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

RESEARCH ARTICLE

The genetic landscape of early-onset Alzheimer's disease in China

Wei Qin[1,2,3,4,5] | **Fang-Yu Li**[1,2,3,4,5] | **Wen-Ying Liu**[1,2,3,4,5] | **Ying Li**[1,2,3,4,5] | **Shu-Man Cao**[1,2,3,4,5] | **Yi-Ping Wei**[1,2,3,4,5] | **Yan Li**[1,2,3,4,5] | **Qi Wang**[1,2,3,4,5] | **Qi-Geng Wang**[1,2,3,4,5] | **Jian-Ping Jia**[1,2,3,4,5] (iD)

[1]Innovation Center for Neurological Disorders and Department of Neurology, Xuanwu Hospital, Capital Medical University, National Clinical Research Center for Geriatric Diseases, Beijing, China

[2]Beijing Key Laboratory of Geriatric Cognitive Disorders, Beijing, China

[3]Clinical Center for Neurodegenerative Disease and Memory Impairment, Capital Medical University, Beijing, China

[4]Center of Alzheimer's Disease, Beijing Institute of Brain Disorders, Collaborative Innovation Center for Brain Disorders, Capital Medical University, Beijing, China

[5]Key Laboratory of Neurodegenerative Diseases, Ministry of Education, Beijing, China

Correspondence
Jian-Ping Jia, 45 Changchun Street, Xicheng District, Beijing 100053, China.
Email: jjp@ccmu.edu.cn, jiajp@vip.126.com

Abstract

INTRODUCTION: Research on somatic and germline mutations in Chinese individuals with early-onset Alzheimer's disease (EOAD) has been limited.

METHODS: We conducted whole-genome sequencing of blood DNA from 108 patients with EOAD and 116 controls. The analysis included somatic and germline mutations across coding and non-coding regions, mutational signature determination, pathway enrichment identification, and predictive model.

RESULTS: The mutational burden was significantly higher in the EOAD group compared to the control group. The prevalence of single-base substitution signature 5, which is strongly associated with aging, was much higher in patients with EOAD than in controls. EOAD-specific somatic mutations were identified in genes such as *MIR31HG*, *TUBB4B*, and *APP*. Germline mutations in *DOCK3*, *PCSK5*, and *PDE4D* were significantly associated with age of dementia onset. Furthermore, a predictive model comprising 15 mutations demonstrated an area under the curve of 0.78.

DISCUSSION: The accumulation of senescence-related somatic mutations may increase the risk of developing EOAD.

KEYWORDS
early-onset Alzheimer's disease, germline mutation, predictive model, somatic mutation, whole genome sequencing

Highlights

- Whole genome sequencing was used to find somatic and germline mutations in Chinese individuals with early-onset Alzheimer's disease (EOAD).
- Total number and burden of blood somatic mutations were significantly higher.
- The prevalence of single-base substitution signature 5 was notably elevated in EOAD.

- EOAD-specific somatic mutations were identified in *MIR31HG*, *TUBB4B*, and *APP*.
- *DOCK3*, *PCSK5*, and *PDE4D* germline mutations were associated with the age of EOAD onset.

1 | INTRODUCTION

Alzheimer's disease (AD) is a profoundly debilitating neurodegenerative disorder and is the most prevalent subtype of dementia. Based on age at onset (AAO), AD is classified into early-onset AD (EOAD; AAO < 65 years) or late-onset AD (LOAD; AAO ≥ 65 years). LOAD is a complex disorder with a heterogenous etiology and heritability of 58% to 79%. However, EOAD is predominantly genetically determined with a heritability of > 90%.[1] Although individuals with EOAD constitute only 5% to 10% of all AD cases, they are more likely to experience a severe clinical course and take longer to receive an accurate diagnosis, thus warranting special consideration and study.

Only 6% to 11% of individuals with EOAD carry established autosomal dominant mutations in the amyloid precursor protein (*APP*) or presenilin 1 and 2 (*PSEN1* and *PSEN2*) genes, recognized as causative factors for AD.[2–4] This suggests that new genetic factors remain to be discovered. Genome-wide association studies, polymerase chain reaction (PCR), and SNaPshot sequencing have identified variants in *SORL1*, *DHCR7*, *TREM2*, and other genes that increase the risk of developing EOAD.[5–8] Whole genome sequencing (WGS) provides the opportunity to obtain the most comprehensive genetic variation of an individual and allows for detailed evaluations of all genetic variations.[9,10] Missense variants, including *TBK1*, *ACAA1*, and *DPP6* genes, have been identified in EOAD using WGS.[11–13] However, these studies have primarily focused on common germline mutations.

The human genome is continually exposed to both external and internal mutagens, leading to the emergence and accumulation of somatic mutations throughout development and aging. This accumulation may stem from DNA replication errors and extensive oxidative stress, compounded by gradual defects in DNA repair mechanisms. Patients lacking germline pathogenic variants in autosomal dominant AD genes may harbor somatic variants. It is plausible that, similar to inherited or de novo germline pathogenic variants, somatic variants with high penetrance can contribute to early onset.[14] Recent studies have highlighted the potential role of somatic mutations in neurodegenerative diseases, including AD.[15] Somatic mutations, also known as mosaic mutations, are acquired postfertilization and are present in a subset of an individual's cells, affecting only those derived from the initially mutated cell.[16] Unlike germline mutations, somatic mutations are postzygotic genetic alterations that are not inherited from the parents, leading to genetically distinct cell populations within an organism.[17] Somatic mutations in the *APP*, *PSEN1*, and *PSEN2* loci have also been reported.[18–20] These mutations demonstrate a gene dosage effect on the age of disease onset and phenotypic severity,[21] and have been implicated in various neurological disorders, including AD, Parkinson's disease, and schizophrenia. Both germline and somatic mutations may influence the incidence and presentation of these disorders.[20,22] Furthermore, the effects of many risk loci vary among ethnic groups.[23] Hence, we reasoned that focusing on a subset of individuals of Chinese ethnicity with EOAD might uncover unique somatic or germline mutations.

Somatic mutations may be carried by postmitotic neurons and expanded by replicative myeloid cells in the central nervous system (CNS) and peripheral tissues.[1] Given that the burden of somatic mutations is five times higher in the blood than in the brain,[17] we conducted WGS on blood tissues from 108 patients with EOAD and 116 normal controls within the Chinese population. This study aimed to identify novel somatic mutations in EOAD among Chinese individuals, compare genomic features, identify differences between EOAD patients and healthy controls, and assess the genome-wide landscape to potentially explain EOAD etiology and discover new genetic markers specific to the Chinese population.

2 | METHODS

2.1 | Patient recruitment

In total, 224 individuals, comprising 108 EOAD and 116 healthy normal controls (NCs), were included in this study. All AD diagnoses were determined according to the recommendations set by the National Institute on Aging–Alzheimer's Association workgroup or National Institute of Neurological[24] and Communicative Disorders and Stroke–Alzheimer's Disease and Related Disorders Association criteria.[25] Individuals with a diagnosis of EOAD and no family history of dementia were included in the study. To include more cases with a genetic component, we enrolled EOAD patients with an age at onset of < 60 years. Participants were recruited following ethical guidelines. All controls were cognitively normal (without subjective memory complaints, Mini-Mental State Examination score of 26–30, and Clinical Dementia Rating scale score of 0), and free of any general or laboratory evidence of diseases that could affect cognition. The EOAD group and control group were generally matched in age and sex. All samples were derived from the China Cognition and Aging Study (COAST), which is a multicenter cohort study comprising clinical diagnosis, disease progression, genetic regulation, and drug trials across 30/31 provinces in China. Signed informed consents were provided by all the patients and control subjects. The study protocol was approved and monitored by the ethics committee of Xuanwu Hospital.

2.2 | Deep WGS

Genomic DNA was obtained from the peripheral blood of all participants. High-quality gDNA samples were sent to the Beijing Genomics Institute (BGI) for deep WGS. Sequencing libraries were prepared using the TruSeq DNA PCR-Free Library Preparation Kit (Illumina) to minimize bias introduced by PCR amplification. Each library was sequenced on the Illumina HiSeq 2000 platform, generating 150 bp paired-end reads. Sequencing was performed to achieve an average coverage depth of $\approx 32 \times$ per sample, ensuring comprehensive and accurate genome representation.

The raw sequencing data were subjected to quality control checks using FastQC (v0.10.1) to assess read quality, guanosine–cytosine content, and the presence of adapter sequences. High-quality reads were then aligned to the human reference genome (hg19) using the BWA-MEM algorithm (v0.7.17). Post-alignment, the Binary Alignment Map files were processed to remove duplicates using Picard tools (v2.18.14), followed by base quality score recalibration and indel realignment with the Genome Analysis Toolkit (GATK, v4.1.4.1) to prepare the data for downstream analyses.

2.3 | Germline mutation calling

Germline mutations were identified using the GATK HaplotypeCaller (v4.1.4.1). Variants were called on each sample individually, followed by joint genotyping across all samples to produce a multisample variant call format (VCF) file. To ensure high confidence in the identified variants, we applied standard hard filters recommended by GATK: single nucleotide polymorphisms (SNPs) were filtered with "QD < 2.0 || FS > 60.0 || MQ < 40.0 || MQRankSum < −12.5 || ReadPosRankSum < −8.0," and indels were filtered with "QD < 2.0 || FS > 200.0 || ReadPosRankSum < −20.0." The resulting variants were annotated using ANNOVAR,[26] and common polymorphisms were filtered out by comparing against the gnomAD (v2.1.1) and dbSNP (build 151) databases. Rare or novel variants not present in these databases were prioritized for further analysis.

2.4 | Somatic mutation calling

Somatic mutations were identified following a multistep approach. Initially, germline variants identified by GATK HaplotypeCaller were used to create a panel of normal (PoN) from control samples using GATK's Create Somatic Panel of Normals tool. This PoN was used to filter out technical artifacts and common germline variants in subsequent somatic mutation calling.

Somatic variants were first called using MuTect2 (v4.1.4.1), with the PoN database applied to increase specificity. Variants tagged as "str_contraction," "triallelic_site," or "t_lod_fstar" were excluded. The preliminary somatic calls were further refined using MosaicForecast, which uses a statistical model to distinguish true somatic variants from sequencing artifacts. MosaicForecast was applied to the MuTect2

RESEARCH IN CONTEXT

1. **Systematic review**: The authors conducted a literature review on the genetics of early-onset Alzheimer's disease (EOAD) using PubMed, meeting abstracts, and presentations. However, there is a notable lack of studies on somatic and germline mutations in Chinese individuals with EOAD.
2. **Interpretation**: The prevalence of single-base substitution signature 5 was found to be significantly higher in Chinese EOAD patients. EOAD-specific somatic mutations were identified in genes such as *MIR31HG*, *TUBB4B*, and *APP*, which are involved in cellular senescence—a major contributor to human aging. Additionally, germline mutations in *DOCK3*, *PCSK5*, and *PDE4D* were strongly associated with the age of dementia onset.
3. **Future directions**: Further validation and more extensive genomic coverage in larger cohorts are required to confirm the role of somatic mutations in AD. Future research should investigate both somatic and germline mutations across the genome, as they may contribute to the risk of EOAD.

output, focusing on candidate somatic variants with variant allele frequencies (VAFs) < 0.4 and excluding those present in the gnomAD database to filter out likely germline variants.

Post calling, somatic variants were further filtered using the following criteria: variants with low confidence scores, those present in common polymorphism databases (gnomAD, dbSNP), and those with VAFs < 0.02 or > 0.4 were excluded. Additionally, variants found in repetitive regions and segmental duplications were removed. Variants passing these filters were considered high-confidence somatic mutations. The resulting mutation annotation format (MAF) files were analyzed using the R package maftools (v2.4.05) to summarize the mutation burden, variant types, and recurrently mutated genes. Driver mutations were identified using the OncodriveCLUST algorithm, focusing on genes known to be involved in neurodegenerative diseases.

2.5 | Mutation annotation and statistical analysis

Variants were annotated using ANNOVAR, a comprehensive tool for annotating variants from VCF files. ANNOVAR provided annotations including mutation types, genomic positions, and gene-based annotations, enabling detailed characterization of mutations across the genome. From the annotated data, MAF files were generated, consolidating information on mutation types, genomic coordinates, and gene identities.

These MAF files served as standardized inputs for subsequent statistical analyses. Statistical analyses were conducted using maftools in R.[27] Maftools facilitated comprehensive statistical exploration of

mutation data, including mutation burden analysis, identification of significantly mutated genes, and comparison of mutation profiles between AD patients and non-AD individuals. Statistical significance was assessed using appropriate tests to discern meaningful differences relevant to AD pathogenesis.

2.6 | Functional annotation and impact prediction

Both somatic and germline mutations were comprehensively annotated by FunSeq2.[28] Because non-coding variants in regulatory elements (promoter, enhancer, etc.) can be associated with potential target genes, this pipeline helps to identify both coding and non-coding variants of a given gene. Additionally, the functional impact of each variant was predicted using PredictSNP2,[29] categorizing variants as neutral, deleterious, or unknown. Only variants predicted to be deleterious, with high functional impact, were selected for gene- and pathway-level analysis.

To predict if variants affect the protein structure and functions, we set up the following pipeline. First, for each protein, the sequence of the wild-type protein was retrieved from UniProt and truncated according to the mutation. For each AlphaFold2 prediction, we used AlphaFold v2.2.0 with the full BFD database for sequence alignment. We predicted five structures for each group and the resulting structures were ranked by predicted local distance difference tests. The top-ranked complex structure was picked for further analysis. For visualization, we use PyMOL 2.3.1 to depict the protein 3D structure and further added labels in Adobe Illustrator.

2.7 | Mutation signature analysis

To elucidate differences in mutational processes between AD and non-AD subjects, somatic mutations from WGS data were pooled separately for each group. Mutational signatures were then constructed using Mutalisk,[30] categorizing mutations into 96 substitution classes. The decomposition of mutational signatures used a linear regression maximum likelihood estimation (MLE) method, with reference to 30 standard mutational signatures from COSMIC (Catalogue of Somatic Mutations in Cancer). The best-fit combination of known signatures for each group was determined based on Cosine similarity scores.

2.8 | Gene-set enrichment analysis

To gain insights into the roles of somatic mutations in underlying biological processes (BPs), we performed functional enrichment analysis on somatic mutations from AD samples using the "clusterProfiler" R package.[31] This analysis encompassed both Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses. GO terms and KEGG pathways with Benjamini–Hochberg adjusted P values < 0.05 were deemed statistically significant.

2.9 | Pathogenic germline mutations in AD risk genes

In total, 290 known AD pathogenic mutation sites associated with AD (e.g., pathogenic, risk modifier, possible risk modifier) in three autosomal dominant (APP: 28 sites, PSEN1: 239 sites, and PSEN2: 16 sites) and two other AD-associated genes (TREM2: 6 sites and MAPT: 1 site) were curated from the AlzGene mutation database (last updated on April 27, 2018).[32] Additionally, individuals carrying two copies of apolipoprotein E (APOE) ε4 (rs429358 [C], rs7412 [C]) were also considered to harbor pathogenic germline mutations. We used the "bamreadcount" R package to quantify the number of reference and alternative alleles in brain samples at each germline SNP site.[33] To ensure the accuracy of identified germline SNPs, we used filtered reads (MQ ≥ 20, BQ ≥ 30) and defined germline SNPs as those with a VAF of ≥ 40%. Suspected heterozygous germline SNPs identified through "bamreadcount" were further validated using GATK HaplotypeCaller (v3.5).

2.10 | Genome-wide association study

Genetic matching of EOAD cases and NC controls was performed using PCAmatchR to mitigate potential confounding from population stratification effects.[34] Linkage disequilibrium filtered variants (R2 < 0.1) were extracted from the array manifest file for the combined set of cases and controls. Principal component (PC) analysis using PLINK identified the first 20 PCs and corresponding eigenvalues, adjusting for covariates including disease status, family history, sex, and age.

Imputation of genetic variants was conducted using the Michigan Imputation Server with the TOPMed reference panel. After imputation, association analyses were carried out under an additive model using SNPTEST (v2.5.6). Any PCs that retained significance after adjusting for covariates were further included in the genome-wide association study (GWAS) association tests. For the primary GWAS, germline variants were filtered based on a control minor allele frequency (> 0.5%) and imputation quality score (> 0.7).

In stratified GWAS analyses by AD and NC groups, variants were subjected to a more stringent control minor allele frequency threshold (> 5%) to mitigate potential spurious associations arising from small sample sizes. Manhattan plots were generated for visualization of association results using the "qqman" and "hudson" R packages.

3 | RESULTS

3.1 | Quantitative comparison of somatic mutations in the AD and control groups

The description of the cohorts used in this study is provided in Table 1. We quantified somatic mutations, VAFs, and mutation subtypes of somatic mutations in patients with EOAD and individuals without AD based on WGS data. After filtering out mutations with low coverage

(A)

(B)

(C)

(D)
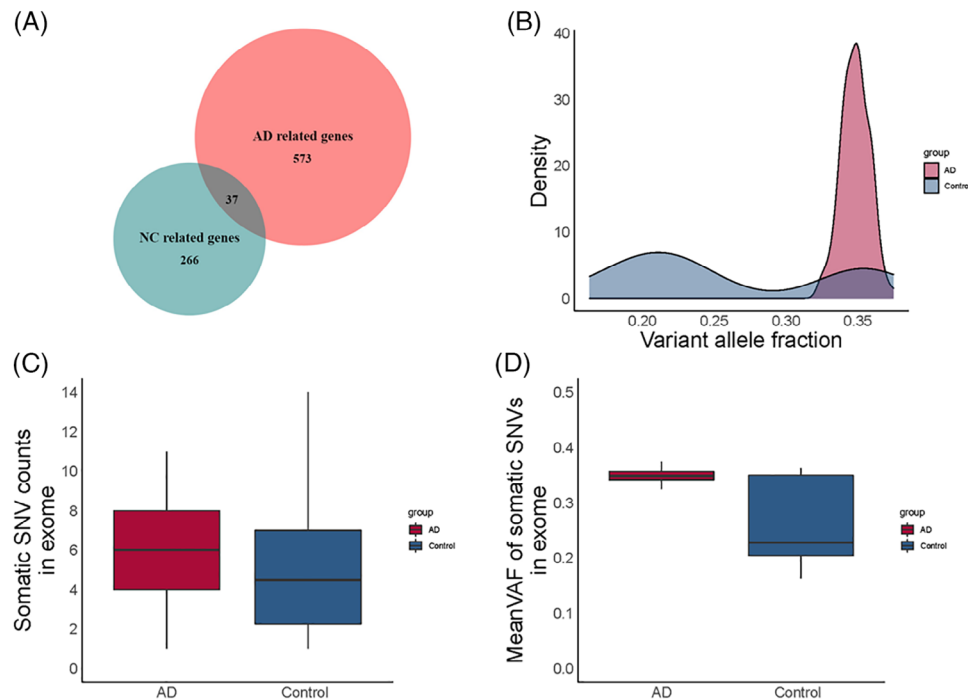
FIGURE 1 Characteristics of somatic mutations in AD and control groups based on WGS data. A, The somatic mutation gene number in AD and control. B, Distribution of VAFs of pooled somatic mutations from AD and control individuals. C, Comparison of average counts in exome of somatic mutations from AD and control individuals. D, Comparison of mean VAFs of somatic mutations in exome in each group, which are shown as box plots (center line, median; box limits, upper and lower quartiles; whiskers, maximum and minimum values). AD, Alzheimer's disease; NC, normal control; SNV, single nucleotide variant; VAFs, variant allele frequencies; WGS, whole genome sequencing

TABLE 1 Description of the cohorts used in the current study.

	AD cases no. (%) (n = 108)	Controls no. (%) (n = 116)
Female	61 (56.5)	56 (48.3)
Age at onset mean (SD)	50.4 (6.2)	NA
Age at examination mean (SD)	54.7 (7.2)	59.3 (7.8)

Abbreviations: AD, Alzheimer's disease; NA, not available; SD, standard deviation.

depth (0.4), we detected an average of 1051.06 ± 173.42 (mean ± standard deviation [SD]) somatic mutations per sample (range, 19–1472) in the EOAD group and an average of 957.11 ± 243.40 (mean ± SD) somatic mutations per sample (range, 524–1515) in the control group. A total of 573 somatic mutation genes were identified in EOAD, compared to 266 in the control group (Figure 1A). The VAFs of the detected somatic mutations ranged from 32.1% to 37.5% (median: 34.9%) in the EOAD group and from 16.3% to 36.3% (median: 22.8%) in the control group (Figure 1B). The average number of somatic mutations was significantly higher in the EOAD group compared to the control group (t test, P = 0.021; Figure 1C). The mean frequency of somatic mutations was significantly higher in the EOAD group than in the control group (t test, P < 0.001; Figure 1D). These findings indicated that both the mutation count and genomic location of the somatic mutations were significantly different between individuals with EOAD and those without AD.

For EOAD samples, the top three variant classifications were missense_mutations, in_frame_insertions, and frameshift_insertions, whereas for control samples, they were missense_mutations, in_frame_deletions, and in_frame_insertions (Figure 2A,B). The number of variant types in blood somatic mutations showed significant differences between individuals with EOAD and those without AD (P < 0.001; Figure 2B). SNPs were the most common variant type in both EOAD patients and controls (Figure 2C). The classification of blood somatic mutations also differed significantly between the two groups (P < 0.001; Figure 2C). C > T mutations accounted for 53.45% (650/1216) in the EOAD group and 54.20% (187/345) in the control group (Figure 2D), with no statistically significant difference observed. For AD-specific variants, the most common variant classification and type were missense_mutation and SNP, and C > T accounted for 53.38% (640/1199) of the mutations (Figure S1 in supporting information).

3.2 | Mutational signature analysis

Somatic cells from different tissues are exposed to various intrinsic (e.g., DNA polymerase error and impairment DNA repair mechanisms) and extrinsic (e.g., tobacco smoking and ultraviolet rays) mutagenic sources. These sources elicit distinct patterns of base alterations and their associated nucleotide contexts, known as mutational signatures. To characterize the mutational processes, we first pooled all putative
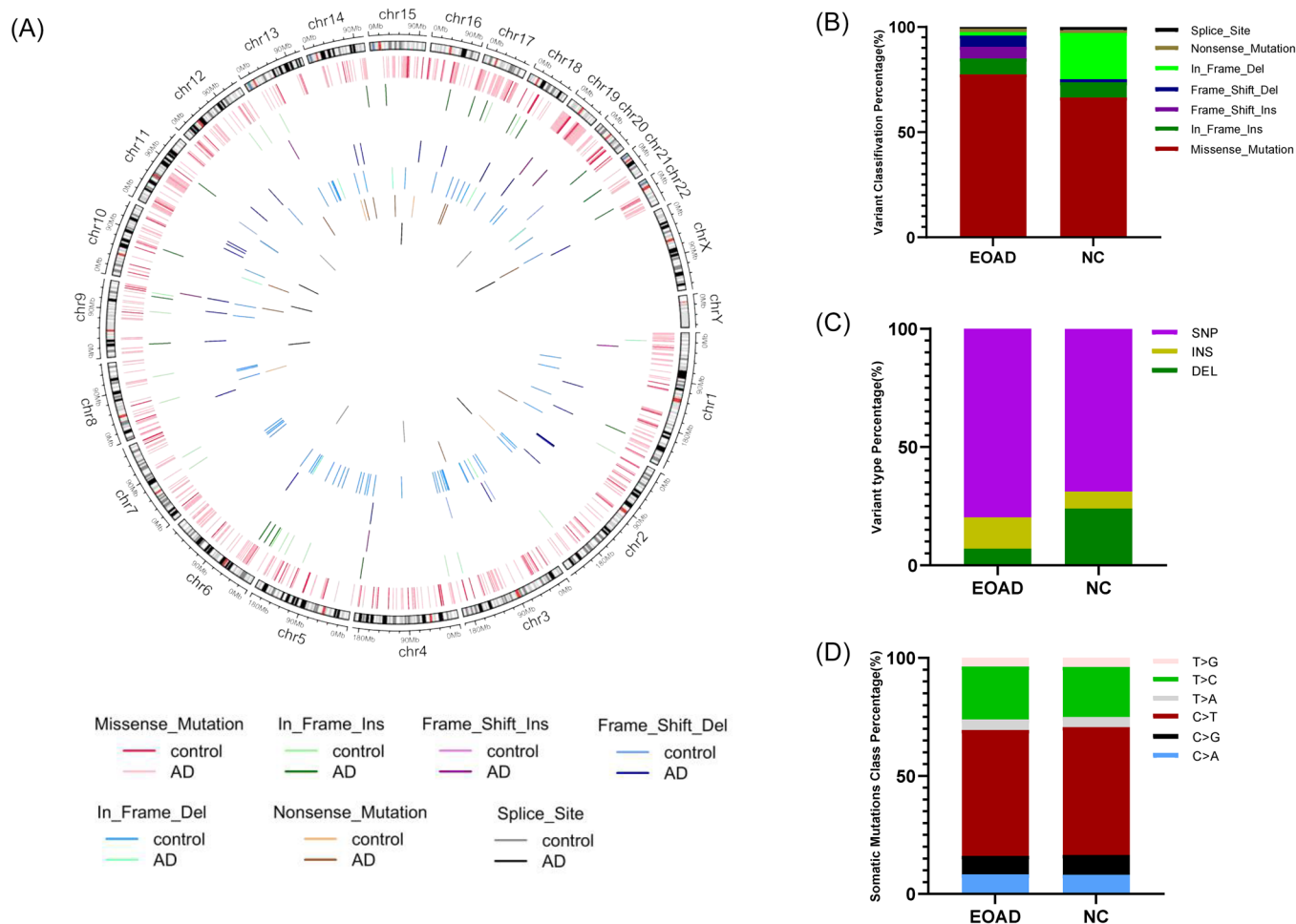
FIGURE 2 Substitution spectrum for somatic mutations. The somatic variant classification distribution in the EOAD and control groups (A), the percentage of each variant classification (B), the percentage of variant type (C), and the percentage of each somatic mutations class (D) in EOAD cases and control individuals. AD, Alzheimer's disease; EOAD, early-onset Alzheimer's disease; NC, normal control

somatic mutations available for signature analysis. We then decomposed all possible combinations of mutation signatures using maximum likelihood estimation (MLE) and identified the best model. Analyses were conducted on 65 single base substitution (SBS) signatures (adjusted for human whole-exome trinucleotide frequencies) from the PCAWG database. For EOAD-specific somatic mutations, SBS signatures 5, 1, and 23 accounted for 61.4%, 25.0%, and 6.4% of all somatic mutations, respectively (Figure 3). However, in the control group, the top three SBS signatures were 5, 15, and 1, accounting for 48.1%, 21.1%, and 18.4%, respectively (Figure S2 in supporting information). SBS signature 5 has recently been found to cause an accumulation of somatic mutations via a universal genomic aging mechanism, although the underlying cause remains unknown.

3.3 | Mutation annotation results

Variants in mutated genes were selected based on the deleteriousness predictions reported by PredictSNP2, encompassing both coding and non-coding mutations. The affected genes were extracted for each sample, and their mutation frequencies were compared between individuals with EOAD and those without AD. A total of 24 AD-specific genes were frequently affected by deleterious somatic mutations. The top 10 genes identified were leukocyte receptor cluster member 1 (*LENG1*); mucin 16 (*MUC16*); MIR31 host gene (*MIR31HG*); tubulin beta 4B class IVb (*TUBB4B*); small nucleolar RNA host gene 14 (*SNHG14*); *APP*; major histocompatibility complex, class II, DR beta 1 (*HLA-DRB1*); G-protein subunit alpha i2 (*GNAI2*); *LOC100506885*, and titin (*TTN*; Figure 4). Notably, the *TTN* gene also appeared among the top 10 genes in the control group (Figure S3 in supporting information). Our analysis of shared gene annotation revealed significant differences between the EOAD and non-AD groups (Fisher test, Bonferroni-adjusted $P < 0.05$). *APP* and *HLA-DRB1* are well-known AD risk genes, while *LENG1*, *MIR31HG*, *TUBB4B*, *SNHG14*, and *GNAI2* are newly discovered somatic mutations in patients with EOAD.

The mutation burden for each sample was obtained by counting the number of mutations across the 24 genes. We compared the mutation burden between the EOAD and NC groups and found it to be significantly higher in the EOAD group than in the control group ($P < 0.001$; Figure S4 in supporting information).
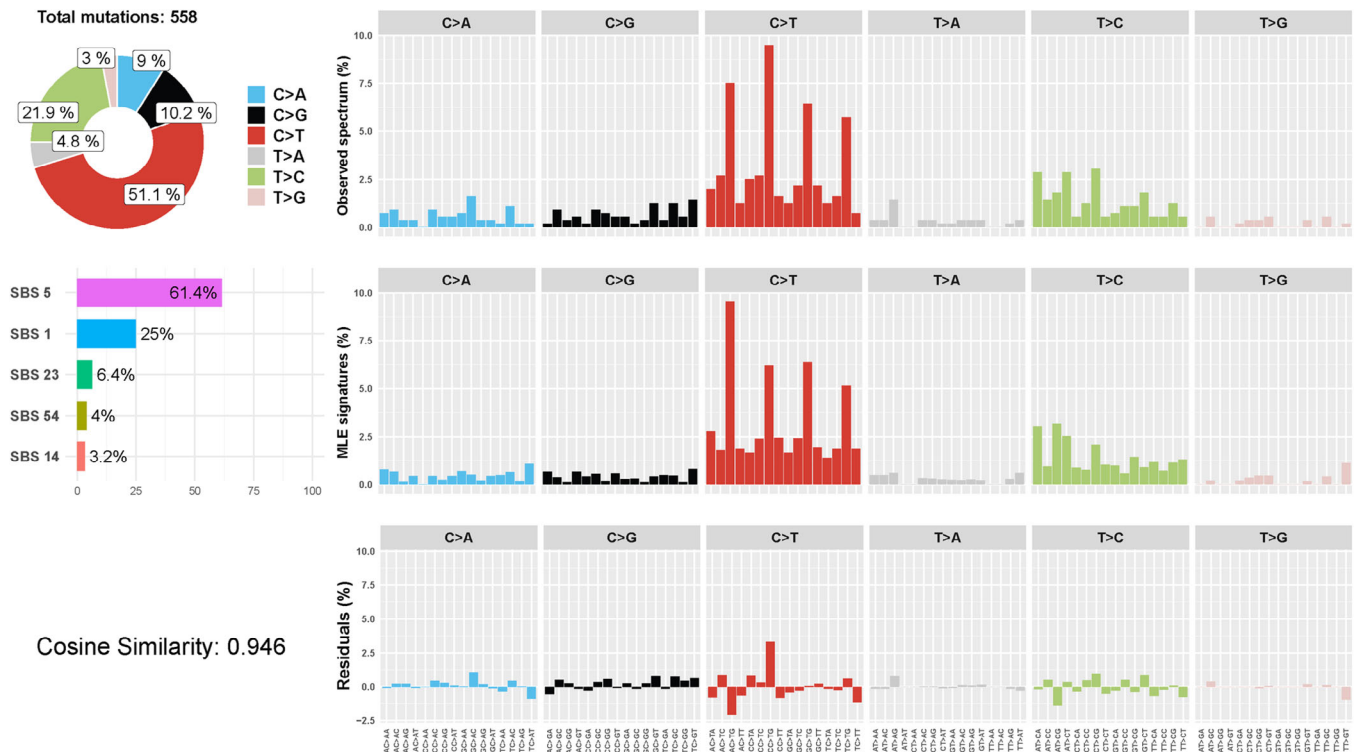
FIGURE 3 Mutation signatures of EOAD-specific somatic mutaions. The best decomposed mutation signature models from multiple likelihood estimation were derived for each tissue along with actual distribution of 96 possible mutation types. SBS signatures 5, 1, 23 and others account for 61.4%, 25.0%, 6.4%, and 7.2% of somatic mutations in EOAD, respectively. AD, Alzheimer's disease; EOAD, early-onset Alzheimer's disease; SBS, single base substitution
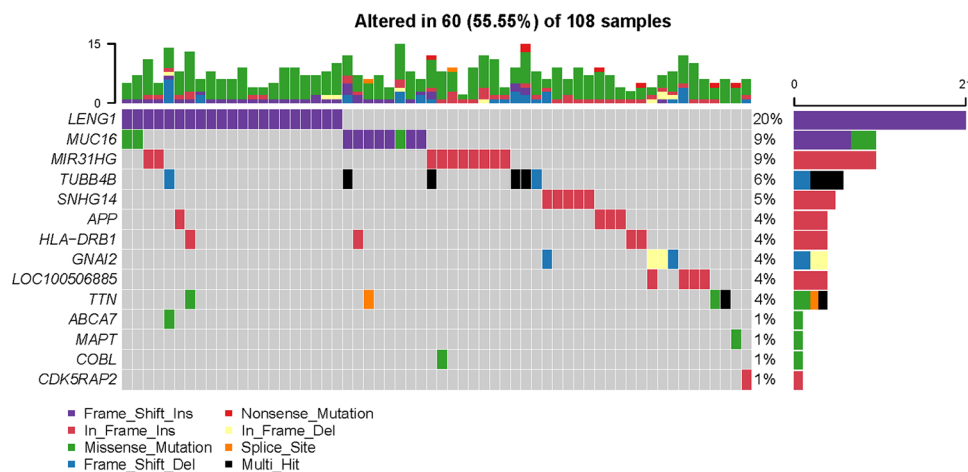

FIGURE 4 Landscape of somatic mutations contributing to early-onset Alzheimer's disease–specific genes

3.4 | Pathway analysis results

Gene set enrichment analysis showed that EOAD-specific somatic mutations were significantly enriched in multiple EOAD-related KEGG pathways, including motor proteins, cytoskeleton in muscle cells, gap junctions, pathogenic Escherichia coli infection, pathways of neurodegeneration–multiple diseases, amyotrophic lateral scle-

rosis, and Huntington disease (Figure 5A). These pathways were not observed in the non-AD population (Figure S5 in supporting information).

GO analysis showed that EOAD-specific somatic mutations were enriched for important BPs such as positive regulation of tumor necrosis factor production, necroptotic processes, microtubule bundle formation, astrocyte activation, negative regulation of amyloid beta

(A)

(B)
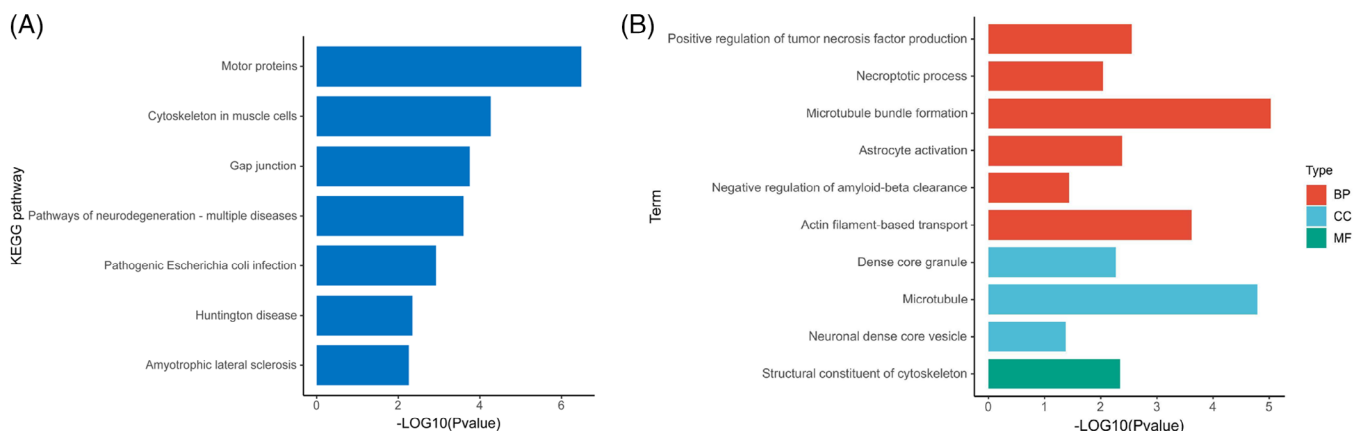
FIGURE 5 Functional analysis of EOAD–specific somatic genes. A, Gene-list enrichment test of putatively pathogenic somatic mutations using the KEGG. B, Results of GO enrichment analysis for AD-specific genes. AD, Alzheimer's disease; BP, biological process; CC, cellular component; EOAD, early-onset Alzheimer's disease; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; MF, molecular function.

(Aβ) clearance, actin filament-based transport, microtubules, dense core granules, neuronal dense core vesicles, and structural constituents of the cytoskeleton (Figure 5B).

3.5 | Drug-gene-KEGG correlation network in AD cases

We constructed drug–gene–KEGG correlation networks to explore how mutations affect the clinical and pathological traits of patients with AD at different omics levels by integrating the drugs, genes, and KEGG pathways. Drug—gene–KEGG correlation network analysis suggested interactions between *TUBB4B*, *HLA-DRB1*, *TUBB6*, *APP*, *MUC16*, *MAPT*, and drugs. Combretastatin A4, nocodazole, colchicine, podofilox, and paclitaxel were correlated with two new genes (*TUBB4B* and *TUBB6*; Figure S6 in supporting information). These drugs all work by targeting microtubules.

3.6 | Landscape of pathogenic germline mutations in AD

Furthermore, we examined germline mutations in EOAD risk factors. We found that several genes were associated with EOAD including potassium two pore domain channel subfamily K member 1 (*KCNK1*), kalirin RhoGEF kinase (*KALRN*), family with sequence similarity 13 member B (*FAM13B*), sarcospan (*SSPN*), NACC family member 2 (*NACC2*), neuronal cell adhesion molecule (*NRCAM*), and pleckstrin homology domain containing A7 (*PLEKHA7*; Figure 6A). Perhaps due to the small sample size, the risk genes did not reach a significance level of 5.0×10^{-8}. We also found that dedicator of cytokinesis 3 (*DOCK3*), proprotein convertase subtilisin/kexin type 5 (*PCSK5*), and phosphodiesterase 4D (*PDE4D*) were significantly associated with age at dementia onset ($P < 5.0 \times 10^{-8}$; Figure 6B).

3.7 | Gene expression results

The expression of the somatic and germline genes that we discovered (*LENG1*, *MIR31HG*, *TUBB4B*, *GNAI2*, *DOCK3*, *PCSK5*, *PDE4D*, *KALRN*, *SSPN*, *NRCAM*) was analyzed using data from the National Center for Biotechnology Information Gene Expression Omnibus dataset (http://www.ncbi.nlm.nih.gov/geo; GSE5281). The GSE5281 dataset includes expression data from 74 controls and 87 AD patients. Samples were obtained from individual brains provided by the Arizona Alzheimer's Disease Center, the Duke University Alzheimer's Disease Center, and the Washington University Alzheimer's Disease Center. Small sample sizes (200 µm sections) from six brain regions, each histopathologically or metabolically relevant to AD and aging, were collected. Prism software (version 8.0.0; GraphPad Software, Inc.) was used to compare gene expression between the CN and AD groups (unpaired *t* test and Welch *t* test) and to generate figures. As shown in Figure S7 in supporting information, except for *MIR31HG* and *GNAI2*, there were statistically significant differences in the expression of other genes between the AD and control groups. Specifically, *LENG1*, *PCSK5*, and *SSPN* showed higher expression levels in the brain tissue of patients with AD than in those of cognitively normal individuals. *TUBB4B*, *DOCK3*, *PDE4D*, *KALRN*, and *NRCAM* were significantly downregulated in patients with AD compared to cognitively normal individuals (Table S1 in supporting information).

3.8 | AlphaFold2 structure prediction

To predict the effect of mutations on protein structure, we used AlphaFold2. The protein structures of *GNAI2*, *LENG1*, and *TUBB4B* were significantly compromised by premature truncation, resulting in the loss of crucial functional domains and disruption of their native conformation, leading to impaired biological activity and potentially detrimental consequences (Figure S8 in supporting information). This
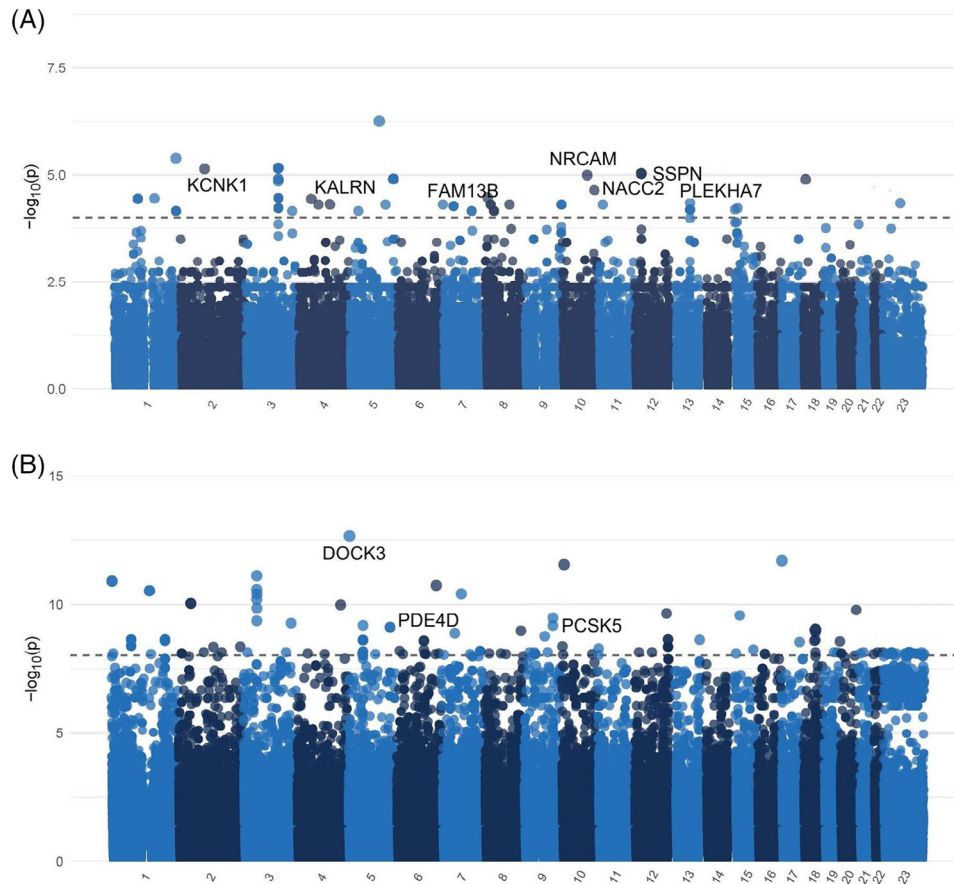
FIGURE 6 Manhattan plot of *P* values on the −log10 scale for SNPs with AD status and age of onset. A, *P* values with AD status. The dashed line represents $P = 1.0 \times 10^{-5}$. $\lambda = 1.002$. B, *P* values with age of onset. The dashed line represents $P = 5.0 \times 10^{-8}$. $\lambda = 1.056$. AD, Alzheimer's disease; SNP, single nucleotide polymorphism

truncation probably disrupts the ability of the protein to interact with other molecules, leading to a cascade of effects on various cellular processes. Furthermore, damaged proteins may accumulate and form toxic aggregates that exacerbate cellular dysfunction.

3.9 | Predictive model results

We randomly divided the participants into a discovery set (two thirds) and a testing set (one third). To assess the cumulative effect of multiple AD risk-associated SNPs in predicting AD risk, we calculated PRS for subjects in the independent testing set based on all 15 implicated SNPs in the discovery set (Table S2 in supporting information). The 15 SNPs were selected based on the criteria below: (1) exceeded the *P* value threshold and (2) when there were several SNPs in strong linkage disequilibrium satisfying these criteria (pairwise r2 ≥ 0.2), we chose the frequently cited one. The performance of the PRS in discriminating patients with AD from controls was evaluated using the area under the receiver operating characteristic curve (AUC). The results showed that the AUC of the PRS for discriminating AD cases from controls was 0.768 (95% confidence interval [CI]: 0.664–0.871; Figure 7A). When only seven SNPs were used, the AUC of the PRS was 0.729 (95% CI:

0.621–0.836; Figure 7B). The AUCs were increased to 0.795 (95% CI: 0.686–0.904) and 0.784 (95% CI: 0.672–0.897) after adding *APOE* ε4 to the model (Figure 7).

4 | DISCUSSION

In this study, we used WGS to identify somatic and germline mutations across the genome in the blood of Chinese individuals with EOAD. This analysis revealed unique genes and patterns that pave the way for future investigations and the identification of potential novel genetic markers affecting the disease.

Using WGS on Chinese patients with EOAD and controls, we found that the total number and mutation burden of blood somatic mutations in patients with EOAD were significantly higher than in controls. Consistent with our results, other studies have shown that carrying somatic mutations in EOAD loci may be associated with earlier disease onset.[20,35] Although the individuals with EOAD were younger than controls, mutation signature analyses showed that EOAD patients had a 61.4% prevalence of SBS5 (notable for C > T variants), which was much higher than the control group. SBS5 was recently shown to be generated via a universal genomic aging mechanism,[17] suggesting
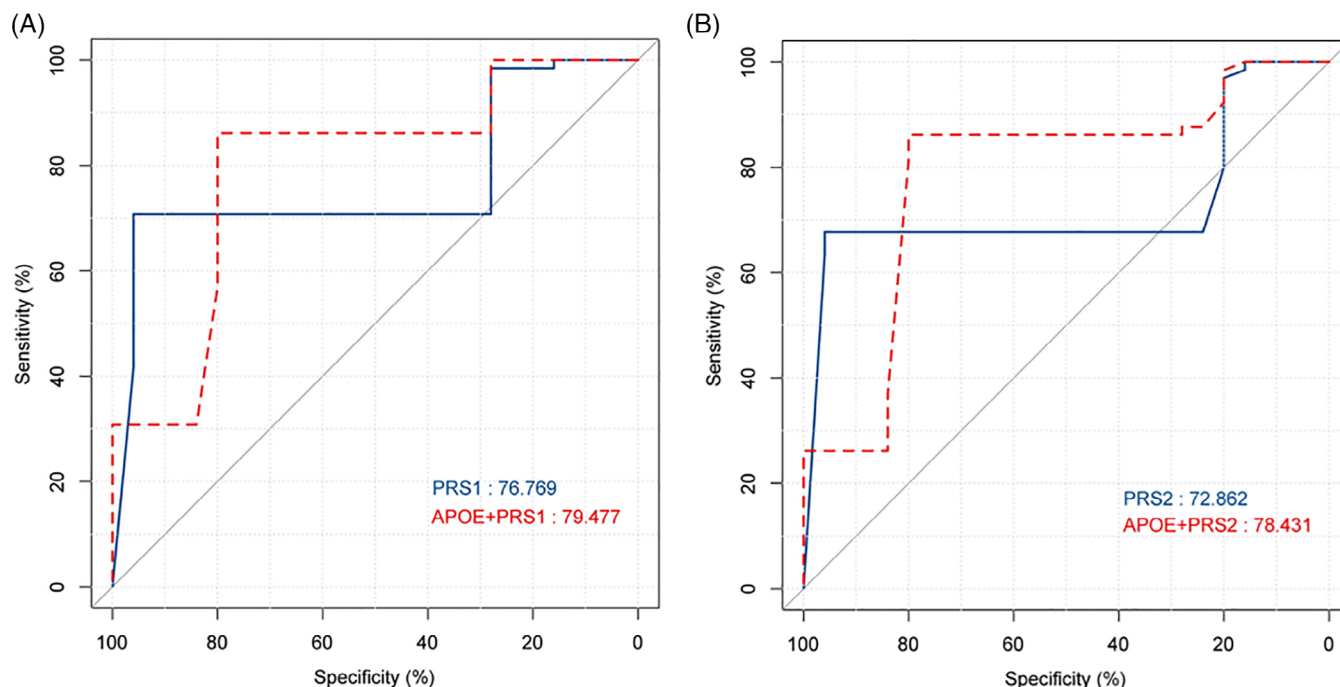
FIGURE 7 ROC curves for two predictive models with different predictors. A, PRS1 model included 15 SNPs. B, PRS2 model included 7 SNPs. ROC, receiver operating characteristic curve; SNP, single nucleotide polymorphism

that the accumulation of age-related somatic mutation signatures may be associated with earlier disease onset. We also identified an EOAD-specific mutation signature, SBS23, which accounts for 6.4% of the somatic variants in EOAD. However, the mechanisms underlying this signature remain unclear. Our findings provide important insights into signature-specific damage related to somatic mutations in EOAD.

We then identified the top 10 candidates for EOAD-specific somatic mutation-harboring genes, including *LENG1*, *MUC16*, *MIR31HG*, *TUBB4B*, *SNHG14*, *APP*, *HLA-DRB1*, *GNAI2*, *LOC100506885*, and *TTN*. We also identified EOAD-specific somatic mutations significantly enriched in motor proteins, gap junctions, and necroptotic pathways. *TUBB4B* is located in the microtubules and is predicted to be involved in the motor protein pathway. The findings revealed the upregulation of *TUBB4B* in Down syndrome and an AD mouse model[36] and the involvement of *TUBB4B* in cellular senescence.[37] Motor proteins, the main components of the axonal transport system, generate directed movements along the cytoskeletal tracks of axons, delivering cargo between the soma and the synapse.[38,39] Previous studies have reported that defects in motor protein–mediated neuronal transport mechanisms are involved in tau homeostasis and implicated in AD.[40,41] Bejarano et al.[42] found that reduced motor protein content in fibroblasts leads to insufficient autophagy and contributes to aging. The *GNAI2* encodes a protein involved in the gap junction pathway. Gap junctions are specialized transmembrane channels that facilitate neuroglial crosstalk in the CNS and play a crucial role in dementia.[43–45] During aging, *GNAI2* was overexpressed. Studies have revealed that gap junctions are important regulators of aging and premature senescence.[46]

Among the other top somatic mutation-harboring genes, long non-coding RNA SNHG14 is reported to be upregulated in AD serum

samples[47] and is involved in AD-related neuroinflammation and nerve cell apoptosis.[48,49] The long non-coding RNA MIR31HG regulates senescence-associated phenotypes.[50] Our results suggest that in addition to Aβ and tau pathways, the putatively specific somatic mutations identified in individuals with EOAD are known to involve aging and premature senescence and may contribute to increased risk for EOAD. However, mutations in these genes have not been reported in AD. For the first time, we report somatic mutations in these genes in Chinese patients with EOAD. This result suggests that Chinese patients with EOAD exhibit a unique mutational profile compared to patients from other ethnic backgrounds. In line with previous findings,[15,17,51] we also identified some candidate somatic variants among the known risk modifiers of AD, such as *APP*, *HLA-DRB1*, *ABCA7*, and *MAPT*.

Furthermore, we examined germline mutations in EOAD risk factors and found some genes associated with EOAD, including *KCNK1*, *KALRN*, *FAM13B*, *SSPN*, *NACC2*, *NRCAM*, and *PLEKHA7*. However, possibly owing to the small sample size, these risk genes did not reach a significance level of 5.0×10^{-8}. Among these, *KALRN* encodes a protein that interacts with huntingtin-associated protein 1, a huntingtin-binding protein that may function in vesicle trafficking. Dysregulation of the *KALRN* gene has been linked to various neurological disorders, including AD.[52] Reducing Kalrn in APPswe/PSEN1dE9 mice resulted in higher levels of synaptic proteins and resilience to the progression of psychosis-associated behaviors.[53] Interestingly, we found higher expression levels of *KALRN* in the brain tissues of patients with AD than in those of CN individuals. *NRCAM* encodes a neuronal cell adhesion molecule that promotes directional signaling during axonal cone growth. A previous study demonstrated that *NRCAM* is a marker for the

substrate-selective activation of ADAM10 in AD.[54] Cerebrospinal fluid *NRCAM* was correlated with the severity of cognitive impairment,[55] and can improve the diagnostic accuracy of Aβ42 and tau for AD.[56]

We also found that *DOCK3*, *PCSK5*, and *PDE4D* were significantly gene-wide associated with age of onset. *DOCK3* encodes a member of the DOCK (dedicator of cytokinesis) family of guanine nucleotide exchange factors. Its interaction with presenilin proteins as well as its ability to stimulate tau/MAPT phosphorylation suggests that it may be involved in AD. *DOCK3* mutations have been reported to be associated with neurodevelopmental disorders and impaired intellectual development.[57] *PCSK5* encodes a member of the subtilisin-like pro-protein convertase family.[58] A GWAS showed *PCSK5* mutations are linked to PHF-tau measurement,[59] neurofibrillary tangle measurement,[59] AD,[60] and cognitive impairment.[58] The *PDE4D* encoded protein has 3′,5′-cyclic adenosine monophosphate (cAMP) phosphodiesterase activity and degrades cAMP, which acts as a signal transduction molecule in multiple cell types. Paes et al.[61] found that increased PDE4D1 and -D3 isoform expression was associated with higher plaque and tau pathology levels, higher Braak stages, and progressive cognitive impairment. Inhibition of specific PDE4D isoforms can enhance memory processes.[62]

Using combinations of germline mutations identified in this study, we established predictive models for EOAD. A polygenic risk score can predict disease and allow the selection of patients with high polygenic risk scores for clinical trials and precision medicine.[63] Chaudhury et al.[64] reported a prediction accuracy of 75.5% for EOAD risk with certain predictors (*APOE*, polygenic risk score calculated from > 9000 SNPs, and sex). However, the model included too many SNPs, preventing its clinical use. Therefore, we established predictive models to determine the risk of EOAD development. Our 15-germline mutation-based model identified patients with EOAD and healthy controls with 77% accuracy. The application of this model is not affected by the sample type and does not require adjustments for other variables. Individualized risk scores can be obtained by SNP microarrays or DNA sequencing. By detecting the status of the 15 mutations in blood or tissue samples, clinicians can distinguish between EOAD and NCs. Therefore, the model has universal applicability and is almost unaffected by the technological differences among medical centers.

This study has some limitations. First, false-positive reads of somatic mutations may occur because of DNA damage during cell lysis and DNA polymerase errors during amplification, making them indistinguishable from naturally occurring somatic variants. Further validation and deeper coverage of larger cohorts are necessary to confirm the involvement of somatic mutations in AD. Second, because brain tissues were unavailable for sequencing, we detected somatic mutations by sequencing DNA isolated from the blood. Considering that the generation rate of somatic mutations in the blood may be five times higher than that in the brain, the interpretation of these mutations requires caution. Future investigations should explore the similarities between the somatic mutations identified in blood and brain tissues of the Chinese population. Third, larger scale studies are needed to better define the somatic and germline genetic landscapes in Chinese patients with

EOAD. Additionally, predictive models should be validated in larger populations. Fourth, functional studies on the candidate genes are lacking. Further investigations are warranted to understand the functional consequences of these risk genes in EOAD.

In this study, we delineated the genomic landscape of EOAD in a Chinese population by WGS of blood samples. We uncovered a distinctive mutation signature associated with EOAD and identified somatic mutations in genes related to cellular senescence, potentially shedding light on its etiology and offering new therapeutic targets. Furthermore, we identified novel germline mutations that predispose individuals to EOAD and developed a predictive model based on these 15 mutations, which holds promise for clinical applications.

AUTHOR CONTRIBUTIONS

ACKNOWLEDGMENTS

CONFLICT OF INTEREST STATEMENT

The authors report no competing interests. Author disclosures are available in the supporting information.

CONSENT STATEMENT

All participants provided informed consent.

ORCID

Jian-Ping Jia 🄳 https://orcid.org/0000-0003-4624-0336

REFERENCES

1. Li Y, Laws SM, Miles LA, et al. Genomics of Alzheimer's disease implicates the innate and adaptive immune systems. *Cell Mol Life Sci.* 2021;78:7397-7426.
2. Bagyinszky E, Youn YC, An SS, Kim S. Mutations, associated with early-onset Alzheimer's disease, discovered in Asian countries. *Clin Interv Aging.* 2016;11:1467-1488.
3. Lanoiselee HM, Nicolas G, Wallon D, et al. APP, PSEN1, and PSEN2 mutations in early-onset Alzheimer disease: a genetic screening study of familial and sporadic cases. *PLoS Med.* 2017;14:e1002270.

4. Griffin P, Apostolova L, Dickerson BC, et al. Developments in understanding early onset Alzheimer's disease. *Alzheimers Dement*. 2023;19 (Suppl 9):S126-S131.

5. Liu G, Sun JY, Xu M, Yang XY, Sun BL. SORL1 Variants show different association with early-onset and late-onset Alzheimer's disease risk. *J Alzheimers Dis*. 2017;58:1121-1128.

6. Liu X, Wu P, Shen L, et al. DHCR7 rs12785878 T>C polymorphism is associated with an increased risk of early onset of Alzheimer's disease in Chinese population. *Front Genet*. 2021;12:583695.

7. Slattery CF, Beck JA, Harper L, et al. R47H TREM2 variant increases risk of typical early-onset Alzheimer's disease but not of prion or frontotemporal dementia. *Alzheimers Dement*. 2014;10:602-608.

8. Cochran JN, Geier EG, Bonham LW, et al. Non-coding and loss-of-function coding variants in TET2 are associated with multiple neurodegenerative diseases. *Am J Hum Genet*. 2020;106:632-645.

9. Nho K, Horgusluoglu E, Kim S, et al. Integration of bioinformatics and imaging informatics for identifying rare PSEN1 variants in Alzheimer's disease. *BMC Med Genomics*. 2016;9 (Suppl 1):30.

10. Werling DM, Brand H, An J-Y, et al. An analytical framework for whole-genome sequence association studies and its implications for autism spectrum disorder. *Nat Genet*. 2018;50:727-736.

11. Luo R, Fan Y, Yang J, et al. A novel missense variant in ACAA1 contributes to early-onset Alzheimer's disease, impairs lysosomal function, and facilitates amyloid-beta pathology and cognitive decline. *Signal Transduct Target Ther*. 2021;6:325.

12. Li P, Yuanyuan Y, Cai H, Zhang H, Zhou Y. Case analysis of early-onset Alzheimer's disease associated with TBK1 p.Tyr235Phe gene mutation. *Front Neurol*. 2022;13:993399.

13. Cacace R, Heeman B, Van Mossevelde S, et al. Loss of DPP6 in neurodegenerative dementia: a genetic player in the dysfunction of neuronal excitability. *Acta Neuropathol*. 2019;137:901-918.

14. Prokopenko D, Morgan SL, Mullin K, et al. Whole-genome sequencing reveals new Alzheimer's disease-associated rare variants in loci related to synaptic function and neuronal development. *Alzheimers Dement*. 2021;17:1509-1527.

15. Nicolas G, Acuña-Hidalgo R, Keogh MJ, et al. Somatic variants in autosomal dominant genes are a rare cause of sporadic Alzheimer's disease. *Alzheimers Dement*. 2018;14:1632-1639.

16. Rodin RE, Dou Y, Kwon M, et al. The landscape of somatic mutation in cerebral cortex of autistic and neurotypical individuals revealed by ultra-deep whole-genome sequencing. *Nat Neurosci*. 2021;24:176-185.

17. Park JS, Lee J, Jung ES, et al. Brain somatic mutations observed in Alzheimer's disease associated with aging and dysregulation of tau phosphorylation. *Nat Commun*. 2019;10:3090.

18. Bushman DM, Kaeser GE, Siddoway B, et al. Genomic mosaicism with increased amyloid precursor protein (APP) gene copy number in single neurons from sporadic Alzheimer's disease brains. *Elife*. 2015;4:e05116.

19. Sala Frigerio C, Lau P, Troakes C, et al. On the identification of low allele frequency mosaic mutations in the brains of Alzheimer's disease patients. *Alzheimers Dement*. 2015;11:1265-1276.

20. Nicolas G, Acuna-Hidalgo R, Keogh MJ, et al. Somatic variants in autosomal dominant genes are a rare cause of sporadic Alzheimer's disease. *Alzheimers Dement*. 2018;14:1632-1639.

21. Spencer PS, Kisby G. Chemicals, somatic mutations and neurodegeneration: evidence from Western Pacific amyotrophic lateral sclerosis-parkinsonism-dementia complex (ALS-PDC): commentary on: Leija-Salazar M, Piette C, Proukakis C. Review: somatic mutations in neurodegeneration. *Neuropathol Appl Neurobiol*. 2019;45:525-527.

22. Wei W, Keogh MJ, Aryaman J, et al. Frequency and signature of somatic variants in 1461 human brain exomes. *Genet Med*. 2019;21:904-912.

23. Park J-H, Park I, Youm EM, et al. Novel Alzheimer's disease risk variants identified based on whole-genome sequencing of APOE epsilon4 carriers. *Transl Psychiatry*. 2021;11:296.

24. McKhann GM, Knopman DS, Chertkow H, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the national institute on aging-Alzheimer's association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement*. 2011;7:263-269.

25. McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA work group under the auspices of department of health and human services task force on Alzheimer's disease. *Neurology*. 1984;34:939-944.

26. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38:e164.

27. Mayakonda A, Lin D-C, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res*. 2018;28:1747-1756.

28. Fu Y, Liu Z, Lou S, et al. FunSeq2: a framework for prioritizing noncoding regulatory variants in cancer. *Genome biol*. 2014;15:480.

29. Bendl J, Musil M, Štourač J, Zendulka J, Damborský J, Brezovský J. PredictSNP2: a unified platform for accurately evaluating SNP effects by exploiting the different characteristics of variants in distinct genomic regions. *PLoS Comput Biol*. 2016;12:e1004962.

30. Lee J, Lee AJ, Lee J-K, et al. Mutalisk: a web-based somatic MUTation AnaLyIS toolKit for genomic, transcriptional and epigenomic signatures. *Nucleic Acids Res*. 2018;46:W102-W108.

31. Wu T, Hu E, Xu S, et al. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innovation*. 2021;2:100141.

32. Bertram L, McQueen MB, Mullin K, Blacker D, Tanzi RE. Systematic meta-analyses of Alzheimer disease genetic association studies: the AlzGene database. *Nature genetics*. 2007;39:17-23.

33. Khanna A, Larson DE, Srivatsan SN, et al. Bam-readcount—rapid generation of basepair-resolution sequence metrics. *Journal of Open Source Software*. 2022;7:3722.

34. Brown DW, Myers TA, Machiela MJ. PCAmatchR: a flexible R package for optimal case-control matching using weighted principal components. *Bioinformatics*. 2021;37:1178-1181.

35. Keogh MJ, Wei W, Aryaman J, et al. High prevalence of focal and multi-focal somatic genetic variants in the human brain. *Nat Commun*. 2018;9:4257.

36. Kelley CM, Ginsberg SD, Alldred MJ, Strupp BJ, Mufson EJ. Maternal choline supplementation alters basal forebrain cholinergic neuron gene expression in the Ts65Dn mouse model of down syndrome. *Dev Neurobiol*. 2019;79:664-683.

37. Yang Z, Gao S, Wong CC, et al. TUBB4B is a novel therapeutic target in non-alcoholic fatty liver disease-associated hepatocellular carcinoma. *J Pathol*. 2023;260:71-83.

38. Vicario-Orri E, Opazo CM, Munoz FJ. The pathophysiology of axonal transport in Alzheimer's disease. *J Alzheimers Dis*. 2015;43:1097-1113.

39. Wang ZX, Tan L, Yu JT. Axonal transport defects in Alzheimer's disease. *Mol Neurobiol*. 2015;51:1309-1321.

40. Hares K, Miners JS, Cook AJ, et al. Overexpression of kinesin superfamily motor proteins in Alzheimer's disease. *J Alzheimers Dis*. 2017;60:1511-1524.

41. Selvarasu K, Singh AK, Iyaswamy A, et al. Reduction of kinesin I heavy chain decreases tau hyperphosphorylation, aggregation, and memory impairment in Alzheimer's disease and tauopathy models. *Front Mol Biosci*. 2022;9:1050768.

42. Bejarano E, Murray JW, Wang X, et al. Defective recruitment of motor proteins to autophagic compartments contributes to autophagic failure in aging. *Aging cell*. 2018;17:e12777.

43. Mesnil M, Defamie N, Naus C, Sarrouilhe D. Brain disorders and chemical pollutants: a gap junction link? *Biomolecules*. 2020;11:51.

44. Pechlivanidou M, Kousiappa I, Angeli S, et al. Glial gap junction pathology in the spinal cord of the 5xFAD mouse model of early-onset Alzheimer's disease. *Int J Mol Sci*. 2022;23:15597.

45. Huang X, Su Y, Wang N, et al. Astroglial connexins in neurodegenerative diseases. *Front Mol Neurosci*. 2021;14:657514.

46. Zeitz MJ, Smyth JW. Gap junctions and ageing. *Sub-cell Biochem*. 2023;102:113-137.

47. Huaying C, Xing J, Luya J, Linhui N, Di S, Xianjun D. A signature of five long non-coding RNAs for predicting the prognosis of Alzheimer's disease based on competing endogenous RNA networks. *Front Aging Neurosci*. 2020;12:598606.

48. Duan R, Wang SY, Wei B, et al. Angiotensin-(1-7) analogue AVE0991 modulates astrocyte-mediated neuroinflammation via lncRNA SNHG14/miR-223-3p/NLRP3 pathway and offers neuroprotection in a transgenic mouse model of Alzheimer's disease. *J Inflamm Res*. 2021;14:7007-7019.

49. Tan Q, Liu L, Wang S, Wang Q, Sun Y. Dexmedetomidine promoted HSPB8 expression via inhibiting the lncRNA SNHG14/UPF1 axis to inhibit apoptosis of nerve cells in AD : the role of dexmedetomidine in AD. *Neurotox Res*. 2023;41:471-480.

50. Montes M, Lubas M, Arendrup FS, et al. The long non-coding RNA MIR31HG regulates the senescence associated secretory phenotype. *Nat Commun*. 2021;12:2459.

51. Lee MH, Siddoway B, Kaeser GE, et al. Somatic APP gene recombination in Alzheimer's disease and normal neurons. *Nature*. 2018;563:639-645.

52. Parnell E, Shapiro LP, Voorn RA, et al. KALRN: a central regulator of synaptic function and synaptopathies. *Gene*. 2021;768:145306.

53. Krivinko JM, Erickson SL, Ding Y, et al. Synaptic proteome compensation and resilience to psychosis in Alzheimer's disease. *Am J Psychiatry*. 2018;175:999-1009.

54. Brummer T, Müller SA, Pan-Montojo F, et al. NrCAM is a marker for substrate-selective activation of ADAM10 in Alzheimer's disease. *EMBO Mol Med*. 2019;11:e9695.

55. Hu WT, Chen-Plotkin A, Arnold SE, et al. Novel CSF biomarkers for Alzheimer's disease and mild cognitive impairment. *Acta Neuropathol*. 2010;119:669-678.

56. Wildsmith KR, Schauer SP, Smith AM, et al. Identification of longitudinally dynamic biomarkers in Alzheimer's disease cerebrospinal fluid by targeted proteomics. *Mol Neurodegener*. 2014;9:22.

57. Iwata-Otsubo A, Ritter AL, Weckselbatt B, et al. DOCK3-related neurodevelopmental syndrome: biallelic intragenic deletion of DOCK3 in a boy with developmental delay and hypotonia. *Am J Med Genet Part A*. 2018;176:241-245.

58. Chen Y-C, Liu Y-L, Tsai S-J, Kuo P-H, Huang S-S, Lee Y-S. LRRTM4 and PCSK5 genetic polymorphisms as markers for cognitive impairment in a hypotensive aging population: a genome-wide association study in Taiwan. *J Clin Med*. 2019;8:1124.

59. Wang H, Yang J, Schneider JA, Jager PLD, Bennett DA, Zhang H-Y. Genome-wide interaction analysis of pathological hallmarks in Alzheimer's disease. *Neurobiol Aging*. 2020;93:61-68.

60. Furney SJ, Simmons A, Breen G, et al. Genome-wide association with MRI atrophy measures as a quantitative trait locus for Alzheimer's disease. *Mol Psychiatry*. 2011;16:1130-1138.

61. Paes D, Lardenoije R, Carollo RM, et al. Increased isoform-specific phosphodiesterase 4D expression is associated with pathology and cognitive impairment in Alzheimer's disease. *Neurobiol Aging*. 2021;97:56-64.

62. Paes D, Schepers M, Willems E, et al. Ablation of specific long PDE4D isoforms increases neurite elongation and conveys protection against amyloid-β pathology. *Cell Mol Life Sci: CMLS*. 2023;80:178.

63. Sims R, Hill M, Williams J. The multiplex model of the genetics of Alzheimer's disease. *Nat Neurosci*. 2020;23:311-322.

64. Chaudhury S, Patel T, Barber IS, et al. Polygenic risk score in postmortem diagnosed sporadic early-onset Alzheimer's disease. *Neurobiol Aging*. 2018;62:244.e1–244.e8.

SUPPORTING INFORMATION

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