




ARTICLE



Functional genomics identify causal variant underlying the protective *CTSH* locus for Alzheimer's disease

Yu Li^{1,2,7}, Min Xu^{1,2,7}, Bo-Lin Xiang^{1,2}, Xiao Li^{1,2}, Deng-Feng Zhang^{1,2} , Hui Zhao^{3,4,5}, Rui Bi^{1,2,6}  and Yong-Gang Yao^{1,2,3,6} 

© The Author(s), under exclusive licence to American College of Neuropsychopharmacology 2023

Alzheimer's disease (AD) is the most prevalent age-related neurodegenerative disease, which has a high heritability of up to 79%. Exploring the genetic basis is essential for understanding the pathogenic mechanisms underlying AD development. Recent genome-wide association studies (GWASs) reported an AD-associated signal in the *Cathepsin H* (*CTSH*) gene in European populations. However, the exact functional/causal variant(s), and the genetic regulating mechanism of *CTSH* in AD remain to be determined. In this study, we carried out a comprehensive study to characterize the role of *CTSH* variants in the pathogenesis of AD. We identified rs2289702 in *CTSH* as the most significant functional variant that is associated with a protective effect against AD. The genetic association between rs2289702 and AD was validated in independent cohorts of the Han Chinese population. The *CTSH* mRNA expression level was significantly increased in AD patients and AD animal models, and the protective allele T of rs2289702 was associated with a decreased expression level of *CTSH* through the disruption of the binding affinity of transcription factors. Human microglia cells with *CTSH* knockout showed a significantly increased phagocytosis of A β peptides. Our study identified *CTSH* as being involved in AD genetic susceptibility and uncovered the genetic regulating mechanism of *CTSH* in pathogenesis of AD.

Neuropsychopharmacology (2023) 48:1555–1566; <https://doi.org/10.1038/s41386-023-01542-2>

INTRODUCTION

Alzheimer's disease (AD) is the most prevalent neurodegenerative disease and is characterized by the accumulation of extracellular amyloid plaques, intracellular neurofibrillary tangles, progressive neuronal loss and neuroinflammation [1, 2]. The molecular mechanisms of AD pathogenesis remain ambiguous and this has obstructed the development of valid treatment strategies [3]. Increasing evidence suggests that genetic factors play a critical role in the pathogenesis of AD [4–6]. The heritability of AD is estimated to be around 60% ~ 80% [7]. More than 100 mutations in three causal genes (*APP*, *PSEN1* and *PSEN2* [4, 5]) have been reported in patients with familial AD, but this only accounts for about 5% or even fewer of total AD cases. For the remaining 95% sporadic cases, high-throughput genome-wide association studies (GWASs), whole-exome sequencing and whole-genome sequencing efforts have identified a number of AD susceptibility genes [8–11]. The association of some GWAS risk genes with AD has been validated in Chinese populations in our previous studies (including *APOE*, *CLU*, *PICALM*, *BIN1* [12]) and others [13, 14]. There are several reports for causal genes including *C7* [15], *ACAA1* [16], and *ECE2* [17] in Han Chinese families with AD. Even though the list of AD risk genes keeps on rising, these AD-related genes can only explain a small portion of the total phenotypic variance, indicating a substantial missing heritability [5, 18, 19]. Moreover,

most of the GWAS risk variants are located in non-coding regions, and their complex regulatory role has been attracting considerable attention recently [20–25]. It is a daunting task to identify the causal genes and characterize the function of their risk variants in the post-GWAS era [26–30].

Cathepsin H (*CTSH*) is one of the cathepsin family members [31, 32]. The most large-scale meta-analysis of AD identified *CTSH* as a new AD-associated gene [8]. Proteome-wide association studies (PWASs) of AD found that variants rs12148472 and rs34593439 in the *CTSH* gene were associated with the *CTSH* protein abundance in human brain tissues [9, 33, 34]. These studies have indicated that the *CTSH* gene is genetically associated with AD. However, the following issues remain to be explored: (1) The available reports suggesting *CTSH* as an AD-associated gene are mainly based on European populations or populations of European origin [8], and whether the association between the *CTSH* gene and AD can be validated in other populations, including the Han Chinese, is unclear. (2) There are multiple AD-associated variants in the *CTSH* gene, and which are causal/functional variants has not been investigated. (3) The exact mechanisms regarding how the *CTSH* variant(s) affect the AD genetic risk remain to be elucidated.

In this study, we aimed to answer the above three questions. We performed a multiple-layer analysis that integrates genetic analysis, functional genomic assays, and cellular assays, to identify

¹Key Laboratory of Animal Models and Human Disease Mechanisms of the Chinese Academy of Sciences & Yunnan Province, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650204 Yunnan, China. ²Kunming College of Life Science, University of Chinese Academy of Sciences, Kunming 650204 Yunnan, China. ³KIZ/CUHK Joint Laboratory of Bioresources and Molecular Research in Common Diseases, Kunming Institute of Zoology, Chinese Academy of Sciences, Yunnan 650204 Kunming, China. ⁴Key Laboratory for Regenerative Medicine, Ministry of Education, School of Biomedical Sciences, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong SAR, China. ⁵Hong Kong Branch of CAS Center for Excellence in Animal Evolution and Genetics, The Chinese University of Hong Kong, Hong Kong SAR, China. ⁶Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences, Shanghai 200031, China. ⁷These authors contributed equally: Yu Li, Min Xu. ✉email: birui@mail.kiz.ac.cn; yaoyg@mail.kiz.ac.cn

the causal variant(s) in the *CTSH* gene that would account for its effect on AD susceptibility. We found the rs2289702 allele T in *CTSH* has a protective effect against AD risk in Han Chinese. Further analyses showed this allele affects the transcriptional factor binding ability of the *CTSH* gene and leads to a down-regulation of *CTSH* expression, which in turn contributes to an increase in the phagocytosis of A β peptides.

MATERIALS AND METHODS

Datasets and independent genetic validation in Han Chinese populations

To investigate whether the *CTSH* gene is genetically associated with AD in different populations, we retrieved data from two large-scale meta-analysis studies of AD GWASs [8, 35]. Data from Jansen et al. [35] contained GWAS data of 71,880 cases and 383,378 controls. Data from Bellenguez et al. [8] included stage I GWAS data of 85,934 clinically diagnosed or proxy (individuals whose parent(s) had dementia) AD cases and 401,577 controls. Single nucleotide polymorphisms (SNPs) located in the -10 kb \sim $+10$ kb region of *CTSH* were retrieved and subjected to subsequent analyses. Regional association plots for GWAS results were generated using the Locus Zoom (<http://locuszoom.org/>) [36]. Genotype data of 503 European individuals (EUR) and 504 individuals from East Asia (EAS) from the 1000 Genomes project (Phase 3) [37] were used to check the linkage disequilibrium (LD) among variants. The haploview v4.1 software (<https://www.broadinstitute.org/haploview/haploview>) [38] was used for LD visualization. The colocalization analysis of GWAS and expression quantitative trait loci (eQTL) signals at the *CTSH* locus was performed using the R package coloc [39] with default settings. GWAS data from Jansen et al. [35] and Bellenguez et al. [8] were integrated with GTEx eQTL of the frontal cortex, cortex, or cerebellar hemisphere [40], respectively.

We validated the association between rs2289702 and AD risk in three independent AD patient cohorts of Han Chinese origin. The first cohort included 221 AD patients with an early age at onset (age \leq 55) and/or an AD familial history (FOAD), in which 169 individuals had been reported in our previous studies [15, 41]. The other two cohorts included 1232 sporadic late-onset AD (LOAD) patients (age of onset $>$ 65) that have been described in our previous studies (615 AD patients from East Han Chinese; 617 AD patients in South Han Chinese) [12, 14, 15]. Sample collection was complied with the declaration of Helsinki, with informed consent being obtained from all the subjects. This study was approved by the institutional review board of the Kunming Institute of Zoology, Chinese Academy of Sciences. Healthy controls from 160 in-house non-dementia individuals [42] and 103 Han Chinese in Beijing and 105 Southern Han Chinese from the 1000 Genome Project (phase 3) [37] were combined as the initial population control group ($N=368$), as described in our previous study [15]. Genotype data of 6051 East Asia individuals from the gnomAD (<https://gnomad.broadinstitute.org/>) [43] were used as another population control group.

We sequenced a small fragment (322 bp) spanning the genomic region containing rs2289702. This fragment was amplified and sequenced using primer pair rs2289702-F 5'-TGGCCTCTACCGGGAAAGCTC-3' / rs2289702-R 5'-CGCCGCTCCTCACGCTCGT-3' (Supplementary Table S1), following the sequencing procedure in our previous study [15]. Frequencies of the rs2289702 allele T in AD patients and controls were compared by using the Fisher's exact test. Standard-error weighted meta-analysis of rs2289702 in combined Han Chinese in this study and European GWAS datasets by Jansen et al. [35] or Bellenguez et al. [8] was performed by using meta (<https://genome.sph.umich.edu/wiki/METAL>) [44].

Functional annotation of genetic variants in the *CTSH* gene

Multi-omics data from AD-related brain tissues or cells were combined to identify potentially functional variants, following the strategy and approach described in our recent study [45]. Briefly, we used chromatin immunoprecipitation followed by sequencing (ChIP-seq) data of histone modifications, assay for transposase-accessible chromatin with high-throughput sequencing (ATAC-seq) data, and ChIP-seq data of transcription factors (TFs) to prioritize variants within the genomic region of the *CTSH* gene that were located in regulatory elements and could affect TF binding. We used brain eQTL data and allele-specific expression (ASE) analyses to further test the regulatory role of identified variants.

For ChIP-seq data, we retrieved the ChIP-Seq data of histone modifications related to active promoters (H3K4me3, H3K9ac) or enhancers (H3K4me1, H3K27ac) from the ENCODE (<https://www.encodeproject.org/>) [46, 47] to

identify whether the tested SNPs of the *CTSH* gene are located in regulatory elements (i.e., promoters or enhancers). Peak files in bed format data of different AD-related brain tissues and cell lines were downloaded, including hippocampus, middle frontal cortex, inferior temporal cortex, neuronal progenitor cultured cells, astrocytes, peripheral blood mononuclear primary cells (monocytes), and brain microvascular endothelial cells. An $FDR < 0.001$ was applied to obtain relatively reliable promoter/enhancer peaks in the genomic region of *CTSH*. For ATAC-seq data analysis, open chromatin data of astrocytes, neurons, and monocytes were retrieved from public sources [48–50] to investigate whether the tested SNPs were located in the active transcription regions. The ATAC-seq peaks were downloaded, and an $FDR < 0.001$ was used to obtain relatively reliable open chromatin peaks.

To test whether different alleles of the tested SNPs affect TF binding, we firstly downloaded TF binding peaks for 623 TFs from the ENCODE [46, 47]. An $FDR < 0.001$ was applied for peak filtration. For each TF, DNA sequences of the top 1000 peaks (ranked by peak height) were subjected to TF binding motifs (position weight matrix, PWM) prediction, and the top three PWMs with the most significant E -values were used for the downstream analyses. The MEME software (<https://meme-suite.org/meme/>) [51] was used for the motif prediction. For these SNPs located within the binding motif of a particular TF, PWMs of the TF were subjected to the atSNP algorithm [52] to predict whether different alleles of the SNP had different binding affinities with this TF.

The association between genotype and *CTSH* mRNA expression level was analyzed by using the available brain eQTL database, the GTEx database (Genotype-Tissue Expression, <http://www.gtexportal.org/home/>) [40]. We reanalyzed the ASE data from the GTEx (phs000424.v7.p2) [40, 53, 54] to verify the *cis* regulatory effects of candidate variants on the *CTSH* gene. Only heterozygous SNPs in GTEx individuals were suitable and reserved for the ASE analyses. Binomial tests were used to evaluate if the ratio of the two alleles of each variant was significantly different from the expected 0.5 [54].

Quantification of *CTSH* mRNA expression in AD patients and murine models

Expression alterations of *CTSH* in microglia and astrocytes were investigated using single-cell RNA-seq (scRNA-seq) data of prefrontal cortex from 24 individuals with mild to severe AD pathology and 24 controls [55]. Alteration of the mRNA level of the *CTSH* gene in AD brains was investigated with the compiled data from the AlzData database (www.alzdata.org) [56], which integrated and re-normalized the original microarray data of 238 AD brain tissues and 232 control brain tissues from four brain regions (73 AD cases and 88 controls from the frontal cortex; 52 AD cases and 39 controls from the temporal cortex; 74 AD cases and 66 controls from the hippocampus; 39 AD cases and 39 controls from the entorhinal cortex).

We performed a correlation analysis between the mRNA level of the *Ctsh* gene and pathology level using AD mouse models [57]. In brief, the *Ctsh* gene expression data in AD transgenic and wild-type (WT) mice were retrieved from the Mouseac database (www.mouseac.org) [57]; this dataset contained the spatial-temporal mRNA expression data of 219 brain tissues from AD mouse models with familial AD mutations and 114 brain tissues of age-matched WT mice. The expression levels of *Ctsh* in the cortex and hippocampus tissues from mice at four different ages (2, 4, 8, and 18 months) were compared between AD HO_TASTPM mice (with homogenous APP^{K670N/M671L} and PSEN1^{M146V} mutations, $N=15$) and WT mice ($N=38$). The correlation between *Ctsh* mRNA level and pathology level of A β plaques was analyzed using data from 44 AD mice (HO_TASTPM mice, with homogenous APP^{K670N/M671L} and PSEN1^{M146V} mutations, $N=15$; HET_TASTPM mice, with heterogeneous APP^{K670N/M671L} and PSEN1^{M146V} mutations, $N=16$; TAS10 mice, with APP^{K670N/M671L} mutation, $N=13$). The correlation between *Ctsh* mRNA level and pathology level of tau was analyzed using data from 15 AD mice (with MAPT^{P301L} mutation, $N=15$).

Cell culture and creation of the *CTSH* knockout HM cell lines

HEK293T, U251 glioma cells, and human microglia (HM) cells were obtained from the Kunming Cell Bank, Kunming Institute of Zoology, Chinese Academy of Sciences. Briefly, the U251 cells were cultured in Roswell RPMI-1640 medium; the HEK293T cells and HM cells were cultured in Dulbecco's modified Eagle's medium (DMEM; Gibco-BRL, 11965-092). All culture medium were supplemented with 10% fetal bovine serum (FBS) (Gibco-BRL; 10099-141), 100 U/mL penicillin and 100 mg/mL streptomycin. Cells were maintained at 37°C in a humidified atmosphere incubator with 5% CO $_2$.

We established the *CTSH* knockout (KO) HM cell line by using the CRISPR/Cas9 system [58]. The sgRNA targeting sequence for the *CTSH* gene was designed by using the CRISPOR website (<http://crispor.tefor.net/>) [59]. Guide RNA *CTSH*-sgRNA-F 5'-GGCCGTAAGACCTACAGTACGG-3' / *CTSH*-sgRNA-R 5'-CCGTACTGTAGTCTTACGGCC-3' (Supplementary Table S1) were designed to target c.125–146 in the *CTSH* gene (with a GG added to the 5'-position of the 20-bp guide sequence to obtain efficient U6 transcription of the sgRNA and an overall efficiency of the CRISPR-Cas9 system) [58, 60]. The sgRNA sequences were cloned into the px330-mCherry vector (Addgene plasmid #98750) [61], which expressing SpCas9 and mCherry as described in our recent study [16]. The wild-type HM cells were transfected with the Cas9 construct targeting the *CTSH* gene by using Lipofectamine 3000 (Thermo Fisher, L300015) following the manufacturer's protocols. In brief, 3.75 μ L of Lipofectamine 3000, 2.5 μ g of plasmid DNA, and 5 μ L of P3000 (Thermo Fisher, L300015) were diluted in 250 μ L Opti-MEM to form Lipofectamine 3000-plasmid DNA-P3000 complex. After an incubation period of 15 min at room temperature, the transfection mixture was added to cells in a dropwise manner. Cells expressing mCherry were sorted by flow cytometry at 48 h after transfection. Single cells were then manually picked with a mouth pipette, and single cell clones were cultured for about 6 weeks. The genomic DNA of the single cell clones was extracted using AxyPrep Multisource Genomic DNA Miniprep Kit (Axygen, 26817KC1). The genomic region spanning the sgRNA targeting site was amplified by primer pair *CTSH*-ID-F 5'-CCTGAGAATCATAGCAGAAAGAA-3' / *CTSH*-ID-R 5'-ATGGGCTGGGTACTAACTG-3' (Supplementary Table S1) and directly sequenced to confirm the knockout of the *CTSH* gene.

Western blot

We followed the procedure described in our previous studies [16, 62, 63] to perform Western blotting for target proteins. In brief, cells were harvested by trypsinization, washed three times with cold PBS and lysed on ice in RIPA lysis buffer (Beyotime, P0013). After centrifugation at 12,000 \times g at 4 $^{\circ}$ C to remove cell debris, the protein concentration was determined by using the BCA protein assay kit (Beyotime, P0012) following the manufacturer's instructions. An equal amount of cellular protein (20 μ g) for each sample was electrophoresed by using 12% (vol/vol) SDS-PAGE and the protein in the gel transferred to a PVDF membrane (Bio-Rad, 1620177). The membranes were blocked in 5% nonfat dry milk dissolved in TBS (Cell Signaling Technology, 9997) containing 0.1% Tween 20 (TBST [0.1%, Sigma, P1379]) at room temperature for 2 h, followed by a wash with TBST, and incubated with primary antibodies of *CTSH* (Abcam, ab128907, 1:1000) and GAPDH (Enogene, E12-042-3, 1:10,000), respectively, overnight at 4 $^{\circ}$ C. After three washes with TBST (each 5 min), the membranes were incubated with secondary antibody (KPL, 1:10,000) for 1 h at room temperature. After three further washes, the membranes were visualized using the ECL reagents (MilliporeSigma, WBKLS0500).

Luciferase reporter assay

We cloned DNA fragments (613 bp) containing different alleles of rs2289702 into the pGL3-basic vector (Promega, E6661). DNA fragments containing different alleles of rs12148472 (813 bp) and rs12592898 (813 bp) were respectively cloned into the pGL3-promoter vector (Promega, E6661). HEK293T cells (2×10^4 /well), U251 cells (2×10^4 /well), or HM cells (2×10^4 /well) were seeded into 24-well plates for growth for 12 h and co-transfected with each luciferase reporter vector (500 ng) and pRL-TK vector (expressing renilla luciferase as an internal control, Promega, E2241; 50 ng) using Lipofectamine 3000 reagent (ThermoFisher, L300015). Cells were harvested at 48 h after transfection. Luciferase activity was determined by using the Dual-Luciferase Reporter Assay System (Promega, E1960) and Luminoskan Ascent instrument (Thermo Scientific). For luciferase assays in the presence of *ZFP69B* knockdown, we co-transfected HM cells with the related luciferase reporters, together with siRNA for *ZFP69B* (siRNA^{ZFP69B}; 50 nM) or control siRNA (siRNA^{NC}, 50 nM) for 48 h. Cells were then harvested for luciferase assay.

Quantitative real-time PCR (qRT-PCR)

The relative mRNA levels of the *CTSH* gene and the *ZFP69B* gene were quantified by qRT-PCR using the $2^{-\Delta\Delta CT}$ method, as described in our previous study [64]. In brief, mRNA levels of the *CTSH* gene and the *ZFP69B* gene were measured using primer pairs RT-*CTSH*-F/RT-*CTSH*-R and KD-*ZFP69B*-F/KD-*ZFP69B*-R (Supplementary Table S1), respectively, and was normalized to a housekeeping gene *ACTB* (Supplementary Table S1). A total of 80 ng of cDNA was subjected to qRT-PCR using iTaq Universal SYBR Green Supermix (172–5125; Bio-Rad Laboratories) with the above indicated

primer pairs on a CFX Connect Real-Time PCR Detection System (Bio-Rad Laboratories). Quantification of mRNA expression of cytokines including *IL8*, *TNF*, *IL1B*, *IL18* and *NLRP1* (Supplementary Table S1) in HM cells with or without treatment of fluorescently-labeled A β 42 in aggregated forms (1 μ g/mL) were also determined by using qRT-PCR.

Electrophoretic mobility shift assay (EMSA)

We performed an EMSA assay for rs2289702 following the manufacturer's protocol. In brief, the nuclear proteins of HEK293T cells and HM cells were extracted by using the nuclear protein extraction kit (Beyotime, P0028). The probes containing different rs2289702 alleles (Supplementary Table S1; EMSA-rs2289702-C-F/EMSA-rs2289702-C-R; EMSA-rs2289702-T-F/EMSA-rs2289702-T-R) were labeled by the EMSA probe biotin labeling kit (Beyotime, GS008). The probes (each 50 fmol) and nuclear extracts (2–8 μ g) were incubated at room temperature for 20 min to allow complexes to form. After electrophoresis and membrane transfer, the complexes were detected using the chemiluminescence EMSA kit (Beyotime, GS009).

RNA-seq and data analyses

We performed RNA-sequencing (RNA-seq) for WT and *CTSH* knockout HM cells, following the procedure described in our previous studies [15, 16]. Briefly, total RNA isolated from WT HM cells and 4 strains of HM cells with knockout of the *CTSH* gene (*CTSH*-KO1, *CTSH*-KO2, *CTSH*-KO3, and *CTSH*-KO4) was subjected to RNA-seq on the Illumina-HiSeq 4000 platform. Principal component analysis and differential expression analysis were performed using the DESeq2 [65] based on the log normalized counts. The *P* values were adjusted (*P*-adjust) by the Benjamini & Hochberg (BH) method. We defined differentially expressed genes (DEGs) between different conditions if *P*-adjust < 0.01. The consistently upregulated or downregulated genes in all 4 *CTSH* knockout cell strains were identified by VENNY analysis (<http://bioinfogp.cnb.csic.es/tools/venny/index.html>). Pathway enrichment analysis of the DEGs was performed using STRING (<https://string-db.org>) [66] to show the overall pattern of *CTSH* knockout on pathway enrichment. The gene set enrichment analyses (GSEA) [67] were performed using the R package clusterProfiler [65] with gene sets from the GOBP, KEGG, WikiPathways (WP), and the Reactome (REACTOME) pathway databases [68]. The RNA-seq data were deposited at GSA (<https://ngdc.cncb.ac.cn/gsa/>) under accession number HRA003434.

Assessment of the endocytosis activity of microglia cells

WT HM cells and *CTSH* knockout HM cells (2×10^5 /well) were seeded into 12-well plates and allowed to grow for 24 h. The plates were then treated with 1 μ g/mL of fluorescently-labeled A β 42 containing oligomeric, aggregated, and fibrillary forms (ChinaPeptides Co., Ltd.), respectively. We harvested cells at the indicated time points: (1) A β treatment for 1 h; (2) A β treatment for 3 h; (3) A β treatment for 3 h, followed by A β removal for 1 h; (4) A β treatment for 3 h, followed by A β removal for 3 h. Fluorescence intensity of cells was measured by using flow cytometry (BD, Influx, USA) at 535 nm. The FlowJo software was used to view and analyze the flow cytometric data. Cells with or without A β treatment for 3 h were harvested for quantification of mRNA levels of cytokines by using qRT-PCR (Supplementary Table S1).

Statistical analysis

The number of samples and AD mouse models was specified in the caption for each experiment. We used the *Pearson's* correlation test to test the correlation between *Ctsh* mRNA expression and A β pathology (scored by A β plaques) and between *Ctsh* mRNA level and pathology level of tau. We used two-way ANOVA with the Sidak's multiple comparisons and the Tukey's multiple comparisons to test significant differences of multiple groups, one-way ANOVA test with adjustment of Tukey's multiple comparisons to test significant differences of multiple comparisons. And we used two-tailed Student's *t* test to compare significant differences between two groups with PRISM software (GraphPad Software, Inc., La Jolla, CA, USA). A *P* < 0.05 was defined to be statistically significant (* *P* < 0.05, ** *P* < 0.01, *** *P* < 0.001, and **** *P* < 1.0×10^{-4}).

RESULTS

Multiple SNPs in *CTSH* confer genetic susceptibility to AD in European populations

In order to fine-map the genetic association of *CTSH* with AD, data from available large-scale meta-analyses of AD GWASs from

Jansen et al. [35] were retrieved and (re)analyzed. We found a total of 377 SNPs within the -10 kb \sim $+10$ kb region of the *CTSH* gene. Among these SNPs, 36 SNPs were associated with genetic risk of AD ($P < 0.05$; Supplementary Table S2). Considering the multiple testing correction ($0.05/377 = 1.33 \times 10^{-4}$), 9 SNPs (including rs2289702) had a P value $< 1 \times 10^{-4}$ and were regarded as showing a significant association with AD (Supplementary Fig. S1A). The association of these SNPs with AD was later validated when we retrieved and (re)analyzed another large-scale meta-analyses of AD GWASs from Bellenguez et al. [8] (Supplementary Fig. S1B, $P < 0.05$; Supplementary Table S3). Most of these AD-associated SNPs were linked with each other according to the LD analysis, especially in European populations (EUR) of the 1000 Genomes project (phase 3) [37] (Supplementary Fig. S1C, D). In order to investigate whether the *CTSH* gene is the only causal gene at this locus, we performed a colocalization analysis for *CTSH* and other genes. Results showed that *CTSH* is the only gene that had a strong evidence of colocalization at the *CTSH* locus (posterior probability of 0.83, 0.86, and 0.92 in GTEx frontal cortex, cortex, and cerebellar hemisphere [40], respectively) in brain tissues (Supplementary Table S5). Therefore, it is likely that the *CTSH* gene is the only causal gene at the *CTSH* locus.

Functional genomics analysis showing rs2289702 in *CTSH* as a functional variant mediating the genetic association between *CTSH* and AD

We performed functional annotations of SNPs in the *CTSH* gene to identify the exact functional variant among these AD-associated variants in the *CTSH* gene. Through integrating the available ChIP-Seq data and ATAC-seq data from AD-related brain tissues and cells [46–50], we found that among those AD-associated SNPs in the *CTSH* gene, only rs2289702 was located in a genomic region with active transcription (Fig. 1A). We used eQTL data from the GTEx project (GTEx, <http://www.gtexportal.org/home>) [40] to investigate whether these AD-associated SNPs were associated with mRNA expression of the *CTSH* gene. We found that the genotype TT of rs2289702 was prominently associated with a decreased mRNA expression level of *CTSH* in the cerebellar hemisphere, frontal cortex, and cortex tissues (Fig. 1B). Similar results could be observed for genotype CC of rs12148472 and genotype GG of rs12592898 (Fig. 1B).

Because the eQTL effect could be affected by the linkage among different variants, we performed a dual-luciferase reporter assay to identify which SNP was the functional variant that regulates gene expression. We found a significant difference between alleles C and T of rs2289702 on luciferase reporter activity (Fig. 1C), whereas no significant difference was observed between alleles T and C of rs12148472 (Fig. 1C) and between alleles A and G of rs12592898 (Fig. 1C). This result indicated that rs2289702 is the functional variant, and the association of *CTSH* mRNA expression with genotype CC of rs12148472 and genotype GG of rs12592898 (Fig. 1C) was most likely caused by the linkage with rs2289702. Taken together, the integrative analysis of the ChIP-seq and ATAC data (Fig. 1A) and the luciferase reporter assay (Fig. 1C) showed that rs2289702 is the main functional variant mediating the association between the *CTSH* gene and AD.

The rs2289702-T allele leads to a decreased binding affinity of transcription factors

We used the EMSA assay to explore the potential effect of the different rs2289702 alleles on the binding affinity of TFs. The EMSA results showed that variant rs2289702 could affect the binding affinity of TFs (Fig. 2A), and the rs2289702-T allele exhibited a significantly weaker binding of TFs than the rs2289702-C allele (Fig. 2A) in HEK293T cells. Similar results were observed in HM cells for rs2289702 alleles (Supplementary Fig. S2). An *in silico* prediction of TF binding motifs downloaded from ENCODE [46, 47] showed four TFs (ZFP69B, MAZ, ZFX, and HNRNPLL) that

could bind to the genomic region covering rs2289702 and their binding affinities were significantly affected by different alleles of rs2289702 ($P_{diff} < 0.05$, Fig. 2B). Notably, ZFP69B was the most significant TF to have its binding site disrupted by rs2289702 (Fig. 2B). Consistent with this prediction, knockdown of *ZFP69B* in HM cells (Supplementary Fig. S3) significantly reduced the luciferase reporter activity of allele C of rs2289702, but have no significant effect on allele T of rs2289702 (Fig. 2C), indicating that C > T change of rs2289702 significantly disrupted the binding affinity of ZFP69B.

To discern the potential effect of different alleles of rs2289702 on gene expression patterns, we analyzed the allele-specific expression (allelic imbalance) of this variant using data from various human tissues in the GTEx database [40, 53, 54]. Consistent with the above results, we found that the rs2289702-C allele, but not the T allele, exhibited a significantly preferential expression in human tissues, especially in brain tissues (Fig. 2D). We investigated the mRNA level of the *CTSH* gene in 293T cells with a heterozygous status of rs2289702 alleles. Compared to the 293T cells with a homozygous CC genotype, cells with a heterozygous CT genotype of rs2289702 showed a significantly reduced mRNA level of *CTSH* (Fig. 2E and Supplementary Fig. S4). These results indicated that rs2289702-T allele, which showed a protective effect on AD risk, could down-regulate the *CTSH* expression through the disruption of TF binding.

The association of rs2289702 with AD can be validated in Han Chinese populations

The above association analysis and functional genomics annotation provided convincing data for an active role of rs2289702 in conferring genetic risk to AD. We attempted to show whether an association between rs2289702 and AD exists in independent Han Chinese populations. We genotyped rs2289702 in three Han Chinese AD populations using Sanger sequencing. Consistent with the GWAS results in the reported European populations [8, 35], a significantly lower frequency of the rs2289702-T allele was observed in the FOAD patients ($N = 221$, frequency = 0.045, odds ratio (OR) [95% confidence interval (CI)] = 0.44 [0.25–0.74]), LOAD patients (Eastern Han, $N = 615$, frequency = 0.050, OR [95% CI] = 0.49 [0.34–0.71]; Southern Han, $N = 617$, frequency = 0.049, OR [95% CI] = 0.48 [0.33–0.69]), and combined AD patients ($N = 1453$, frequency = 0.049, OR [95% CI] = 0.48 [0.35–0.65]) compared with controls ($N = 368$, frequency = 0.098; Table 1). The association of rs2289702 with AD could be consistently observed when 6051 East Asian individuals from the gnomAD (<https://gnomad.broadinstitute.org/>) [43] were used as the population control (Table 1). Meta-analysis for rs2289702 with data from Han Chinese populations in this study and data from European populations [8, 35] further validated the association between rs2289702 and AD. These results confirmed the association of the *CTSH* gene with AD in both Han Chinese populations and European populations and suggested that this gene may have a key role in AD development.

Increased *CTSH* mRNA expression in brain tissues of AD patients and AD mouse models

To investigate whether the mRNA expression levels of the *CTSH* gene are changed in brain tissues of AD patients, we compared the *CTSH* mRNA expression levels in brain tissues of AD patients ($N = 238$) and controls ($N = 232$), using the data retrieved from AlzData (www.alzdata.org) [56]. We found a significantly increased mRNA expression level of *CTSH* in three brain regions of AD patients including the hippocampus, temporal cortex, and frontal cortex as compared to those of controls (Fig. 3A). We also analyzed the expression alterations of *CTSH* in microglia and astrocytes using single-cell RNA-seq (scRNA-seq) of prefrontal cortex from 24 individuals with mild to severe AD pathology and 24 controls [55]. Results showed that the *CTSH* expression was consistently

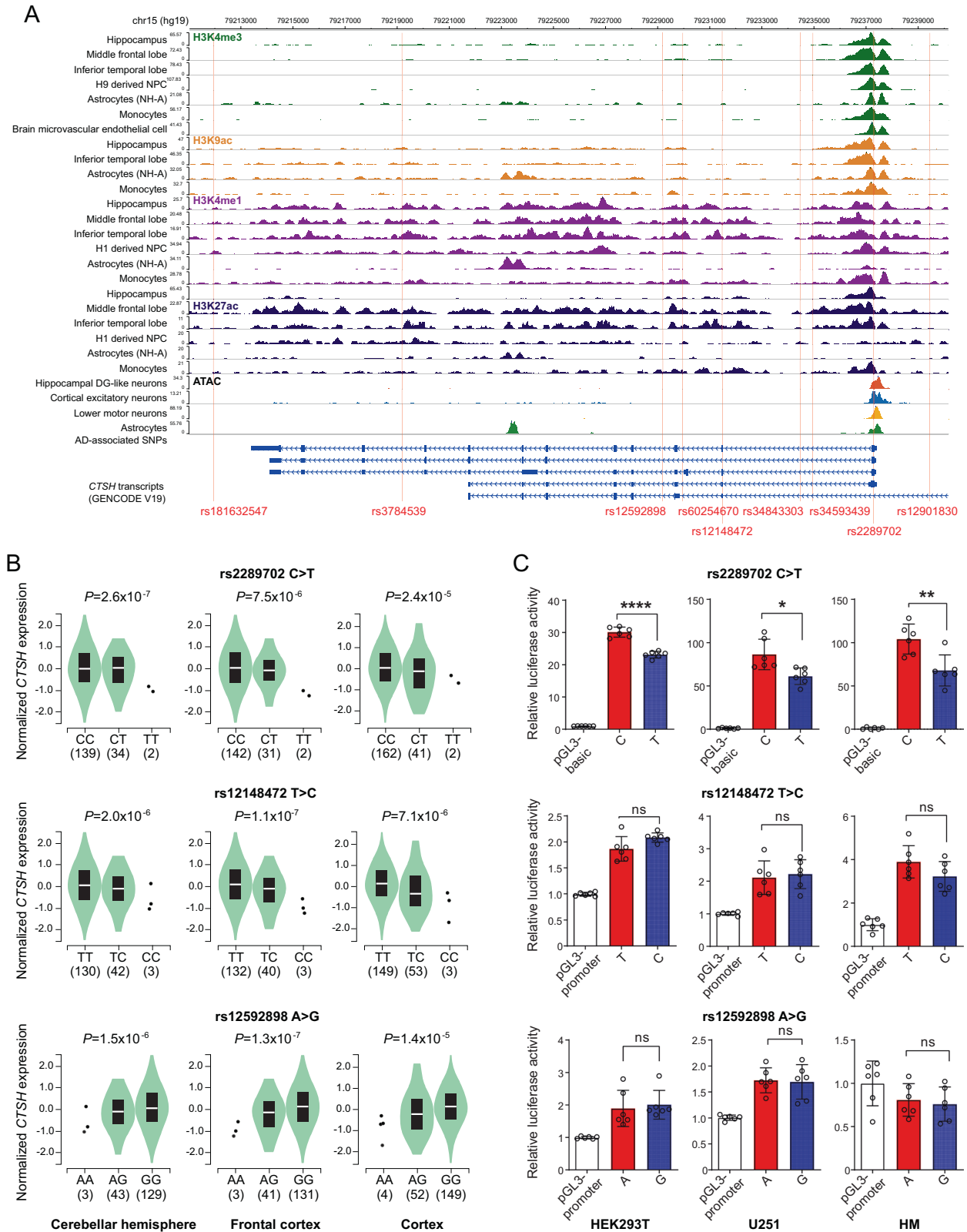


Fig. 1 Functional genomics analysis identifies rs2289702 as a functional variant in the CTSH gene. **A** Functional annotation of SNPs in the CTSH gene using ChIP-seq data and ATAC-seq data [46–50]. **B** Association of allele C of rs2289702 (upper), allele T of rs12148472 (middle), and allele G of rs12592898 (below) with an increased mRNA expression of CTSH in the cerebellar hemisphere, frontal cortex, and cortex. Data from the GTEx project [40] were reanalyzed by using a linear regression test. **C** Luciferase reporter assays showing a higher luciferase activity for the allele C (risk allele) than the allele T of rs2289702, whereas no significant difference was observed between alleles T and C of rs12148472 and between alleles A and G of rs12592898 in HEK293T cells, U251 cells, and HM cells. Data are presented as mean ± standard deviation (SD), and were analyzed by using two-tailed Student's *t* test. **P* < 0.05; ***P* < 0.01; *****P* < 1.0 × 10⁻⁴.

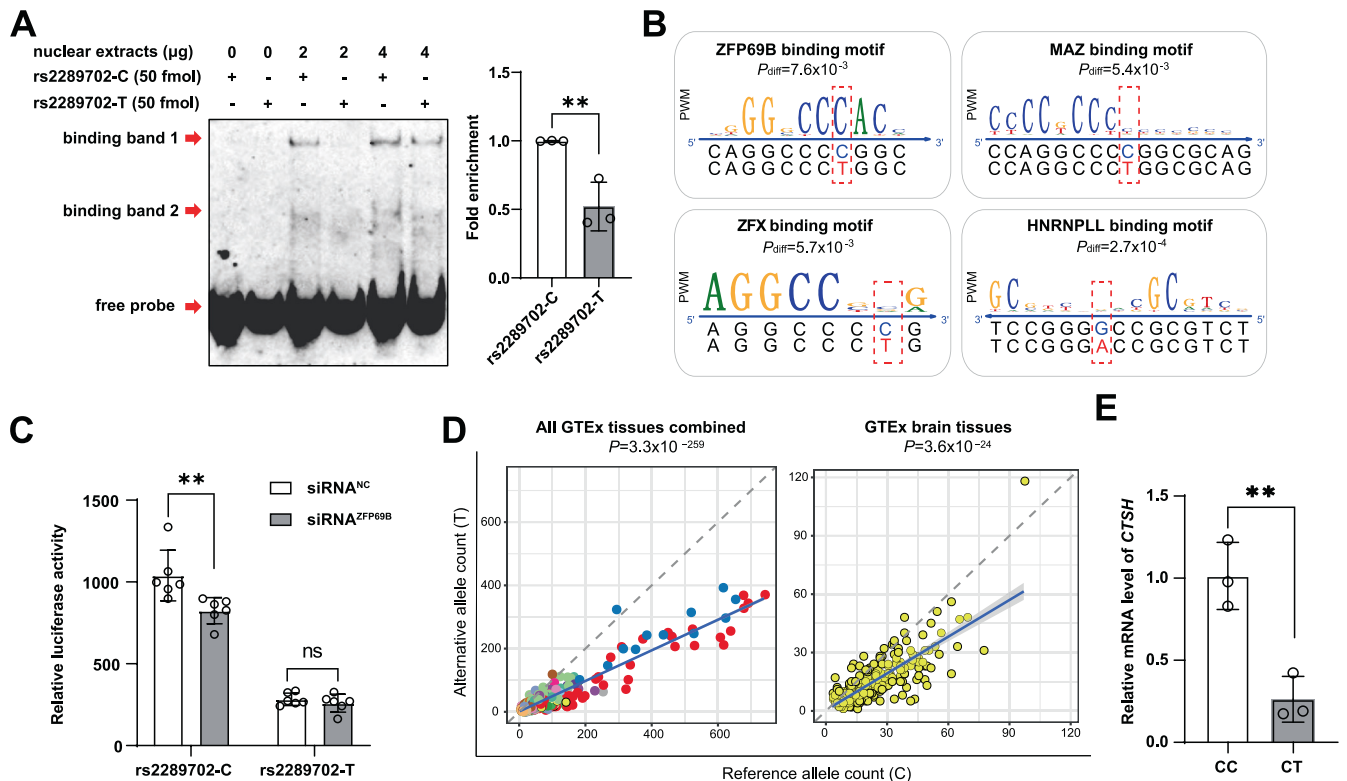


Fig. 2 SNP rs2289702 affects the binding affinity of transcription factors. **A** SNP rs2289702 affected transcription factor binding affinity. The EMSA assays showed that allele C had a stronger binding affinity than allele T of rs2289702 in HEK293T cells (left) and quantification of the binding affinity (right). The arrows indicated two binding bands and the free probe. **B** Program affiliated prediction showing the location of rs2289702 in the binding motifs of transcription factors ZFP69B, ZFX, MAZ and HNRNPLL. The P_{diff} value of the differential binding affinities of allele C and T of rs2289702 with each TF was computed by the atSNP algorithm [52]. **C** Knockdown of *ZFP69B* in human microglia (HM) cells affected the luciferase activity of the reporter vectors with different rs2289702 alleles. **D** Allele-specific expression of rs2289702 in GTEx human tissues [40]. The allele counts for the reference (Ref) allele and the alternative (Alt) allele were plotted, with each dot representing individual samples. Different tissues were marked by different colors. The blue line referred to the linear regression of all samples, and the gray shade represented the 95% confidence interval for the regression. The dashed line referred to the diagonal line. P value was measured by binomial tests. **E** Relative mRNA expression level of the *CTSH* gene in HEK293T cells with genotypes CC and CT of rs2289702. Values in **A** and **E** are presented as mean \pm SD. ns not significant; * $P < 0.05$; ** $P < 0.01$; two-tailed Student's t test for **A** and **E**; two-way ANOVA test adjusted by Sidak's multiple comparisons test for **C**.

Table 1. Association of rs2289702 with AD in independent Han Chinese populations and meta-analysis.

Samples	AD cases	Han Chinese controls ^d				GnomAD controls ^d				
		AC/AN	AC/AN	P	OR	95% CI	AC/AN	P	OR	95% CI
FOAD ^a	20/442	72/736		1.05×10^{-3}	0.44	0.25–0.74	789/12102	1.14×10^{-1}	0.68	0.41–1.07
Eastern Han with LOAD ^b	62/1230			9.02×10^{-5}	0.49	0.34–0.71		4.33×10^{-2}	0.76	0.57–0.99
Southern Han with LOAD ^b	61/1234			5.69×10^{-5}	0.48	0.33–0.69		3.20×10^{-2}	0.75	0.56–0.98
Combined ^c	143/2906			2.67×10^{-6}	0.48	0.35–0.65		1.13×10^{-3}	0.74	0.61–0.89
Meta-analysis (Jansen et al. ^e)	$N_{cases} = 73333$	$N_{ctl} = 383746$		7.48×10^{-6}	0.99	0.98–0.99	$N_{ctl} = 389429$	2.19×10^{-5}	0.99	0.98–0.99
Meta-analysis (Bellenguez et al. ^e)	$N_{cases} = 87387$	$N_{ctl} = 401945$		3.10×10^{-10}	0.92	0.90–0.94	$N_{ctl} = 407628$	7.99×10^{-8}	0.93	0.91–0.96

AC/AN allele count/total number of alleles, OR odds ratio, 95% CI 95% confidence interval, N_{cases} number of cases, N_{ctl} number of controls.

^aFOAD: 221 AD patients with an early age at onset (age ≤ 55) and/or an AD familial history, including 169 patients reported in our previous studies [15, 41].

^bEastern Han Chinese with late onset Alzheimer's disease (LOAD): 615 sporadic AD patients from East China; Southern Han Chinese with LOAD: 617 sporadic AD patients in South China [12, 14, 15].

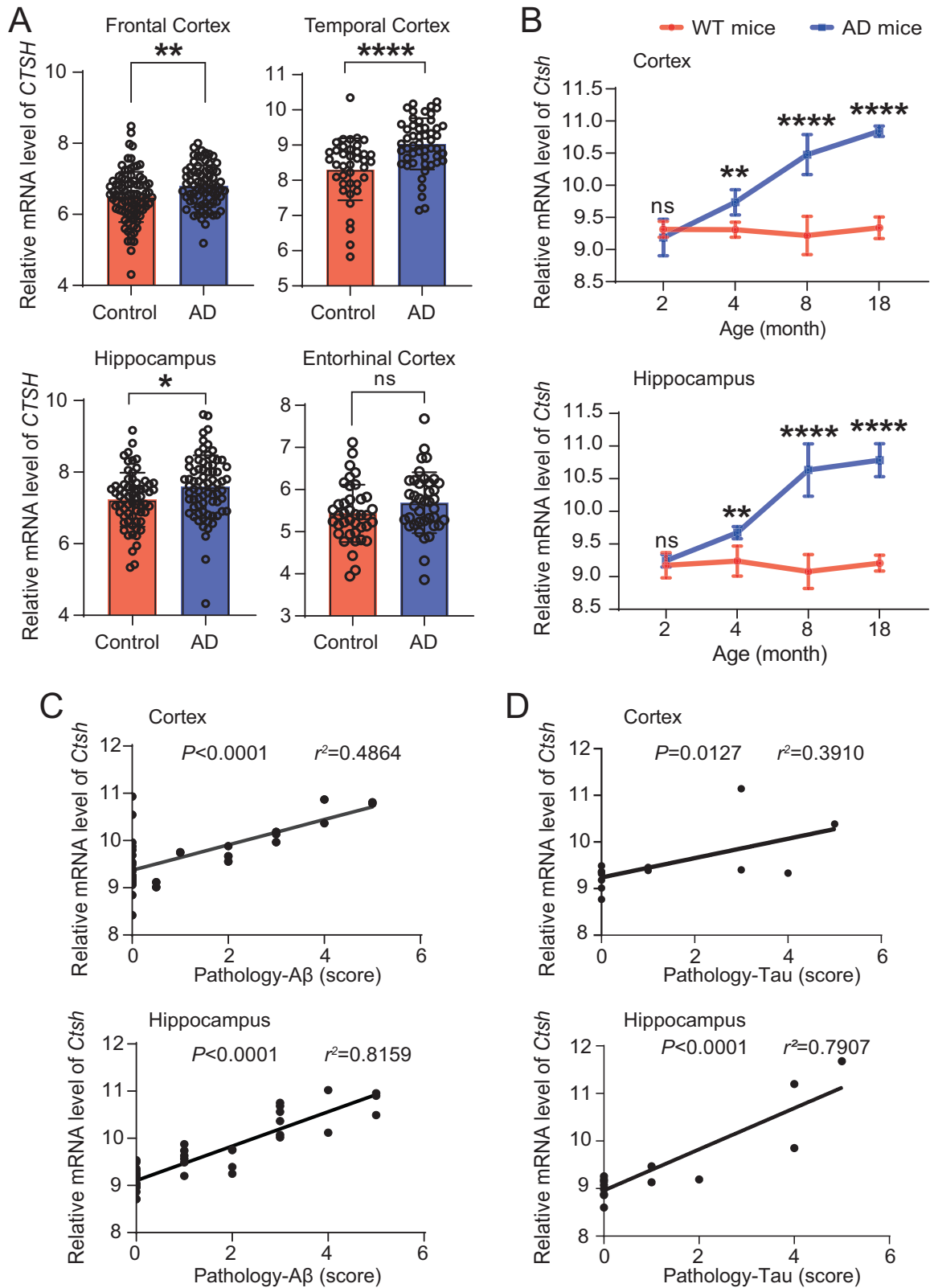
^cWe combined FOAD, Eastern Han with LOAD, and Southern Han with LOAD together as a population for comparison.

^dHan Chinese controls, 368 healthy controls as described in our previous study [15]; GnomAD controls, genotype data of 6051 East Asians from the gnomAD (<https://gnomad.broadinstitute.org/>) [43] as another population control. P values were calculated by Fisher's exact test.

^eMeta-analyses of rs2289702 in combined Han Chinese in this study and European GWAS dataset by Jansen et al. [35], and in combined Han Chinese in this study and European GWAS dataset by Bellenguez et al. [8]. P values were calculated by standard-error weighted meta-analysis.

upregulated in AD patients compared with controls in both microglia and astrocytes (Supplementary Table S6). Similarly, the mRNA expression levels of the *Ctsh* gene were significantly increased in the cortex and hippocampus tissues from AD mouse

models based on the data from the Mouseac (www.mouseac.org) [57] (Fig. 3B). Moreover, the mRNA expression level of the *Ctsh* gene was highly correlated with the pathological levels of A β and tau in the hippocampus and cortex tissues of AD mouse models



(Fig. 3C, D). Overall, we found that the mRNA expression level of the *CTSH* gene was upregulated in AD brain tissues and was positively correlated with the A β pathology in AD mouse models. As rs2289702 allele C was associated with increased gene expression, we suggested that an increased *CTSH* mRNA expression might be a key factor for AD development.

***CTSH* knockout enhances the microglial phagocytosis of A β peptides**

The *CTSH* gene was strongly expressed in glial cells in the brain according to the available single-cell sequencing data [55, 69, 70]. Microglia are the most abundant immune cell type in the human brain [71], and play important roles in the clearance of pathological

Fig. 3 Alterations of the *CTSH* mRNA level in the brain tissues of AD patients and AD mouse models. **A** The mRNA expression of the *CTSH* gene in the entorhinal cortex, frontal cortex, temporal cortex, and hippocampus tissues [56]. The difference between AD patients and controls was quantified by using two-tailed Student's *t* test. Values are presented as mean \pm SD. ns not significant; **P* < 0.05; ***P* < 0.01; *****P* < 1.0×10^{-4} . **B** The *Ctsh* mRNA expression in brain tissues of mice at different ages (2, 4, 8, and 18 months) from wild type mice (WT, *N* = 38) and HO_TASTPM mice (with homogenous APP^{K670N/M671L} and PSEN1^{M146V} mutations, *N* = 15). Values are presented as mean \pm SD and were quantified by using the two-way ANOVA with the Sidak's multiple comparisons test. ns, not significant; ***P* < 0.01; *****P* < 1.0×10^{-4} . Correlation between *Ctsh* mRNA expression and A β level (**C**) and between *Ctsh* mRNA expression and Tau pathology (**D**) in AD mouse models. The *Ctsh* gene expression data and A β and Tau pathology data of APP mutant mice (13 mice with homozygous APP^{K670N/M671L} mutation, 15 mice with homozygous APP^{K670N/M671L}-PSEN1^{M146V} mutations, 16 mice with heterozygous APP^{K670N/M671L}-PSEN1^{M146V} mutations) and 15 mice with MAPT^{P301L} mutation were retrieved from Mouseac [57] for the Pearson's correlation test.

proteins including A β and hyperphosphorylated tau [72, 73]. To understand the potential function of CTSH in AD, we established *CTSH* knockout HM cell lines and were able to obtain four cell strains (CTSH-KO1, CTSH-KO2, CTSH-KO3, and CTSH-KO4) with successful knockout of this gene, as confirmed by Western blotting (Fig. 4A) and Sanger sequencing (Supplementary Fig. S5). We performed RNA-seq for these KO cell strains. We found that differentially expressed genes (DEGs; *P*-adjust < 0.01) in the *CTSH* KO cells (Fig. 4B) were enriched in the vesicle related pathway (*P* = 1.99×10^{-6} , Fig. 4C), which is important for endocytosis and exocytosis functions [74–76]. Consistently, in GSEA analysis, 25 pathways were commonly enriched in all *CTSH* KO cell strains (*P* < 0.05, Supplementary Table S4). Except for pathways related with synaptic functions, metabolism process, and vesicle transportation, several immune-related pathways were also enriched. As shown in the Venn diagram, a total of 839 genes were consistently upregulated, and 927 genes were consistently downregulated in all four *CTSH* KO cell strains (Fig. 4B). Among these consistently regulated DEGs, we found that genes involved in positive regulation of endocytosis (e.g., *PICK1*, *HYAL2*) [77, 78] were significantly upregulated in *CTSH* KO cells, whereas genes involved in negative regulation of endocytosis (e.g., *CAV1*, *RAC1*, and *MTMR2*) [79–81] were significantly downregulated in *CTSH* KO cells (Fig. 4D).

Based on the above observation for altered expression for genes involved in endocytosis and vesicle-related pathway, we investigated whether *CTSH* knockout would affect the microglial phagocytosis of A β peptides. After the WT HM cells and *CTSH* KO HM cells were treated with different forms of fluorescently-labeled A β 42 (oligomeric, aggregated, and fibrillary), we found a significantly higher signal of fluorescently-labeled A β 42 in *CTSH* KO cells than in WT cells, especially for the oligomeric form of A β according to flow cytometry assay (Fig. 4E and Supplementary Fig. S6). After the removal of fluorescently-labeled A β 42, we observed a similar level of fluorescent signal between the WT and KO cells for a continual observation up to 3 h (Fig. 4F), indicating that *CTSH* KO would not affect the degradation of A β . This result suggested an increased phagocytic ability for microglia cells with *CTSH* KO and was consistent with a protective effect of reduced *CTSH* expression in AD brain tissues. Moreover, we found that the above short-term treatment of A β to HM cells did not induce remarkable mRNA expression of pro-inflammatory cytokines that might affect the phagocytic ability of HM cells (Supplementary Fig. S7). Collectively, these results indicated that *CTSH* played a critical role in the microglial phagocytosis of A β 42 and a reduced expression of this gene had a beneficial effect.

DISCUSSION

Up to now, GWASs and proteome-wide association studies (PWASs) have suggested the *CTSH* gene to be associated with AD [8, 9, 33–35], but the exact functional variant and mechanism are still to be identified. Moreover, whether the association of the *CTSH* gene with AD can be confirmed in other non-European populations is an open question. In this study, by analyzing the genetic data from whole-exome sequencing of the Han Chinese AD patients from our previous studies [15, 41], combined with the available data

from the large-scale AD GWASs of European populations or populations of European origin [8, 35], we found that genetic variants in the *CTSH* gene were significantly associated with AD risk, and this genetic association could be validated in independent populations of different origins (Table 1). Further functional genomics analyses showed that variant rs2289702 of the *CTSH* gene was the causal variant mediating the association between the *CTSH* gene and AD (Fig. 1), with a protective effect of rs2289702 allele T on AD risk that leads to a decreased level of TF binding and a decreased mRNA expression of the *CTSH* gene (Fig. 2). We also found that the level of *CTSH* expression was significantly lower in the brain tissues of healthy controls compared to those of AD patients (Fig. 3A). Moreover, *CTSH* knockout affected these genes related to endocytosis and significantly increased A β 42 phagocytosis in microglia cells (Fig. 4). All these lines of evidence provided robust evidence to indicate that rs2289702-T in the *CTSH* gene confers a protective effect on AD by downregulating the expression level of *CTSH*.

CTSH is one of the members of the cathepsin superfamily and is mainly located in the lysosome [82, 83]. The cathepsin superfamily is a large group of proteins that consists of 11 cysteine proteases (Cathepsin B, C, F, H, K, L, O, S, V, X, and W), two serine proteases (Cathepsin A and G), and two aspartic proteases (Cathepsin E and D) [84]. The main function of the cathepsin [85] is the degrading of proteins by proteolysis in the lysosome. Previous studies have revealed that cathepsin proteins play active roles in the pathogenesis of the AD [84, 86–89]. The elevated protein level or activity of Cathepsin B (CTSB) was reported to be associated with increased amyloid plaques, neuroinflammation in AD, and related neurodegenerative diseases [90, 91]. Recently, it has been found that the accumulation of high molecular weight kinogen in the AD brains may affect microglial function by altering phagocytosis and cathepsin (CTSS and CTSL) activity [92]. Our study found that increased *CTSH* expression is a risk factor for AD. This finding was consistent with a recent study that showed increased CTSE expression affected AD pathogenesis through upregulating neuroinflammation [93]. It is of note that the most significantly changed pathways affected by *CTSH* KO in HM cells are related to nervous system development, cell morphogenesis, metabolic process, cell migration, cell junction, and vesicle (Fig. 4C). This observation indicated that CTSH might play multiple roles in regulating the function of glia cells, which would eventually affect the function of neurons and AD pathogenesis. It would be worthwhile to explore whether the expression level of the *CTSH* gene was negatively associated with cognition/plasticity in the future study. The exact regulating mechanisms of CTSH in AD pathogenesis are deserved to be further explored.

Our study had several limitations. First, rs2289702 is located in the first exon of the *CTSH* gene and encodes a missense variant that changes glycine (Gly) to arginine (Arg) in some isoforms of *CTSH* (Fig. 1A), which is in line with a recent study showing that a missense variant in NDUFA6 would affect schizophrenia risk by affecting YY1 binding and NAGA expression [94]. Although this variant is not located in the key domains that affect the Cathepsin H activity, we did not know whether this variant would play other cellular functions as we did not perform focused assays. One

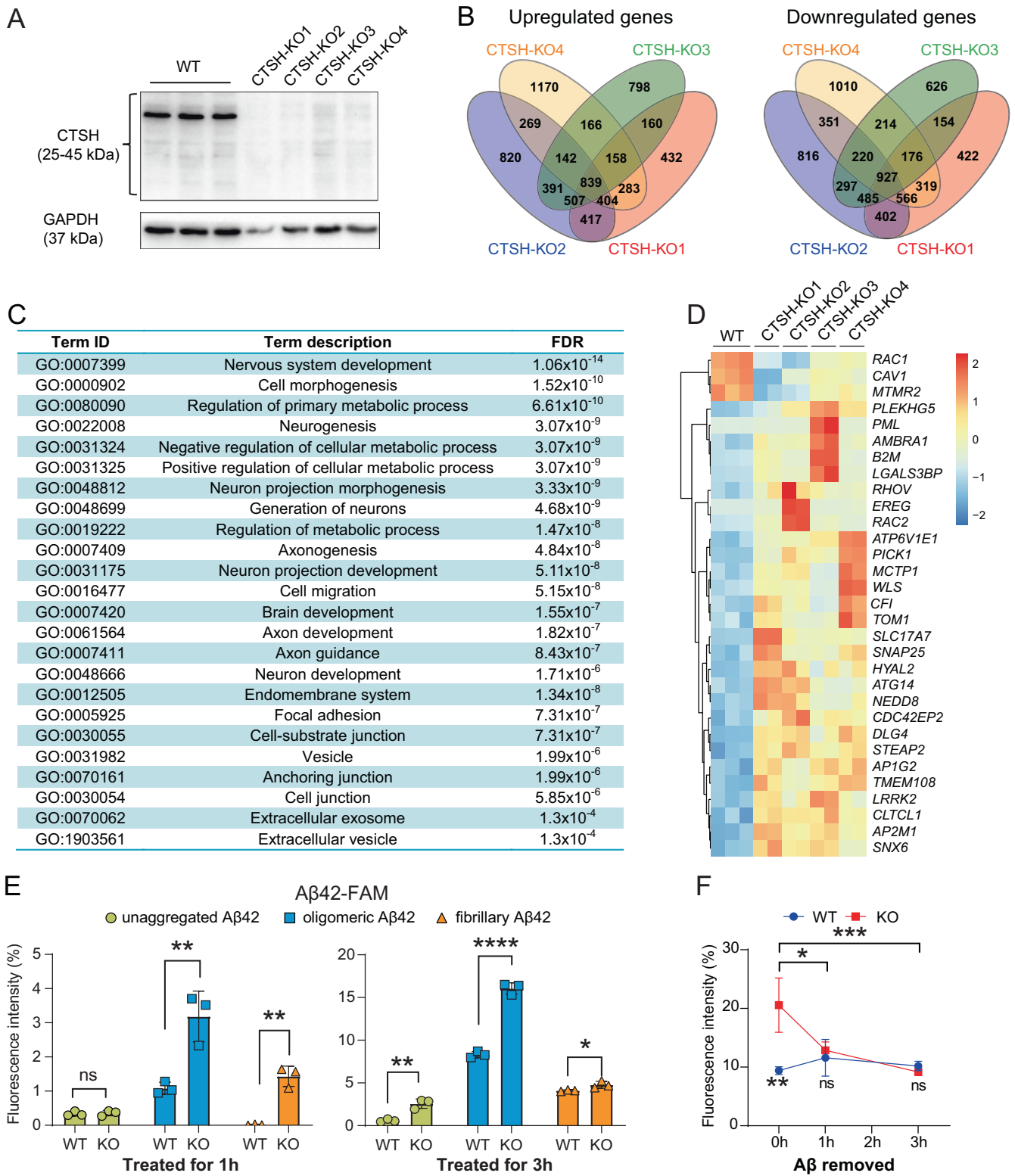


Fig. 4 Knockout of the *CTSH* gene in human microglial cells alters gene expression pattern and increases phagocytosis of Aβ42. **A** Successful knockout of the *CTSH* gene in wild type human microglia (HM) cells by using the CRISPR/Cas9 system. Four *CTSH* knockout (KO) cell strains (CTSH-KO1, CTSH-KO2, CTSH-KO3, CTSH-KO4) were selected for the assays. The GAPDH was used as the loading control. **B** Venn diagram of differentially expressed genes among the four *CTSH* knockout cell strains. **C** Pathway enrichment analyses of differentially expressed genes between wild type and *CTSH* KO in HM cell strains. FDR, false discovery rate. **D** Heatmap of genes enriched in endocytosis pathway. **E** Phagocytosis of Aβ42 in WT and *CTSH* KO HM cells. Cells were treated with 1 μg/mL of fluorescently-labeled Aβ42 in three different forms (oligomeric, aggregated, and fibrillary). Cells were harvested at 1 h and 3 h after the treatment and were analyzed by flow cytometry at 535 nm. Data are presented as mean ± SD and are quantified using a two-tailed Student's *t* test. ns not significant; **P* < 0.05; ***P* < 0.01; *****P* < 1.0 × 10⁻⁴. **F** Fluorescent intensity of fluorescently-labeled Aβ42 in HM WT cells and *CTSH* KO cells after the removal of the fluorescently-labeled Aβ42. Data are presented as mean ± SD. ns not significant; **P* < 0.05; ***P* < 0.01; *****P* < 1.0 × 10⁻³; two-way ANOVA with the Tukey's multiple comparisons test.

workable way is to test the effect of this variant in human-induced pluripotent stem cells with different alleles of rs2289702, which deserves a test in future study. Second, cultured microglia may behave differently from microglia *in vivo*, especially in response to stimuli. It would be desirable to confirm that *CTSH* KO would have an ameliorating effect on AD development using animal models [95].

In summary, we performed a functional genomics assay and identified causal variant underlying the protective *CTSH* locus for AD. Further studies with this gene are definitely needed to fully uncover the involvement of *CTSH* in AD pathogenesis.

REFERENCES

- De Strooper B, Karran E. The cellular phase of Alzheimer's disease. *Cell*. 2016;164:603–15.
- Querfurth HW, LaFerla FM. Alzheimer's disease. *N Engl J Med*. 2010;362:329–44.
- Guo T, Zhang D, Zeng Y, Huang TY, Xu H, Zhao Y. Molecular and cellular mechanisms underlying the pathogenesis of Alzheimer's disease. *Mol Neurodegener*. 2020;15:40.
- Karch CM, Goate AM. Alzheimer's disease risk genes and mechanisms of disease pathogenesis. *Biol Psychiatry*. 2015;77:43–51.
- Bertram L, Lill CM, Tanzi RE. The genetics of Alzheimer disease: back to the future. *Neuron*. 2010;68:270–81.
- Zhang DF, Xu M, Bi R, Yao YG. Genetic analyses of Alzheimer's disease in China: achievements and perspectives. *ACS Chem Neurosci*. 2019;10:890–901.
- Gatz M, Reynolds CA, Fratiglioni L, Johansson B, Mortimer JA, Berg S, et al. Role of genes and environments for explaining Alzheimer disease. *Arch Gen Psychiatry*. 2006;63:168–74.
- Bellenguez C, Kucukali F, Jansen IE, Kleindam L, Moreno-Grau S, Amin N, et al. New insights into the genetic etiology of Alzheimer's disease and related dementias. *Nat Genet*. 2022;54:412–36.
- Wingo AP, Liu Y, Gerasimov ES, Gockley J, Logsdon BA, Duong DM, et al. Integrating human brain proteomes with genome-wide association data implicates new proteins in Alzheimer's disease pathogenesis. *Nat Genet*. 2021;53:143–6.
- Jia L, Li F, Wei C, Zhu M, Qu Q, Qin W, et al. Prediction of Alzheimer's disease using multi-variants from a Chinese genome-wide association study. *Brain*. 2021;144:924–37.
- Jiao B, Xiao X, Yuan Z, Guo L, Liao X, Zhou Y, et al. Associations of risk genes with onset age and plasma biomarkers of Alzheimer's disease: a large case-control study in mainland China. *Neuropsychopharmacology*. 2022;47:1121–7.
- Wang HZ, Bi R, Hu QX, Xiang Q, Zhang C, Zhang DF, et al. Validating GWAS identified risk loci for Alzheimer's disease in Han Chinese populations. *Mol Neurobiol*. 2016;53:379–90.
- Tan L, Yu JT, Zhang W, Wu ZC, Zhang Q, Liu QY, et al. Association of GWAS-linked loci with late-onset Alzheimer's disease in a northern Han Chinese population. *Alzheimers Dement*. 2013;9:546–53.
- Bi R, Zhang W, Zhang DF, Xu M, Fan Y, Hu QX, et al. Genetic association of the cytochrome c oxidase-related genes with Alzheimer's disease in Han Chinese. *Neuropsychopharmacology*. 2018;43:2264–76.
- Zhang DF, Fan Y, Xu M, Wang G, Wang D, Li J, et al. Complement C7 is a novel risk gene for Alzheimer's disease in Han Chinese. *Natl Sci Rev*. 2019;6:257–74.
- Luo R, Fan Y, Yang J, Ye M, Zhang DF, Guo K, et al. A novel missense variant in *ACAA1* contributes to early-onset Alzheimer's disease, impairs lysosomal function, and facilitates amyloid- β pathology and cognitive decline. *Signal Transduct Target Ther*. 2021;6:325.
- Liao X, Cai F, Sun Z, Zhang Y, Wang J, Jiao B, et al. Identification of Alzheimer's disease-associated rare coding variants in the *ECE2* gene. *JCI Insight*. 2020;5:135119.
- Khaira AS, Wimberly CE, Semmes EC, Hurst JH, Walsh KM. An integrated genome and phenome-wide association study approach to understanding Alzheimer's disease predisposition. *Neurobiol Aging*. 2022;118:117–23.
- Campion D, Pottier C, Nicolas G, Le Guennec K, Rovelet-Lecrux A. Alzheimer disease: modeling an A β -centered biological network. *Mol Psychiatry*. 2016;21:861–71.
- Kikuchi M, Hara N, Hasegawa M, Miyashita A, Kuwano R, Ikeuchi T, et al. Enhancer variants associated with Alzheimer's disease affect gene expression via chromatin looping. *BMC Med Genomics*. 2019;12:128.
- Nott A, Holtman IR, Coufal NG, Schlachetzki JCM, Yu M, Hu R, et al. Brain cell type-specific enhancer-promoter interactome maps and disease-risk association. *Science*. 2019;366:1134–9.
- Johnson ECB, Dammer EB, Duong DM, Ping L, Zhou M, Yin L, et al. Large-scale proteomic analysis of Alzheimer's disease brain and cerebrospinal fluid reveals early changes in energy metabolism associated with microglia and astrocyte activation. *Nat Med*. 2020;26:769–80.
- Corces MR, Shcherbina A, Kundu S, Gloudemans MJ, Fresard L, Granja JM, et al. Single-cell epigenomic analyses implicate candidate causal variants at inherited risk loci for Alzheimer's and Parkinson's diseases. *Nat Genet*. 2020;52:1158–68.
- Wingo AP, Fan W, Duong DM, Gerasimov ES, Dammer EB, Liu Y, et al. Shared proteomic effects of cerebral atherosclerosis and Alzheimer's disease on the human brain. *Nat Neurosci*. 2020;23:696–700.
- Nativio R, Lan Y, Donahue G, Sidoli S, Berson A, Srinivasan AR, et al. An integrated multi-omics approach identifies epigenetic alterations associated with Alzheimer's disease. *Nat Genet*. 2020;52:1024–35.
- Ioannidis JP, Thomas G, Daly MJ. Validating, augmenting and refining genome-wide association signals. *Nat Rev Genet*. 2009;10:318–29.
- Gallagher MD, Chen-Plotkin AS. The post-GWAS era: from association to function. *Am J Hum Genet*. 2018;102:717–30.
- Cannon ME, Mohlke KL. Deciphering the emerging complexities of molecular mechanisms at GWAS Loci. *Am J Hum Genet*. 2018;103:637–53.
- Kumasaka N, Knights AJ, Gaffney DJ. High-resolution genetic mapping of putative causal interactions between regions of open chromatin. *Nat Genet*. 2019;51:128–37.
- lotchkova V, Ritchie GRS, Geijs M, Morganello S, Min JL, Walter K, et al. GARFIELD classifies disease-relevant genomic features through integration of functional annotations with association signals. *Nat Genet*. 2019;51:343–53.
- Rojnik M, Jevnikar ZR, Doljak B, Turk S, Zidar N, Kos J. The influence of differential processing of procathepsin H on its aminopeptidase activity, secretion and subcellular localization in human cell lines. *Eur J Cell Biol*. 2012;91:757–64.
- Ruiz-Blazquez P, Pistorio V, Fernandez-Fernandez M, Moles A. The multifaceted role of cathepsins in liver disease. *J Hepatol*. 2021;75:1192–202.
- Ou YN, Yang YX, Deng YT, Zhang C, Hu H, Wu BS, et al. Identification of novel drug targets for Alzheimer's disease by integrating genetics and proteomes from brain and blood. *Mol Psychiatry*. 2021;26:6065–73.
- Kibinge NK, Relton CL, Gaunt TR, Richardson TG. Characterizing the causal pathway for genetic variants associated with neurological phenotypes using human brain-derived proteome data. *Am J Hum Genet*. 2020;106:885–92.
- Jansen IE, Savage JE, Watanabe K, Bryois J, Williams DM, Steinberg S, et al. Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nat Genet*. 2019;51:404–13.
- Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics*. 2010;26:2336–7.
- The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature*. 2015;526:68–74.
- Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*. 2005;21:263–5.
- Wallace C. Eliciting priors and relaxing the single causal variant assumption in colocalisation analyses. *PLoS Genet*. 2020;16:e1008720.
- Consortium GT. The Genotype-Tissue Expression (GTEx) project. *Nat Genet*. 2013;45:580–5.
- Wang G, Zhang DF, Jiang HY, Fan Y, Ma L, Shen Z, et al. Mutation and association analyses of dementia-causal genes in Han Chinese patients with early-onset and familial Alzheimer's disease. *J Psychiatr Res*. 2019;113:141–7.
- Wang D, Fan Y, Malhi M, Bi R, Wu Y, Xu M, et al. Missense variants in *HIF1A* and *LACC1* contribute to leprosy risk in Han Chinese. *Am J Hum Genet*. 2018;102:794–805.
- Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. 2016;536:285–91.
- Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genome-wide association scans. *Bioinformatics*. 2010;26:2190–1.
- Xu M, Liu Q, Bi R, Li Y, Zeng C, Yan Z, et al. A multiple-causal-gene-cluster model underlying GWAS signals of Alzheimer's disease. *bioRxiv*. 2021;<https://doi.org/10.1101/2021.05.14.444131>.
- Encode Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489:57–74.
- Davis CA, Hitz BC, Sloan CA, Chan ET, Davidson JM, Gabdank I, et al. The Encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Res*. 2018;46:D794–D801.
- Song M, Yang X, Ren X, Maliskova L, Li B, Jones IR, et al. Mapping cis-regulatory chromatin contacts in neural cells links neuropsychiatric disorder risk variants to target genes. *Nat Genet*. 2019;51:1252–62.
- Fullard JF, Hauberg ME, Bendl J, Egervari G, Cirmaru MD, Reach SM, et al. An atlas of chromatin accessibility in the adult human brain. *Genome Res*. 2018;28:1243–52.
- Novakovic B, Habibi E, Wang SY, Arts RJW, Davar R, Megchelenbrink W, et al. β -Glucan reverses the epigenetic state of LPS-induced immunological tolerance. *Cell*. 2016;167:1354–68.e14.

51. Bailey TL, Elkan C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol*. 1994;2:28–36.
52. Zuo C, Shin S, Keles S. atSNP: transcription factor binding affinity testing for regulatory SNP detection. *Bioinformatics*. 2015;31:3353–5.
53. Consortium GT, Laboratory DA, Coordinating Center -Analysis Working G, Statistical Methods groups-Analysis Working G, Enhancing Gg, Fund NIHC. et al. Genetic effects on gene expression across human tissues. *Nature*. 2017;550:204–13.
54. Consortium GT. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science*. 2015;348:648–60.
55. Mathys H, Davila-Velderrain J, Peng Z, Gao F, Mohammadi S, Young JZ, et al. Single-cell transcriptomic analysis of Alzheimer's disease. *Nature*. 2019;570:204–7.
56. Xu M, Zhang DF, Luo R, Wu Y, Zhou H, Kong LL, et al. A systematic integrated analysis of brain expression profiles reveals YAP1 and other prioritized hub genes as important upstream regulators in Alzheimer's disease. *Alzheimers Dement*. 2018;14:215–29.
57. Matarin M, Salih DA, Yasvoina M, Cummings DM, Guelfi S, Liu W, et al. A genome-wide gene-expression analysis and database in transgenic mice during development of amyloid or tau pathology. *Cell Rep*. 2015;10:633–44.
58. Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. Genome engineering using the CRISPR-Cas9 system. *Nat Protoc*. 2013;8:2281–308.
59. Concordet JP, Haeussler M. CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. *Nucleic Acids Res*. 2018;46:W242–W5.
60. Gao Z, Harwig A, Berkhout B, Herrera-Carrillo E. Mutation of nucleotides around the +1 position of type 3 polymerase III promoters: the effect on transcriptional activity and start site usage. *Transcription*. 2017;8:275–87.
61. Wu Y, Liang D, Wang Y, Bai M, Tang W, Bao S, et al. Correction of a genetic disease in mouse via use of CRISPR-Cas9. *Cell Stem Cell*. 2013;13:659–62.
62. Jiao L, Su LY, Liu Q, Luo R, Qiao X, Xie T, et al. GSNOR deficiency attenuates MPTP-induced neurotoxicity and autophagy by facilitating CDK5 S-nitrosation in a mouse model of Parkinson's disease. *Free Radic Biol Med*. 2022;189:111–21.
63. Li H, Su LY, Yang L, Li M, Liu Q, Li Z, et al. A cynomolgus monkey with naturally occurring Parkinson's disease. *Natl Sci Rev*. 2021;8:nwaa292.
64. Yao YG, Kajigaya S, Young NS. Mitochondrial DNA mutations in single human blood cells. *Mutat Res*. 2015;779:68–77.
65. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012;16:284–7.
66. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*. 2019;47:D607–D13.
67. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* 2005;102:15545–50.
68. Dolgalev I msigdb: MSigDB gene sets for multiple organisms in a tidy data format. R package. 2022. <https://CRAN.R-project.org/package=msigdb>.
69. Lau SF, Cao H, Fu AKY, Ip NY. Single-nucleus transcriptome analysis reveals dysregulation of angiogenic endothelial cells and neuroprotective glia in Alzheimer's disease. *Proc Natl Acad Sci USA* 2020;117:25800–9.
70. Grubman A, Chew G, Ouyang JF, Sun G, Choo XY, McLean C, et al. A single-cell atlas of entorhinal cortex from individuals with Alzheimer's disease reveals cell-type-specific gene expression regulation. *Nat Neurosci*. 2019;22:2087–97.
71. Korin B, Ben-Shaan TL, Schiller M, Dubovik T, Azulay-Debby H, Boshnak NT, et al. High-dimensional, single-cell characterization of the brain's immune compartment. *Nat Neurosci*. 2017;20:1300–9.
72. Heckmann BL, Teubner BJW, Tummers B, Boada-Romero E, Harris L, Yang M, et al. LC3-associated endocytosis facilitates β -amyloid clearance and mitigates neurodegeneration in murine Alzheimer's disease. *Cell*. 2019;178:536–51.e14.
73. McAlpine CS, Park J, Gričič A, Kim E, Choi SH, Iwamoto Y, et al. Astrocytic interleukin-3 programs microglia and limits Alzheimer's disease. *Nature*. 2021;595:701–6.
74. Kaksonen M, Roux A. Mechanisms of clathrin-mediated endocytosis. *Nat Rev Mol Cell Biol*. 2018;19:313–26.
75. Wu LG, Hamid E, Shin W, Chiang HC. Exocytosis and endocytosis: modes, functions, and coupling mechanisms. *Annu Rev Physiol*. 2014;76:301–31.
76. Shin W, Wei L, Arpino G, Ge L, Guo X, Chan CY, et al. Preformed omega-profile closure and kiss-and-run mediate endocytosis and diverse endocytic modes in neuroendocrine chromaffin cells. *Neuron*. 2021;109:3119–34.
77. Zhao B, Wang Q, Du J, Luo S, Xia J, Chen YG. PICK1 promotes caveolin-dependent degradation of TGF-beta type I receptor. *Cell Res*. 2012;22:1467–78.
78. Bourguignon V, Flamion B. Respective roles of hyaluronidases 1 and 2 in endogenous hyaluronan turnover. *FASEB J*. 2016;30:2108–14.
79. Lee HW, Kim Y, Han K, Kim H, Kim E. The phosphoinositide 3-phosphatase MTMR2 interacts with PSD-95 and maintains excitatory synapses by modulating endosomal traffic. *J Neurosci*. 2010;30:5508–18.
80. Malecz N, McCabe PC, Spaargaren C, Qiu R, Chuang Y, Symons M. Synaptojanin 2, a novel Rac1 effector that regulates clathrin-mediated endocytosis. *Curr Biol*. 2000;10:1383–6.
81. Cheng ZJ, Singh RD, Holicky EL, Wheatley CL, Marks DL, Pagano RE. Co-regulation of caveolar and Cdc42-dependent fluid phase endocytosis by phosphocaveolin-1. *J Biol Chem*. 2010;285:15119–25.
82. Saftig P, Klumperman J. Lysosome biogenesis and lysosomal membrane proteins: trafficking meets function. *Nat Rev Mol Cell Biol*. 2009;10:623–35.
83. Braulke T, Bonifacino JS. Sorting of lysosomal proteins. *Biochim Biophys Acta*. 2009;1793:605–14.
84. Nakanishi H. Cathepsin regulation on microglial function. *Biochim Biophys Acta Proteom*. 2020;1868:140465.
85. Lopez-Otin C, Bond JS. Proteases: multifunctional enzymes in life and disease. *J Biol Chem*. 2008;283:30433–7.
86. Oberstein TJ, Utz J, Spitzer P, Klafki HW, Wiltfang J, Lewczuk P, et al. The role of cathepsin B in the degradation of A β and in the production of A β peptides starting with Ala2 in cultured astrocytes. *Front Mol Neurosci*. 2020;13:615740.
87. Chai YL, Chong JR, Weng J, Howlett D, Halsey A, Lee JH, et al. Lysosomal cathepsin D is upregulated in Alzheimer's disease neocortex and may be a marker for neurofibrillary degeneration. *Brain Pathol*. 2019;29:63–74.
88. Suire CN, Abdul-Hay SO, Sahara T, Kang D, Brizuela MK, Saftig P, et al. Cathepsin D regulates cerebral A β 42/40 ratios via differential degradation of A β 42 and A β 40. *Alzheimers Res Ther*. 2020;12:80.
89. Hook V, Funkelstein L, Wegrzyn J, Bark S, Kindy M, Hook G. Cysteine Cathepsins in the secretory vesicle produce active peptides: Cathepsin L generates peptide neurotransmitters and cathepsin B produces beta-amyloid of Alzheimer's disease. *Biochim Biophys Acta*. 2012;1824:89–104.
90. Sundelof J, Sundstrom J, Hansson O, Eriksdotter-Jonhagen M, Giedraitis V, Larsson A, et al. Higher cathepsin B levels in plasma in Alzheimer's disease compared to healthy controls. *J Alzheimers Dis*. 2010;22:1223–30.
91. Kim KR, Cho EJ, Eom JW, Oh SS, Nakamura T, Oh CK, et al. S-Nitrosylation of cathepsin B affects autophagic flux and accumulation of protein aggregates in neurodegenerative disorders. *Cell Death Differ*. 2022;29:2137–50.
92. Zamolodchikov D, Duffield M, Macdonald LE, Alessandri-Haber N. Accumulation of high molecular weight kininogen in the brains of Alzheimer's disease patients may affect microglial function by altering phagocytosis and lysosomal cathepsin activity. *Alzheimers Dement*. 2022;18:1919–29.
93. Xie Z, Meng J, Kong W, Wu Z, Lan F, Narengaowa, et al. Microglial cathepsin E plays a role in neuroinflammation and amyloid β production in Alzheimer's disease. *Aging Cell*. 2022;21:e13565.
94. Li Y, Ma C, Li W, Yang Y, Li X, Liu J, et al. A missense variant in NDUFA6 confers schizophrenia risk by affecting YY1 binding and NAGA expression. *Mol Psychiatry*. 2021;26:6896–911.
95. Chen ZY, Zhang Y. Animal models of Alzheimer's disease: applications, evaluation, and perspectives. *Zool Res*. 2022;43:1026–40.

ACKNOWLEDGEMENTS

We thank Dr. Ian Logan for his helpful comments and language editing. We would like to thank the Institutional Center for Shared Technologies and Facilities of Kunming Institute of Zoology, Chinese Academy of Sciences for providing us with confocal microscopy image acquisition and flow cytometric analysis. We are grateful to Cong Li for his technical support.

AUTHOR CONTRIBUTIONS

YL, YGY, and RB conceived and designed the study. YL, YGY, and RB compiled the figures and wrote the manuscript with help and input from all authors. YL, RB, BLX, MX and XL performed experiments. MX, DFZ, and RB performed the data processing and analysis. HZ revised the draft. All authors revised the manuscript and approved the publication.

FUNDING

This study was supported by the National Natural Science Foundation of China (32230021, 31970560, and 31730037), The STI2030-Major Projects (2021ZD0200900), Yunnan Province (202003AD150009, 2019FA027, and 202101AT070285), and Strategic Priority Research Program (B) of the Chinese Academy of Sciences (CAS) (XDB32020200), the International Partnership Program of Chinese Academy of Sciences (152453KY5B20170031), the Bureau of Frontier Sciences and Education, CAS (QYZDJ-SSW-SMC005) and the Youth Innovation Promotion Association of CAS (2020000017).

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41386-023-01542-2>.

Correspondence and requests for materials should be addressed to Rui Bi or Yong-Gang Yao.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.