensive characterization of protein-protein interactions perturbed by disease mutations. *Nat Genet*. 2021;53:342-353
79. Suomivuori CM, Latorraca NR, Wingler LM, et al. Molecular mechanism of biased signaling in a prototypical G protein-coupled receptor. *Science*. 2020;367:881-887
80. Likic VA, Gooley PR, Speed TP, Strehler EE. A statistical approach to the interpretation of molecular dynamics simulations of calmodulin equilibrium dynamics. *Protein Sci*. 2005;14:2955-2963
81. Short AK, Maras PM, Pham AL, Ivy AS, Baram TZ. Blocking CRH receptors in adults mitigates age-related memory impairments provoked by early-life adversity. *Neuropsychopharmacology*. 2020;45:515-523

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Ali A, Milman S, Weiss EF, et al. Genetic variants associated with age-related episodic memory decline implicate distinct memory pathologies. *Alzheimer's Dement*. 2024;1-17. https://doi.org/10.1002/alz.14379