SUPPLEMENTARY INFORMATION

Coexistence of Multiple Functional Variants and Genes Underlie Genetic Risk Locus 11p11.2 of Alzheimer's Disease

Xu *et al.*

Supplementary Materials and Methods

Functional genomic strategy for fine-mapping of potentially functional variants and likely causal genes at 11p11.2

A regulatory *functional variant* (fVar) was defined as a genomic variant capable of modulating gene expression by affecting the binding of certain transcription factor (TF) to the *active regulatory element* (ARE) in which the variant was located. Multi-omics data at bulk brain tissue or single-cell level from brain tissues, neural cells, and monocytes were included for the fine-mapping of fVars. Alzheimer's disease (AD)-associated single nucleotide polymorphisms (SNPs) from three recent large-scale AD genome-wide association studies (GWASs), the Lambert study (1), the Kunkle study (2), and the Jansen study (3) were initially subjected to the functional genomic analysis. Bulk brain tissue and monocyte histone modifications data (Chromatin Immunoprecipitation Sequencing, ChIP-seq) associated with active promoters and enhancers (H3K4me3, H3K9ac, H3K4me1 and H3K27ac) (4, 5), and open chromatin data (Assay for Transposase-Accessible Chromatin using sequencing, ATAC-seq) (6, 7) were integrated to identify SNPs located within AREs. Single-cell ChIP-seq (8) and ATAC-seq (8-10) data for different neural cells (including neurons, astrocytes, microglia, and oligodendrocytes) were also used to identify SNPs that may have a function in certain types of neural cells, especially microglia. ChIP-seq data of 623 TFs (4, 5) and the atSNP algorithm (11) were applied to further test whether SNPs within the AREs were able to affect the binding affinities of TFs to the AREs. Allelic effects of potential fVars were confirmed by allele imbalance analysis, dual-luciferase reporter assays, and base-editing.

We defined target gene of an fVar on basis of two criteria. First, its expression is associated with that fVar (eGene). Second, that gene has direct chromatin interactions

with the ARE in which the fVar is located. We integrated expressional quantitative trait loci (eQTL) datasets of bulk brain tissue (12, 13), microglia (14), and monocytes (15, 16) with chromatin interaction data of neurons (8, 10), astrocytes (10), oligodendrocytes (8), microglia (8), and monocytes (4, 5), respectively.

The convergent functional genomics (CFG) strategy (17, 18), which incorporated multiple lines of AD-related evidence, was used to assess a gene's relevance to AD pathogenesis (17, 19, 20). Briefly, one point was assigned if the target gene: i) was predicted to be the causal gene by integrative analyses of AD GWAS and eQTL datasets (14, 21, 22); ii) was differentially expressed at the early stage or at the late stage of AD at the single-cell (23) or bulk brain level (17); iii) was differentially acetylated (H3K27ac) (24), or iv) had a differential protein level (25) in AD patients compared to controls. The sum of all lines of evidence (the integrative analysis, early-stage mRNA change, late-stage mRNA change, epigenetic change, and protein change) resulted in a total CFG score ranging from 0 (no association) to 5 (the strongest association). Genes with CFG > 3 were considered to be more likely to be involved in AD pathogenesis. The effects of prioritized genes on AD-related molecular phenotypes were further confirmed by cell assays.

AD GWASs

The three recent large-scale AD GWASs, i.e. the Lambert study (1), the Kunkle study (2), and the Jansen study (3) were included for fine-mapping the AD-associated SNPs at 11p11.2. The Lambert study included 17,008 AD patients and 37,154 controls (Nsnps = 7,055,881) (1). The Kunkle study (2) had an updated dataset of the Lambert study (1), with 17 newly added sample datasets, which resulted in 21,982 AD patients and 41,944 controls (Nsnps = 36,648,992) (2). The Jansen study (3) was based on

clinical diagnosed AD patients and individuals with one or two parents with AD (AD-by-proxy). The Phase 3 of the Jansen study (3) was a meta-analysis of stage 1 data from the Lambert study (N = 54,162) (1), AD GWAS data of the Psychiatric Genomics Consortium (PGC, N = 17,477), whole-exome sequencing (WES) data of the Alzheimer's Disease Sequencing Project (ADSP, N = 7,506), and GWAS data of AD-by-proxy subjects and controls from the UK Biobank (N = 376,113). A total of 13,367,299 variants were genotyped or imputed in the Jansen study (3).

For association analysis of AD endophenotypes including age-at-onset of AD and β-amyloid 42 ($A\beta_{42}$) level in cerebrospinal fluid (CSF), we used two reported datasets (26, 27). Briefly, the GWAS of age-at-onset of AD included 14,406 AD patients and 25,849 controls (Nsnps = 8,253,925) (26). The GWAS data of amyloid beta ($A\beta_{42}$), tau, and phosphorylated tau (pTau-181) levels in CSF were conducted with 3146 subjects, and had 7,358,575 variants for analyses (27). Considering the fact that the sample size for GWAS of AD endophenotypes including $A\beta_{42}$ level in CSF (N=3146) was relatively small for achieving a robust statistical power, we arbitrarily used a loose threshold ($P < 0.001$) to define the association of 11p11.2 with AD endophenotypes.

Assignment of AD-associated SNPs at 11p11.2

Summary statistics from the Lambert study (stage 1 data) (1), the Kunkle study (2) (stage 1 data), and the Jansen study (3) were used to identify the AD-associated SNPs at 11p11.2. As the significant level is affected by sample size under study, some suggestively or marginally significant SNPs in a single GWAS report can achieve a genome-wide significance ($P < 5 \times 10^{-8}$) with increased sample size in further GWASs or meta-analyses. This has been testified by many large-scale meta-analyses of AD (2,

3) and other complex diseases (28-30). There is a high likelihood that these suggestive SNPs may also be biologically functional and relevant to AD, albeit with minor effects (31). Therefore, in order to maximize the coverage of subsequent functional mapping with an acceptable statistical credibility and to capture as many potentially risk SNPs as possible, we used a relatively moderate GWAS *P* cutoff ($P < 1 \times 10^{-5}$, which was also used by others (32, 33) to distill candidate AD risk SNPs at 11p11.2. The false discovery rate (FDR) corresponding to the chosen $P < 1 \times 10^{-5}$ threshold is $1x10^{-5}$ x 4049 SNPs at 11p11.2 = 0.04. SNPs with a minor allele frequency (MAF) > 0.01 and reached a suggestive genome-wide significance ($P < 1x10^{-5}$) (32, 33), and SNPs that were in tight linkage ($r^2 > 0.8$) with the above SNPs at 11p11.2 (chr11: 46.5 megabases [Mb]-48 Mb, hg19) were extracted from the three GWASs (1-3). Genotype data from the 1000 Genomes project Phase 3 (503 European individuals [EUR]) (34) were used to compute the linkage disequilibrium (LD) among the variants at 11p11.2. The SNPs from all three studies (1-3) were combined, and a total of 452 AD-associated SNPs were obtained for 11p11.2 (**Supplementary Table S2**). The effects of SNPs across the three studies (1-3) were checked and SNPs with inconsistent effects were discarded. We conducted gene-based annotations of these 452 SNPs by using ANNOVAR (35). The Combined Annotation Dependent Depletion (CADD) database (https://cadd.gs.washington.edu/) was used to evaluate the deleteriousness of missense variants (36).

Detection of LD blocks (LBs) at 11p11.2

LD-based clumping of GWAS summary statistics at 11p11.2 from the Lambert study (1) were conducted to detect LBs using Plink v1.9 (www.cog-genomics.org/plink/1.9/) (37). Genotyping data of 503 EUR individuals from the 1000 Genomes Project (phase

3) (34) were used as the reference. The significance threshold for index SNPs was arbitrarily set as $1x10^{-5}$ (--clump-p1 $1x10^{-5}$). The SNPs that were located within ±1 Mb (--clump-kb 1000) from the index SNP and in tight linkage (--clump-r2 0.8) with the index SNP were identified.

As defining LD blocks is quite challenging, we also used R package bigLD (38) to detect LD blocks at 11p11.2 using the 1000 Genomes EUR genotyping data (34) and compared with the LBs detected by using Plink v1.9 (37). The threshold for the correlation value |r| was set to 0.9 for bigLD (38), corresponding to $r^2 > 0.8$ that was used in Plink v1.9 (37).

QTL data of bulk brain tissues, microglia, and monocyte

For bulk brain eQTL data, we used the eMeta dataset (12) and the PsychENCODE dataset (13). The eMeta dataset was a meta-analysis (12) of bulk brain eQTL data from the Genotype-Tissue Expression (GTEx) project (39, 40), the CommonMind Consortium (CMC) (41), and the Religious Orders Study and Memory and Aging Project (ROSMAP) (42), with an effective sample size of 1194. The PsychENCODE dataset had 1387 postmortem prefrontal cortex samples (13). Microglia eQTL dataset was taken from the Microglia Genomic Atlas (MiGA, n = 216), which contains 216 primary human microglia samples isolated from medial frontal gyrus, superior temporal gyrus, subventricular zone, and thalamus of 90 subjects with neurological and psychiatric diseases, as well as unaffected subjects (14). We downloaded the results of meta-analysis (fixed effects) across different brain regions (14). The monocyte eQTL data were taken from Raj et al. (16) (N=461) and Kim-Hellmuth et al. (15) (N=134). Only *cis*-eQTL, i.e., SNPs within ±1 Mb from the gene transcriptional start site (TSS) or transcriptional end site (TES) were included. When referring to

significant eQTL genes for all SNPs at 11p11.2, we used a *P* threshold that was corrected by the total number of SNPs and genes within ±1Mb of 11p11.2 ($P < 0.05/35$ genes/452 SNPs = 3.2×10^{-6}). For eGene of a single SNP, we used an eQTL *P* threshold corrected by the total number of genes ($P < 0.05/35$ genes $\approx 1 \times 10^{-3}$).

Transcriptome-wide association study (TWAS) of eQTL and AD GWASs

We performed TWAS to infer potentially causal genes at 11p11.2. Two algorithms, i.e. MetaXcan (21) and SMR (22), were used in this study. The SMR (22) and MetaXcan (21) analyses need the reference eQTL panels, which were constructed by genotype and expression data. It is ideal that the TWAS could be repeated with microglial or monocyte reference panels. However, the available datasets for monocytes and microglia contained no such information. Therefore, we performed the SMR and MetaXcan analyses with the reference eQTL panels of whole blood and bulk brain tissues, respectively. In brief, summary statistics from the three GWASs (the Lambert study (1), the Kunkle study (2), and the Jansen study (3)) at 11p11.2 were integrated with eQTL datasets of different brain regions from the GTEx project (39, 40) and of peripheral blood (43, 44). Thirteen GTEx brain regions and nervous tissues, including amygdala (N = 100), anterior cingulate cortex (BA24) (N = 121), caudate (N = 160), cerebellar hemisphere (N = 136), cerebellum (N = 173), cortex (N = 158), frontal cortex (BA9) (N = 129), hippocampus (N = 123), hypothalamus (N = 121), nucleus accumbens (N = 147), putamen (N = 124), spinal cord (cervical c-1) (N = 91), and substantia nigra (N = 88), were included for analyses.

For MetaXcan analyses (21), pre-calculated databases for Depression Genes and Network's (DGN) whole blood (N = 922) (43) and GTEx were downloaded from the PredictDB Data Repository (http://predictdb.org/) (45). For SMR analyses (22), we

downloaded peripheral blood (N = 2765) (44) and GTEx eQTL data from the SMR website (http://cnsgenomics.com/software/smr/#eQTLsummarydata) (22). The significant threshold values were set as Bonferroni-corrected $P < 1x10^{-4}$ for the MetaXcan analyses (14 tested tissues and 35 genes within ±1 Mb of 11p11.2 captured in eQTL datasets), and as $P_{SMR} < 1x10^{-4}$ (Bonferroni-corrected) and $P_{HEIDI} > 0.05$ for the SMR analyses, respectively.

We compared our TWAS results with colocalization results reported by Lopes *et al*. (14) for cross validation. We obtained the results of colocalization analyses from Lopes et al. (14), which integrated eQTL data of bulk brain, monocyte, and microglia with AD GWASs (Ref. (14) and references therein). Genes with PP.H4 > 0.7 were considered to be colocalized. The detailed information regarding colocalization analyses can be found in the original publication (14).

Histone modification data of brain tissues, neural cells, and monocytes

ChIP-seq data for histone modifications related to active promoters (H3K4me3 and H3K9ac) (46), enhancers (H3K4me1 and H3K27ac) (47), and repressors (H3K27me3) (48) for eight brain regions (layer of hippocampus, temporal lobe, angular gyrus, caudate nucleus, cingulate gyrus, middle frontal area 46, substantia nigra, and embryonic brain) and six neural cells (astrocyte, bipolar neuron, neural cell, neural stem progenitor cell, neuron, and radial glial cell) were downloaded from the Encyclopedia of DNA Elements (ENCODE) (https://www.encodeproject.org) (**Supplementary Table S1**) (4, 5). Histone modification data for monocytes were also included for analysis in consideration of its critical role in AD (26). H3K4me3 and H3K27ac ChIP-seq peak files for neurons (NEUN+), astrocytes (NEUNneg LHX2+), microglia (PU.1+), and oligodendrocytes (OLIG2+) isolated from resected cortical

brain tissues (N = 6) (8) were obtained from the UCSC (https://genome.ucsc.edu/s/nottalexi/glassLab_BrainCellTypes_hg19). Peak files in bed format were obtained, and a FDR < 0.001 was applied to obtain relatively reliable peaks (49).

ATAC-seq data of bulk brain tissues, neural cells, and monocytes

ATAC-seq peaks of induced pluripotent stem cell (iPSC)-induced excitatory neurons, iPSC-derived hippocampal dentate gyrus (DG)-like neurons, and primary fetal astrocytes were downloaded from Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/), with accession number GSE113483 (10). The ATAC peaks of neuron and glia cells from 14 brain regions were downloaded from the Brain Open Chromatin Atlas (BOCA, http://icahn.mssm.edu/boca) (6). The ATAC-seq of monocytes was downloaded from the GEO database with accession number GSE87218 (7). ATAC-seq peak files for neurons (NEUN+), astrocytes (NEUNneg LHX2+), microglia (PU.1+), and oligodendrocytes (OLIG2+) isolated from resected cortical brain tissues (N = 6) (8) were obtained from the UCSC browser (https://genome.ucsc.edu/s/nottalexi/glassLab_BrainCellTypes_hg19). Single-cell ATAC-seq (scATAC-seq) data of isocortex (N = 3), striatum (N = 3), hippocampus (N = 2), and substantia nigra (N = 2), were downloaded from the GEO with accession number GSE147672 (9). A FDR < 0.001 was applied to filter ATAC peaks.

ChIP-seq of transcription factors (TFs) and differential TF binding analyses

To identify variants located in TF binding sites (TFBS), a total of 1,322 ChIP-seq datasets for 623 TFs were downloaded from the ENCODE database (**Supplementary Table S1**) (4, 5). Among these ChIP-seq datasets, only 37 datasets for 23 TFs were

collected from AD-related tissues or cells. These AD-related datasets had an insufficient coverage for the entire TFs. To remedy this limitation, we used the GTEx database (39, 40) to evaluate the expression levels of TFs in brain tissues. As we found that about 91% of TFs (565 out of 623 TFs with transcripts per million (TPM) > 1) are expressed in brain tissues (**Supplementary Figure S11**), we used ChIP-seq data of all TFs from all available tissues and cells for the subsequent TF binding analysis in order to achieve a higher coverage.

DNA sequences of the top 1000 peaks (ranked by peak height in bed files) for each TF were subjected to the motif-based sequence analysis tool MEME (https://meme-suite.org/meme/) (50) to predict the DNA binding motifs (position weight matrix, PWM) (-mod zoops -nmotifs 3 -minw 6 -maxw 30). Top 3 PWMs with the smallest *E*-values for each TF were subjected to R package atSNP (11) to predict whether different alleles of certain variant within the TFBS could affect binding affinities of this TF. A variant was considered to disrupt the TF binding affinity if DNA sequence with reference allele ($P\_ref < 0.05$) or alternative allele ($P\_alt < 0.05$) of this variant was able to bind to the target TF, and their binding affinities were significantly different ($P\_rank < 0.05$) (11).

Functional genomic fine-mapping of potential fVars

To decide whether the target SNP was located in any potential regulatory elements (promoter or enhancer), open chromatin, or TFBS, peaks of histone modification, TFs, and ATAC-seq were intersected with 452 AD-associated SNPs, respectively, using bedtools (51). A SNP was considered to locate in ARE if it was overlapped with the histone modification peaks and the open chromatin peaks (ATAC-seq) at the same time. If an ARE SNP was also located in the binding peak of a TF, and was also able

to affect the binding affinity of this TF to the peak, which was predicted by the atSNP algorithm (11), the SNP was regarded as a potentially fVar.

Allele specific expression (ASE) analyses

ASE measures allelic imbalance during transcription, which reflects the expression regulation activity of certain variant (39). We used ASE data for 53 tissues from the GTEx (phs000424.v7.p2) (39, 40, 52) to verify the *cis*-regulatory effects of candidate fVars. Only SNPs that were heterozygous in GTEx individuals and were captured by RNA sequencing (RNA-seq) were suitable for the ASE analyses. Among the 24 fVars identified in the above analyses, only 12 variants met this criterion for the ASE analyses. As it was unable to draw meaningful conclusions for variants with relatively low capture rates, we only included variants that were detected in > 10 samples in the ASE analysis. Binomial tests were used to test if the ratio of the two alleles of target variant was significantly different from the expectation (52). ASE analyses were performed by using a pooled data of all GTEx tissues, brain tissues only, and whole blood (39, 40, 52), respectively.

Cell lines and cell culture

HEK293T cells, U251 cells, human microglia HMC3 cells, and HM cells were obtained from Kunming Cell Bank, Kunming Institute of Zoology. HEK293T cells and HM cells were cultured in Dulbecco's modified Eagle's medium (DMEM; Gibco-BRL, 11965-092). U251 cells were cultured in Roswell RPMI-1640 medium (Gibco-BRL, C11875500BT). HMC3 cells were cultured in MEM medium (Procell, PM150410). All culture media were supplemented with 10% fetal bovine serum (FBS,

Gibco-BRL, 10099-141), 100 U/mL penicillin and 100 mg/mL streptomycin. Cells were cultured at 37 °C in a humidified atmosphere incubator with 5% CO_2.

Vector construction and dual-luciferase reporter assays

The DNA fragments containing the target SNPs were amplified from in-house human DNA samples (20, 53) (**Supplementary Table S11)**. A DNA fragment containing rs1542321, rs11039200, and rs10734557 was commercially synthesized (Tsingke Biotechnology Co. Ltd., Nanjing, China). The DNA fragments were inserted into the pGL3-basic (Promega, for promoter assays) or pGL3-promoter (Promega, for enhancer assays) luciferase reporter vector. PCR-mediated point mutagenesis was used to generate DNA vectors containing the respective alleles of each target SNP (**Supplementary Table S11**). All inserted DNA fragments were verified by Sanger sequencing.

We validated the allelic regulatory effects by using dual-luciferase reporter assays, which were performed in the above four cell lines. We used HEK293T cells and U251 cells to test all 11 fVars (including 7 fVars with significant ASE and 4 fVars without ASE data). We chose these two cell lines based on two reasons. First, most genes at 11p11.2 were ubiquitously expressed in different cells (**Supplementary Figure S7**). Second, most of active regulatory elements (AREs) containing fVars were also active in HEK293T and U251 cell lines (**Supplementary Figure S12**). For the three fVars (rs10734557, rs1542321 and rs11039200) in the enhancer of *SPI1*, which was primarily expressed in microglia and monocytes (26, 54, 55), we repeated luciferase reporter assays by using human microglia cell lines HMC3 and HM.

HEK293T cells were grown in 48-well plates with six replicates for each vector. U251 cells were grown in 24-well plates with four replicates per vector. HMC3 and

HM cells were grown in 24-well plates with six replicates per vector. The pGL3 vector (250 ng per well for the 48-well plate, and 500 ng per well for the 24-well plate) and the internal control vector phRL-TK (25 ng per well for the 48-well plate, and 50 ng per well for the 24-well plate) were co-transfected into the cells. The X-tremeGene HP DNA transfection reagent (ROCHE, 6366236001) was used for transfection. HEK293T cells were harvested at 24 h post transfection, U251, HMC3, and HM cells were harvested at 48 h using passive lysis buffer (Promega). Luminoskan Ascent instrument (Thermo Fisher Scientific Inc.) was used to measure the firefly and Renilla luciferase activities with the Dual-Luciferase Reporter Assay System (Promega, E1910) following the manufacture's instruction.

Assign candidate target genes to potential fVars with chromatin interaction and eQTL data

Promoter capture HiC (pc-HiC) data of iPSC-derived hippocampal DG-like neurons, iPSC-induced cortical excitatory neurons, and human primary fetal astrocytes were downloaded from the GEO database with accession number GSE113481 (10). Proximity ligation-assisted ChIP-seq (PLAC-seq) data of microglia, neurons and oligodendrocytes were obtained from the UCSC (https://genome.ucsc.edu/s/nottalexi/glassLab_BrainCellTypes_hg19) (8). HiC data of CD14-positive monocyte were downloaded from the ENCODE (accession IDs: ENCSR236EYO and ENCSR444SKT) (4, 5). Consensus regulatory elements at 11p11.2 were obtained by merging promoter and enhancer peaks from all histone modification datasets included in this study (**Supplementary Table S1**) (4, 5) with the mergePeaks function (http://homer.ucsd.edu/homer/ngs/mergePeaks.html, -d given). A gene was considered to interact with the potentially fVar if its promoter or

enhancer significantly interacted (interaction score > 3, corresponding to $P < 0.001$ (10)) with the regulatory element containing the fVar. Bulk-brain eQTL were integrated with HiC data from all types of neural cells (neurons, astrocytes, microglia, and oligodendrocytes) to assign candidate target genes for potentially fVars. Microglia eQTL and microglia PLAC-seq, monocyte eQTL and monocyte HiC were integrated, respectively, to assign microglia-specific and monocyte-specific target genes for fVars. In order to obtain relatively reliable target genes for fVars, only a gene that physically interacted (chromatin interaction) and were expressionally associated (eGene) with the fVar at the same time was regarded as the target gene for the fVar. Because two different levels of data were applied to ensure the reliability, we thus used less stringent cutoffs for eGenes and chromatin interactions. We used an eQTL $P<0.001$ to define an eGene and a HiC score>3 (corresponded to $P < 1x10^{-3}$) to define a significant chromatin interaction. In addition, more stringent cutoffs for eGenes ($P < 3.2x10^{-6}$) and chromatin interactions (HiC score > 5, corresponding to $P < 1x10^{-5}$ (10)) were also applied. If an fVar was located in a gene, which was also labeled as the eGene of this particular fVar, we defined this gene as the target gene of this fVar.

Cell-type expression specificity analyses

Cell-type expression specificity of certain gene was tested using the scRNA-seq data from prefrontal cortex samples of AD patients (N = 24, including patients at the early-stage and the late-stage pathology) and controls (N = 24) (23). Cell-type specificity (i.e. proportion of total expression of a gene in one cell type compared to all other cell types) metric was calculated for certain gene using the generate.celltype.data function from the expression-weighted cell-type enrichment (EWCE) R package (56).

Base-editing of target variants

Base-editing was used to generate precise point mutations in cellular DNA for the target variants. Guide RNAs (gRNAs) targeting to the genomic regions of the variants were designed (**Supplementary Table S11**), and were sub-cloned into the pGL3-U6-sgRNA-PGK-puromycin (57) (Addgene plasmid # 51133) plasmid. Constructs containing different gRNAs (500 ng per well for the 6-well plate) were co-transfected with pCMV-ABE7.10 (58) (2 μg per well for the 6-well plate, Addgene plasmid # 102919) into HEK293T cells by using Lipofectamine^{TM} 3000 (Thermo Fisher Scientific Inc.). Culture medium was changed daily with fresh medium supplemented with 2 μg/mL puromycin after transfection for 24 h, and cells were selected by puromycin for 5 days. Single cells resistant to puromycin were seeded in the 96-well plate and were cultured for 2-3 weeks to obtain single cell clones. For each clone, the target region was amplified and sequenced to confirm successful editing of the target variants.

ATAC-seq library preparation and data analyses

ATAC-seq libraries were prepared using the TruePrep® DNA Library Prep Kit (Vazyme, TD501) following the manufacturer's instruction. Briefly, 1×10^5 HEK293T cells pellet was re-suspended in 50 μL of cold lysis buffer (Sigma-Aldrich, NUC101) to generate nuclei, followed by centrifuging at 500 ×g for 10 min at 4 ºC to remove the supernatant. The nuclei pellet was immediately continued to transposition reaction with Tn5 transposome at 37 ºC for 30 min and was purified using the KAPA Pure Beads (KAPA Biosystems, ks8002). The transposed DNA fragments were amplified following by 72 °C for 3 min, 98 °C for 30 sec, and 9 cycles (each cycle: 98 °C for 15 sec, 60 °C for 30 sec and 72 °C for 30 sec), followed by a final incubation at 72 °C for

5 min. The amplified PCR products were purified using the KAPA Pure Beads to get the ATAC-Seq libraries. Library qualities were assessed by gel electrophoresis and Agilent 2100 Bioanalyzer.

The ATAC-seq libraries were sequenced on the Novaseq 6000 platform, and 150 bp paired-ends reads were generated. The ENCODE ATAC-seq pipeline (https://github.com/ENCODE-DCC/atac-seq-pipeline) (4, 5) with default settings was used for the quality control and processing of ATAC-seq data. Briefly, adaptors and low-quality reads were trimmed and the remaining reads were mapped to human reference genome (GRCh38). PCR duplicates and reads mapping to mitochondrial DNA were filtered. Narrow peaks were called and peaks within blacklist regions (https://storage.googleapis.com/encode-pipeline-genome-data/hg38/hg38.blacklist.bed.gz) were discarded. Peaks called from different samples were merged by the mergePeaks function from the HOMER tool set (http://homer.ucsd.edu/homer/ngs/mergePeaks.html). The maximum distance between peak centers to merge was set as 1000 bp. Consensus peaks were obtained by extracting peaks detected in at least 3 samples. Read counts for all consensus peaks were quantified by featureCounts (59) and were normalized by counts per million (CPM) by the calculateCPM function in R package scater (60), with adjustment of library size for each sample. For visualization, coverage of peaks was normalized using CPM in the bamCoverage function (binsize = 10) in deeptools (61).

Real-time quantitative PCR (RT-qPCR)

Total RNA was extracted by using the RNAeasy kit (TIANGEN Biotech Co. Ltd., Beijing, China) according to the manufacturer's instructions. The quality of total RNA was measured on a NanoDrop biophotometer (Thermo Fisher Scientific Inc.). Total

RNA (1 μg) was used to synthesize cDNA by using oligo-dT18 primer and Moloney murine leukemia virus reverse transcriptase (M1701, Promega). The RT-qPCR was performed using iTaq Universal SYBR Green Supermix (Bio-Rad Laboratories, 172-5125) and the gene-specific primer pairs (**Supplementary Table S11**) on a CFX Connect Real-Time PCR Detection System (Bio-Rad Laboratories). The *ACTB* transcript was used for normalizing the expression of each target gene.

Brain transcriptomic, epigenomic, and proteomic data of AD patients and controls

Bulk brain tissue mRNA expression data of AD patients and controls were obtained from our previous study (AlzData: www.alzdata.org) ((17) and references therein). In brief, renormalized expression data for four brain regions were included, including the entorhinal cortex (EC, N_{AD} = 39, N*control* = 39), hippocampus (HP, N_{AD} = 74, N*control* = 65), frontal cortex (FC, N_{AD} = 104, N*control* = 128), and temporal cortex (TC, N_{AD} = 52, N*control* = 39). A gene with Benjamini-Hochberg's (BH) adjusted P (FDR) < 0.05 was considered as the differential expressed gene (DEG) in AD patients compared to controls. Single cell RNA sequencing (scRNA-seq) data from prefrontal cortex of AD patients and controls were taken from Mathys et al. (23). Briefly, a total of 15 AD patients with early-stage pathology, 9 AD patients with late-stage pathology, and 24 controls were included in this study. Genes with a FDR < 0.05 in both the two-sided Wilcoxon rank-sum test and the Poisson mixed-model test were defined as DEGs at the single-cell level (23). Raw H3K27ac count data in postmortem EC samples from 24 AD patients and 23 controls were downloaded from the GEO database with accession number GSE102538 (24). Counts were normalized by CPM using the calculateCPM function in R package scater (60). The H3K27ac levels for

each regulatory element in AD and controls were compared by using the Student's *t* test. Peak files for visualization were downloaded from https://epigenetics.essex.ac.uk/AD_H3K27ac/ (24). The protein abundance data for dorsolateral prefrontal cortex from 91 controls and 230 AD patients were obtained from the original proteomic study (25). The differences of H3K27ac levels and protein abundance between AD patients and controls were compared using the Student's *t* test.

Knockdown or overexpression of likely causal genes

U251 cells with a stable expression of mutant *APP* constructed in our previous studies (U251-APP cells) (62, 63), were used to test the effect of expressional change of target gene on $A\beta_{42}$ and pTau (pTau396) levels. For knockdown assay, siRNA of each gene (20 nM per well for the 6-well plate) was transfected into cells by using Lipofectamine^{TM} 3000 (Thermo Fisher Scientific Inc.). For overexpression assay, expression vector of target gene (2 μg per well for the 6-well plate) was transfected by using the X-tremeGene HP DNA transfection reagent (ROCHE, 6366236001). After transfection for 24 h, culture supernatant in each well was replaced with equal volume of fresh growth medium, and 1 μg/mL doxycycline (Sigma, D9891) was added to induce APP expression. Cells and culture supernatant were harvested at 72 h after transfection.

Western blot and enzyme-linked immunosorbent assay (ELISA)

Cells were lysed by protein lysis buffer (Beyotime, P0013) on ice and were centrifuged at 12 000 ×g at 4 °C for 10 min to remove cell debris. Protein concentration was determined using a BCA Protein Assay Kit (Beyotime, P0012). A

total of 20 μg protein was separated by 12% (vol/vol) sodium dodecyl sulphate (SDS)-polyacrylamide gel and electrophoretically transferred to a polyvinylidene difluoride membrane (Bio-Rad Laboratories, L1620177). The membranes were soaked with 5% (w/v) skim milk for 2 h at room temperature, and were incubated with the primary antibodies against MADD (1:1000; abcam, ab134117), MTCH2 (1:1000; absin, abs143485), PSMC3 (1:1000; abcam, ab171969), Flag (1:5000; Abmart, M20008), and Tubulin (1:20000; EnoGene, E1C601) overnight at 4 °C, respectively. After three washes with Tris-buffered saline with 0.1% Tween (TBST, 5 min each), the membranes were incubated with the respective anti-mouse or anti-rabbit secondary antibody (1:10000, KPL, USA) for 1 h at room temperature. The membranes were visualized using enhanced chemiluminescence reagents (Millipore, WBKLS0500).

The levels of $A\beta_{40}$ (Elabscience, E-EL-H0542c) and $A\beta_{42}$ (Elabscience, E-EL-H0543c) in culture supernatant, and phosphorylated tau (pTau-396, Elabscience, E-EL-H5314c) in cell lysate of U251-APP cells with different transfections were measured by using commercial ELISA kits. A total of 100 μL culture supernatant or cell lysate were used to perform the ELISA assays according to the manufacturer's instructions, respectively. The protein level of $A\beta_{42}$ and pTau-396 were further normalized by the total amount of protein of each sample. However, $A\beta_{1\text{-}40}$ was undetectable in U251-APP cell line and was excluded in the subsequent analysis.

Stepwise conditional analysis for 11p11.2

A stepwise model selection procedure was performed by using GCTA-COJO (64, 65) to independently select AD-associated SNPs (--cojo-slct) at 11p11.2. Briefly, GWAS summary statistics from the Lambert study (1), the Kunkle study (2), and the Jansen

study (3) were used in the analysis. Genotype data from 4410 individuals from the Alzheimer's Disease Genetics Consortium (ADGC, NG00032) (66) were used as a population reference. Because only SNPs with a genome-wide ($P < 5x10^{-8}$) or suggestive genome-wide significance ($P < 1x10^{-5}$) were used in this analysis, we set a loose *P* threshold for parameter (--cojo-p) for GWASs as $1x10^{-5}$.

RNA-seq data of PU.1 knockout B cells

The RNA-seq data of B cells from PU.1 (*SPI1*) knockout (KO) mice (n = 2) and wide-type controls (n = 2) were downloaded from the GEO with accession number GSE90094 (67). Fragments per kilobase of transcript per million mapped reads (FPKM) normalized expression for target genes were obtained, and expressional difference between PU.1 KO and control groups were analyzed by two-tailed Student's *t* test.

Statistical analysis and data visualization

The Locuszoom (http://locuszoom.org/) (68) was used to visualize GWAS results. Functional annotations for target genomic regions were visualized using the WashU epigenome browser (http://epigenomegateway.wustl.edu/) (69) or the Integrative Genomics Viewer (IGV) (70). Network was visualized by using the Cytoscape v3.7.1 (71). The comparisons of relative luciferase activities, mRNA levels, chromatin accessibility levels, or protein levels between two different groups were performed by using the PRISM software (GraphPad Software, Inc., La Jolla, CA, USA) with the Student's *t* test. A $P < 0.05$ was considered to be statistically significant. We performed Bonferroni correction for multiple testing for *P* values whenever this correction should be applied.

Data availability

Publically available data used in this study were listed in **Supplementary Table S1**. The ATAC-seq data generated in this study were available at GSA (https://ngdc.cncb.ac.cn/gsa/) under accession number HRA004084. Related results and codes were available at the Alzdata webserver (http://www.alzdata.org/file/11p11.2_related_data_and_scripts.rar).

Supplementary Figures

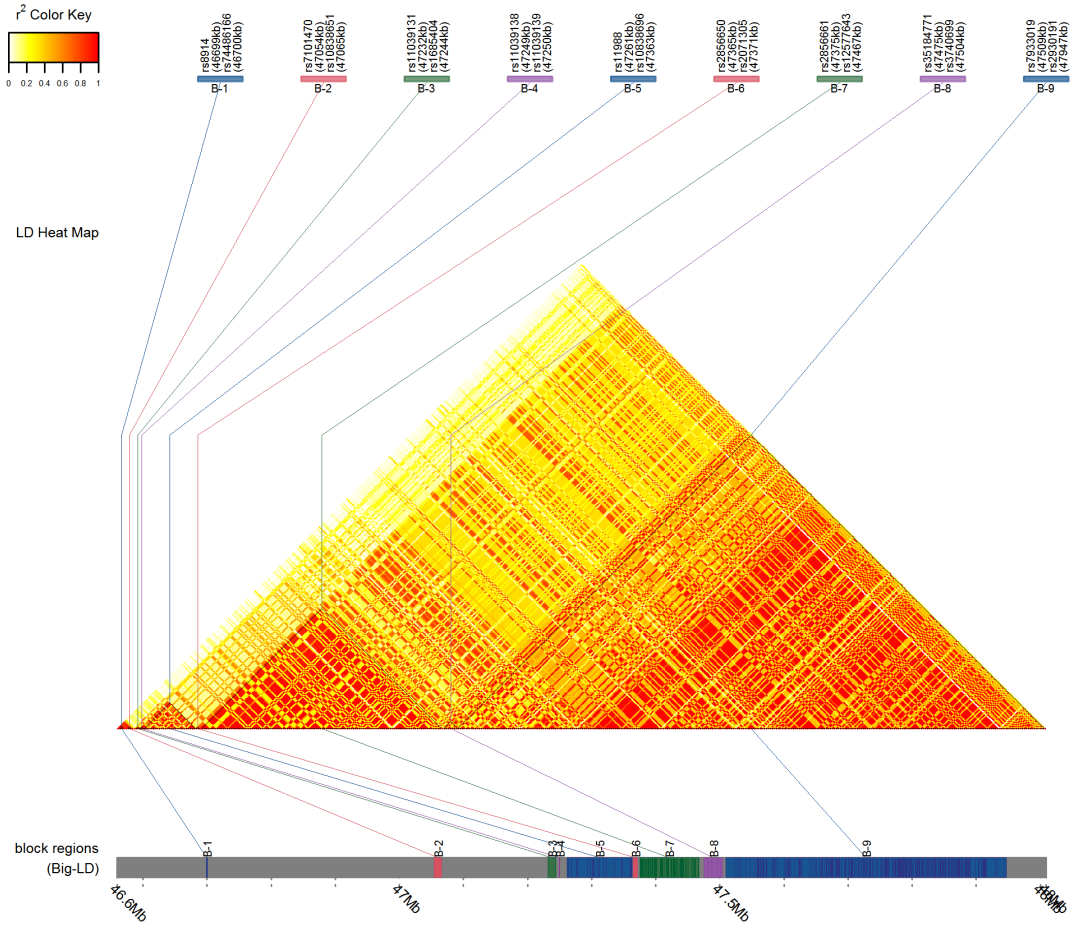

Figure S1. Linkage disequilibrium (LD) blocks at 11p11.2 detected by the bigLD. LD detection was performed by the bigLD (38) based on genotype data of 503 European individuals from the 1000 Genomes project (phase 3) (34). Each LD block (from B-1 to B-9) was represented by a different color.

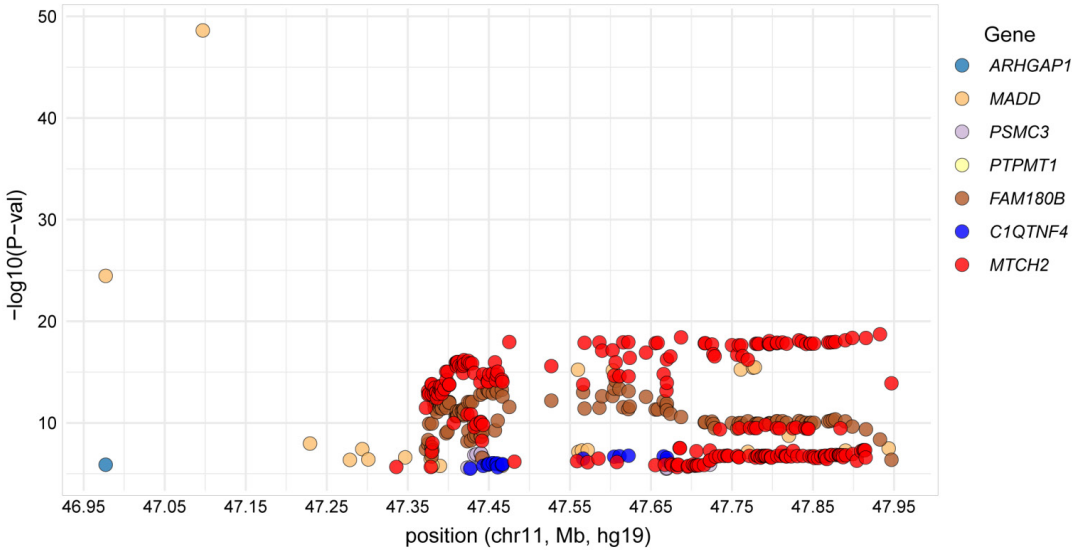

Figure S2. AD-associated SNPs at 11p11.2 were associated with expression of multiple genes in prefrontal cortex. The eQTL data were taken from the psychENCODE dataset (13). eQTLs with $P < 3.2 \times 10^{-6}$ (Bonferroni-corrected) were shown.

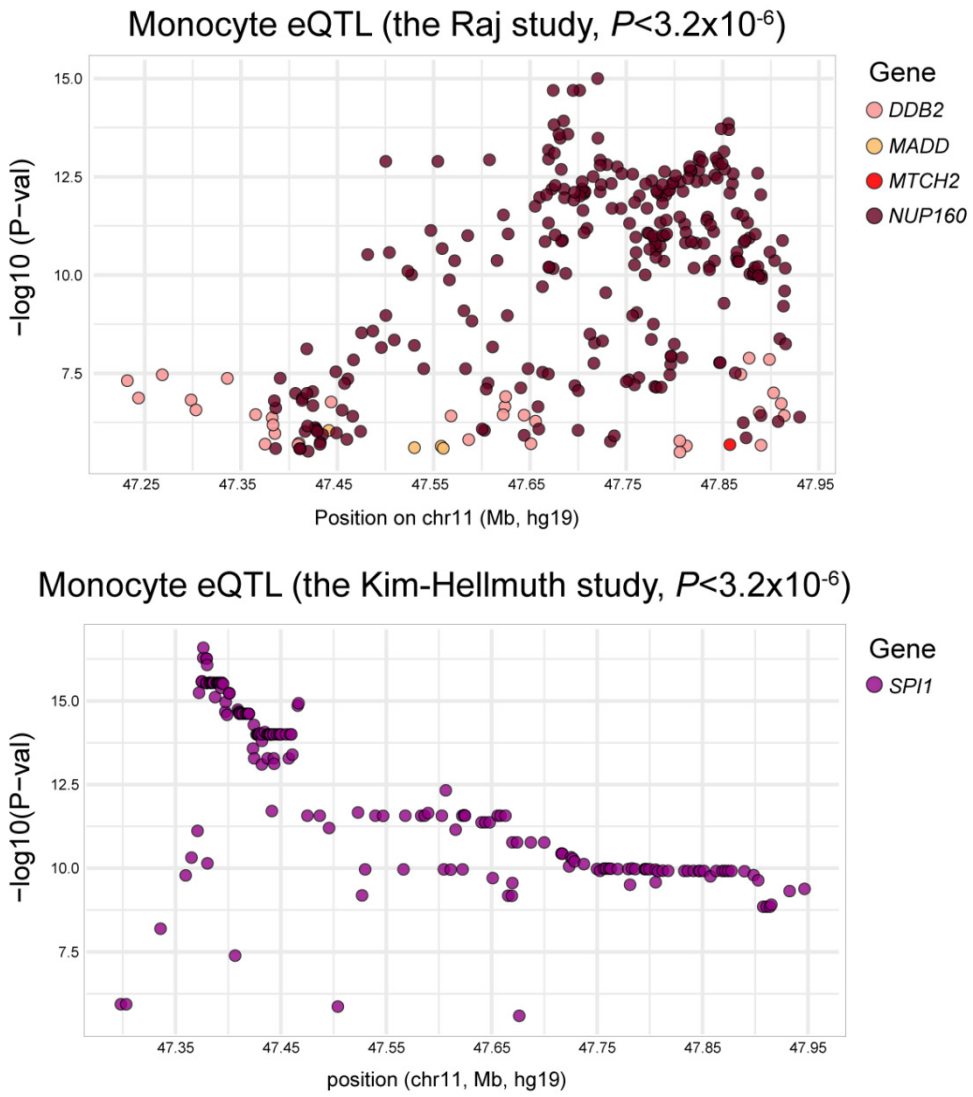

Figure S3. AD-associated SNPs at 11p11.2 were associated with expression of multiple genes in monocytes. The Raj study, monocyte eQTL dataset from Raj et al. (16); The Kim-Hellmuth study, monocyte eQTL dataset from Kim-Hellmuth et al. (15); eQTLs with $P < 3.2 \times 10^{-6}$ (Bonferroni-corrected) were shown.

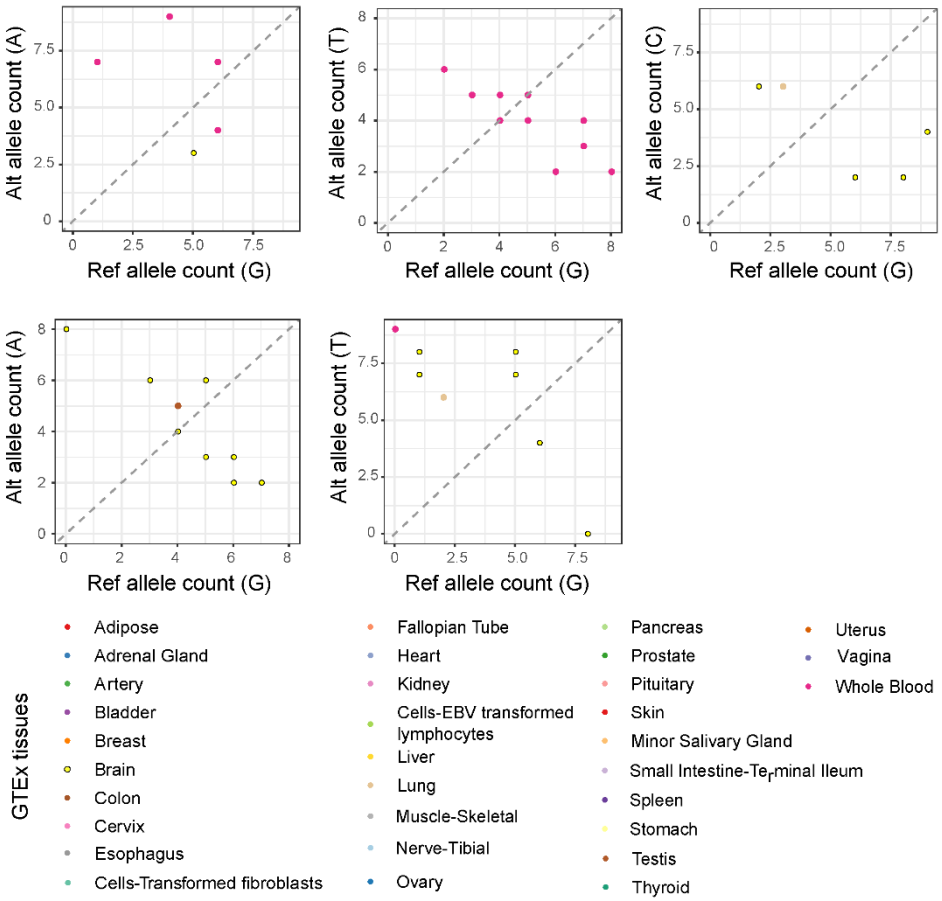

Figure S4. Potentially functional variants with a low capture rate by the GTEx ASE data (39, 40, 52)**.** Allele counts for the reference (Ref) allele and the alternative (Alt) allele were plotted for each fVar. Each dot represented an individual sample, and was colored by tissue. As the capture rates of these variants were inadequate (i.e., were detected in very few samples) to draw meaningful conclusions, ASE *P*-values were not calculated for these variants.

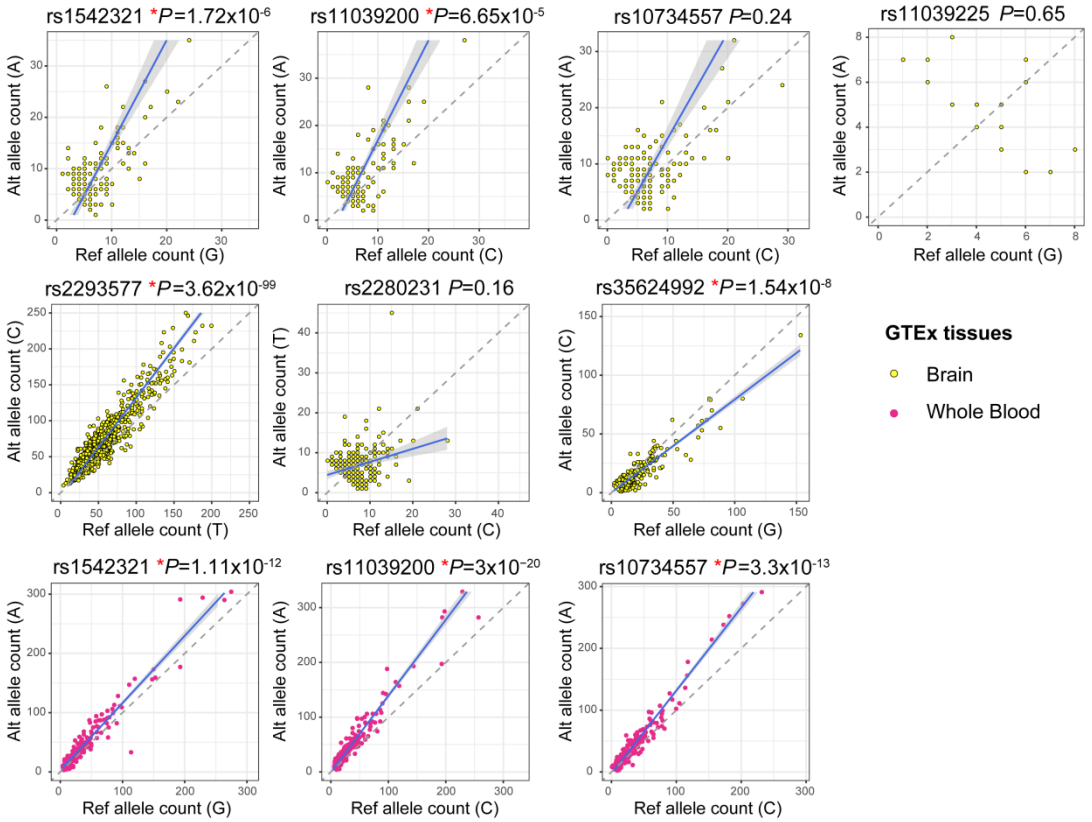

Figure S5. ASE of potentially functional variants in GTEx brain tissues and blood (39, 40, 52)**.** Allele counts for the reference (Ref) allele and the alternative (Alt) allele were plotted for each fVar. Each dot represented an individual sample, and was colored by tissue. *P* values were measured by binomial tests. *P* values < 0.005 after Bonferroni correction for the total number of comparisons (0.05/10) were marked with red asterisks (*).

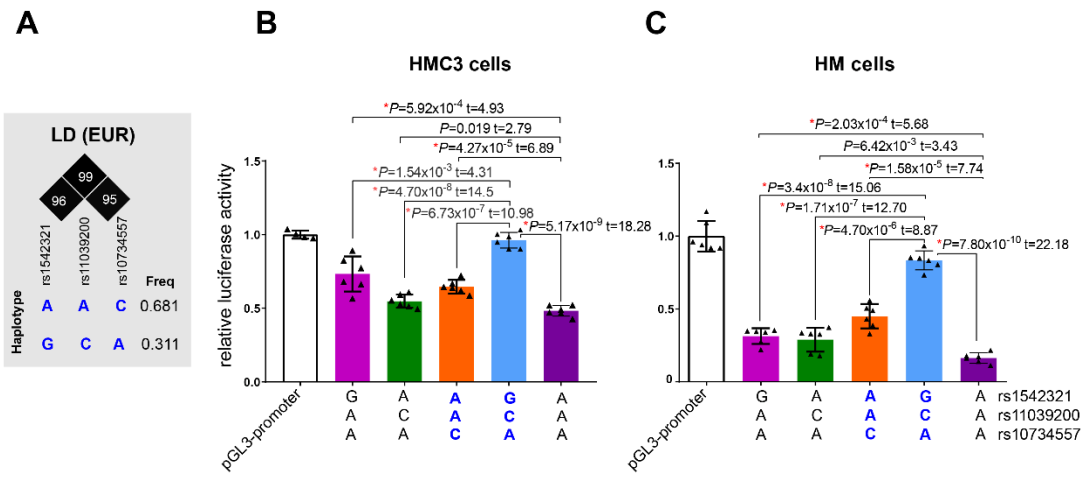

Figure S6. Three *SPI1* potential fVars showed individual and addictive effects in human microglia cell lines. (A) Linkage disequilibrium (LD) of three fVars in *SPI1* and their haplotypes. Result was performed by Haploview 4.1 (72) based on genotype data of 503 European individuals (EUR) from the 1000 Genomes project Phase 3 (34). r^2 was used for the LD color scheme. Haplotypes with frequencies > 0.1 in EUR were shown. (B-C) Dual-luciferase reporter assays for the three *SPI1* potential fVars and their common haplotypes in EUR using human microglia cell lines HMC3 (B) and HM (C) cells. Two common haplotypes of the three *SPI1* fVars in EUR were marked in blue. Shown results were representative of three independent experiments with similar results. Bars represent mean ± SD (n = 6 biological replicates for HMC3 cells and HM cells, respectively). *P* values were calculated by two-sided Student's *t* test, together with the t-statistics (the degrees of freedom (df) = 10). *P* values < 0.003 after Bonferroni correction for the total number of comparisons (0.05/16) were marked with red asterisks (*).

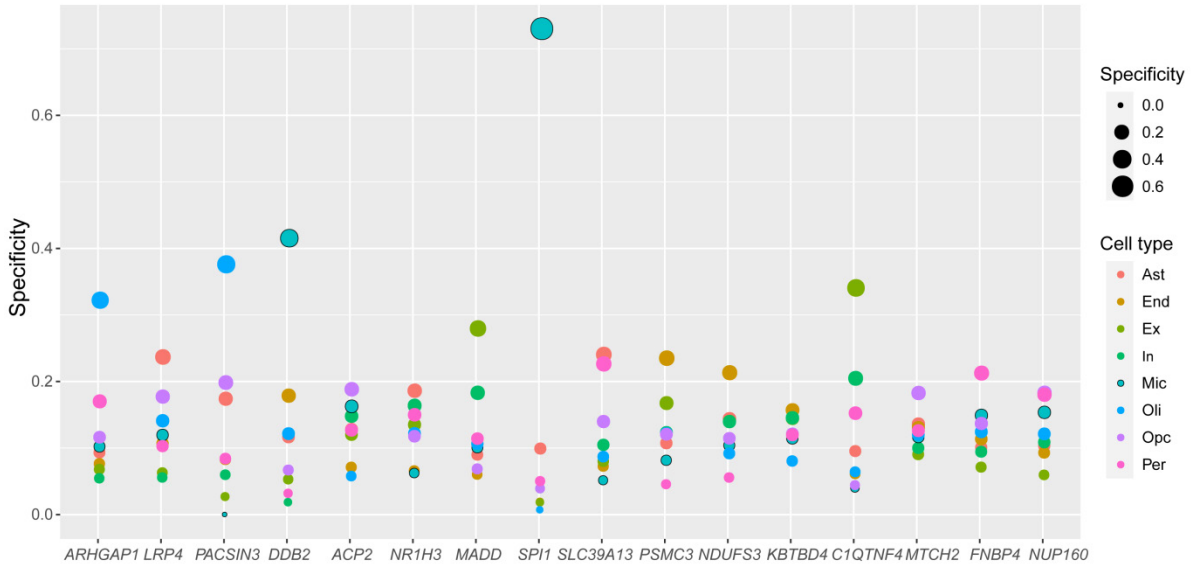

Figure S7. Cell-type expression specificity of eGenes of potentially functional variants at 11p11.2. Cell-type specificity for each gene was calculated using R package EWCE (56), based on single cell RNA-seq data from Mathys et al. (23) (frontal cortex, n = 48). Ast: astrocytes; End: endothelial cells; Ex: excitatory neurons; In: inhibitory neurons; Mic: microglia; Oli: oligodendrocytes; OPC: oligodendrocytes precursor cells; Per: pericytes.

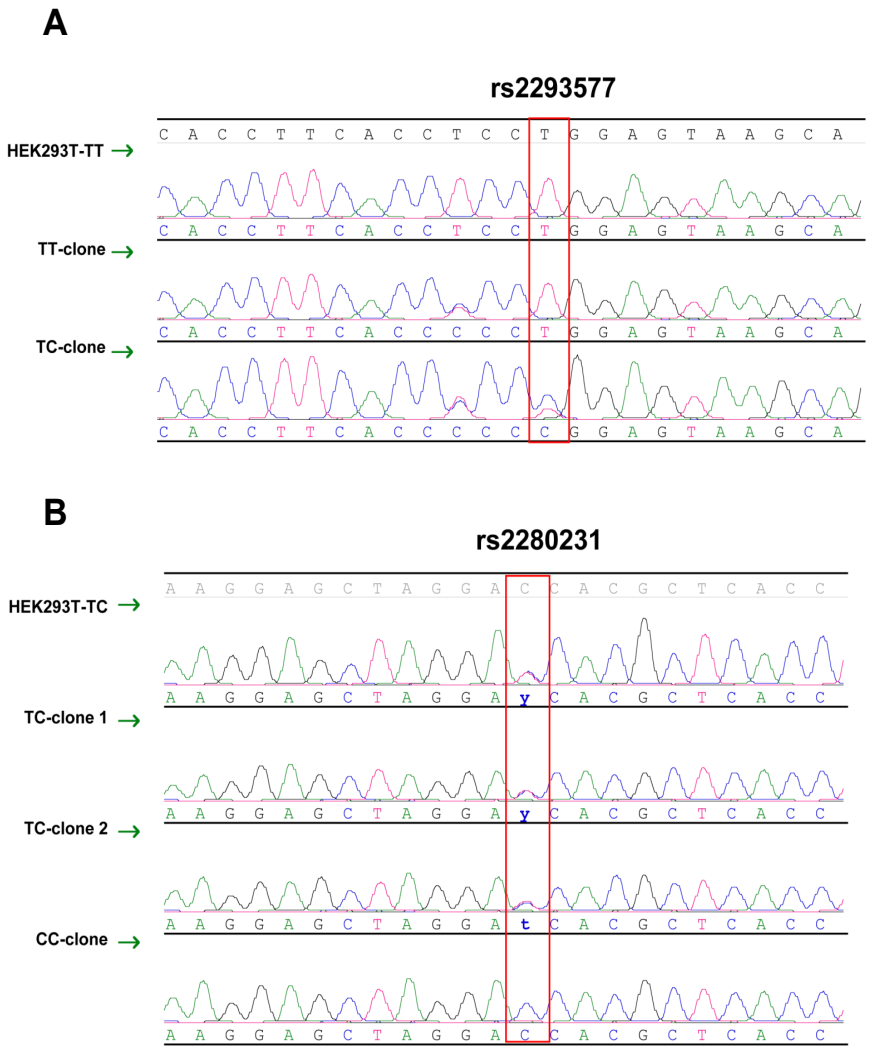

Figure S8. Sanger sequencing validated a successful base-editing of rs2293577 (A) and rs2280231 (B) in HEK293T cells. The original HEK293T cell line has a genotype TT for rs2293577 (HEK293T-TT) and TC for rs2280231 (HEK293T-TC). We obtained HEK293T cell clones with genotypes TT (TT-clone) and TC (TC-clone) for rs2293577, and with genotypes TC (TC-clone 1 and TC-clone 2) and CC (CC-clone) for rs2280231, respectively.

Figure S9. GWAS association signals before and after the conditional analysis at 11p11.2. Each panel showed the original GWAS association *P* values (green dots) and *P* values conditioned (orange dots) on the SNP selected by GCTA-COJO (64, 65) for each GWAS (the Lambert study (1), the Kunkle study (2), and the Jansen study (3)).

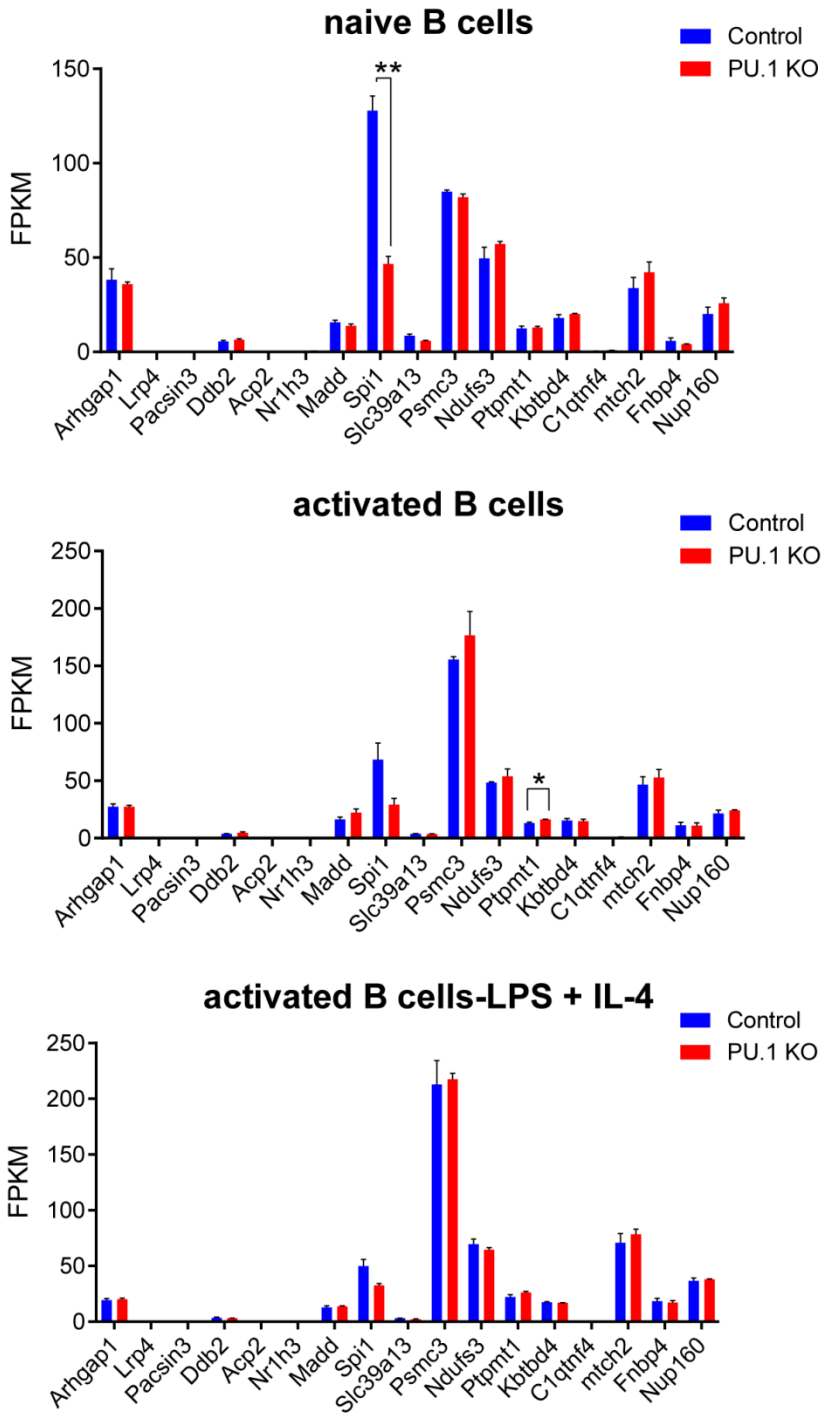

Figure S10. Expressional changes of eGenes at 11p11.2 in B cells isolated from PU.1 (*SPI1*) knockout mice (67). The *P* value was calculated by the Student's *t* test (two-tailed) to evaluate expressional difference of each gene between control and PU.1 KO cells. FPKM, fragments per kilobase of transcript per million mapped reads. LPS, lipopolysaccharide; Control, wide-type mice; PU.1 KO, PU.1 knockout mice; Bars represent mean ± SD. *, $P < 0.05$; **, $P < 0.01$.

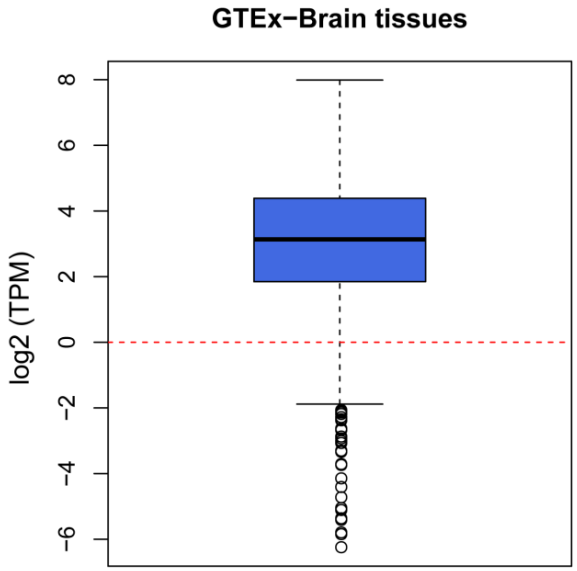

Figure S11. Mean expression of TFs in GTEx brain tissues. Original RNA-seq data of bulk brain tissues were obtained from the GTEx (39, 40), and were normalized by transcription per million (TPM).

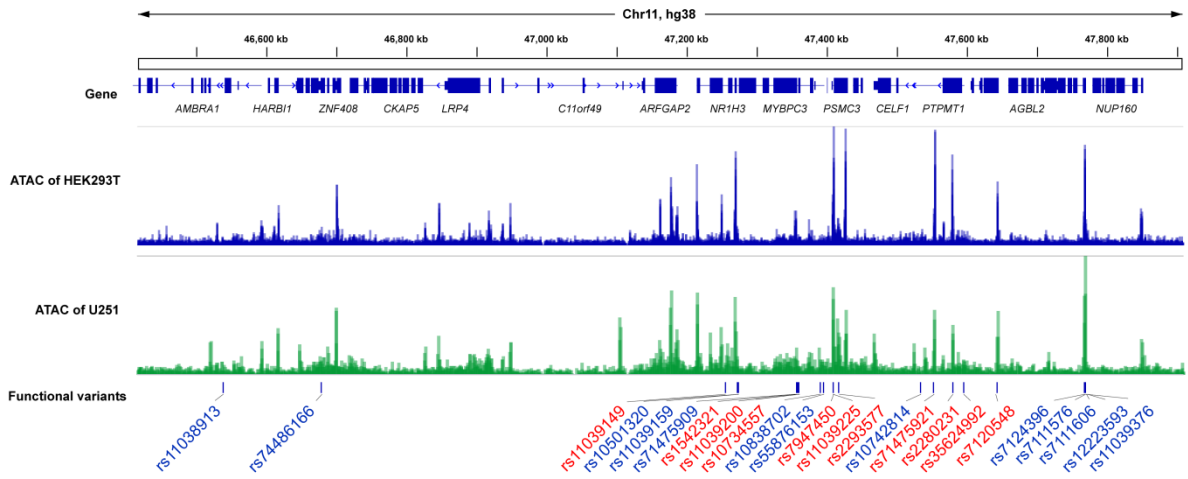

Figure S12. Open chromatin peaks in HEK293T and U251 cells at 11p11.2. Potentially functional variants (fVars) tested by the dual-luciferase reporter assays were marked in red.

Supplementary References

1. Lambert JC, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C, et al. (2013): Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet* 45:1452-1458.
2. Kunkle BW, Grenier-Boley B, Sims R, Bis JC, Damotte V, Naj AC, et al. (2019): Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Abeta, tau, immunity and lipid processing. *Nat Genet* 51:414-430.
3. Jansen IE, Savage JE, Watanabe K, Bryois J, Williams DM, Steinberg S, et al. (2019): Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nat Genet* 51:404-413.
4. Encode Project Consortium (2012): An integrated encyclopedia of DNA elements in the human genome. *Nature* 489:57-74.
5. Davis CA, Hitz BC, Sloan CA, Chan ET, Davidson JM, Gabdank I, et al. (2018): The Encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Res* 46:D794-D801.
6. Fullard JF, Hauberg ME, Bendl J, Egervari G, Cirnaru MD, Reach SM, et al. (2018): An atlas of chromatin accessibility in the adult human brain. *Genome Res* 28:1243-1252.
7. Novakovic B, Habibi E, Wang SY, Arts RJW, Davar R, Megchelenbrink W, et al. (2016): beta-glucan reverses the epigenetic state of LPS-induced immunological tolerance. *Cell* 167:1354-1368 e1314.
8. Nott A, Holtman IR, Coufal NG, Schlachetzki JCM, Yu M, Hu R, et al. (2019): Brain cell type-specific enhancer-promoter interactome maps and disease-risk association. *Science* 366:1134-1139.
9. Corces MR, Shcherbina A, Kundu S, Gloudemans MJ, Fresard L, Granja JM, et al. (2020): Single-cell epigenomic analyses implicate candidate causal variants at inherited risk loci for Alzheimer's and Parkinson's diseases. *Nat Genet* 52:1158-1168.
10. Song M, Yang X, Ren X, Maliskova L, Li B, Jones IR, et al. (2019): Mapping cis-regulatory chromatin contacts in neural cells links neuropsychiatric disorder risk variants to target genes. *Nat Genet* 51:1252-1262.
11. Zuo C, Shin S, Keles S (2015): atSNP: transcription factor binding affinity testing for regulatory SNP detection. *Bioinformatics* 31:3353-3355.
12. Qi T, Wu Y, Zeng J, Zhang F, Xue A, Jiang L, et al. (2018): Identifying gene targets for brain-related traits using transcriptomic and methylomic data from blood. *Nat Commun* 9:2282.
13. Wang D, Liu S, Warrell J, Won H, Shi X, Navarro FCP, et al. (2018): Comprehensive functional genomic resource and integrative model for the human brain. *Science* 362:eaat8464.
14. Lopes KP, Snijders GJL, Humphrey J, Allan A, Sneeboer MAM, Navarro E, et al. (2022): Genetic analysis of the human microglial transcriptome across brain regions, aging and disease pathologies. *Nat Genet* 54:4-17.
15. Kim-Hellmuth S, Bechheim M, Putz B, Mohammadi P, Nedelec Y, Giangreco N, et al. (2017): Genetic regulatory effects modified by immune activation contribute to autoimmune disease associations. *Nat Commun* 8:266.
16. Raj T, Rothamel K, Mostafavi S, Ye C, Lee MN, Replogle JM, et al. (2014): Polarization of the effects of autoimmune and neurodegenerative risk alleles in leukocytes. *Science* 344:519-523.
17. Xu M, Zhang DF, Luo R, Wu Y, Zhou H, Kong LL, et al. (2018): 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* 14:215-229.
18. Niculescu AB, Le-Niculescu H (2010): Convergent Functional Genomics: what we have learned and can learn about genes, pathways, and mechanisms. *Neuropsychopharmacology* 35:355-356.
19. Wu Y, Yao YG, Luo XJ (2017): SZDB: a database for schizophrenia genetic research. *Schizophr Bull* 43:459-471.
20. Bi R, Zhang W, Zhang DF, Xu M, Fan Y, Hu QX, et al. (2018): Genetic association of the cytochrome c oxidase-related genes with Alzheimer's disease in Han Chinese. *Neuropsychopharmacology* 43:2264-2276.
21. Barbeira AN, Dickinson SP, Bonazzola R, Zheng J, Wheeler HE, Torres JM, et al. (2018): Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *Nat Commun* 9:1825.
22. Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE, et al. (2016): Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet* 48:481-487.
23. Mathys H, Davila-Velderrain J, Peng Z, Gao F, Mohammadi S, Young JZ, et al. (2019): Single-cell

transcriptomic analysis of Alzheimer's disease. *Nature* 570:332-337.
24. Marzi SJ, Leung SK, Ribarska T, Hannon E, Smith AR, Pishva E, et al. (2018): A histone acetylome-wide association study of Alzheimer's disease identifies disease-associated H3K27ac differences in the entorhinal cortex. *Nat Neurosci* 21:1618-1627.
25. Johnson ECB, Dammer EB, Duong DM, Ping L, Zhou M, Yin L, et al. (2020): 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* 26:769-780.
26. Huang KL, Marcora E, Pimenova AA, Di Narzo AF, Kapoor M, Jin SC, et al. (2017): A common haplotype lowers PU.1 expression in myeloid cells and delays onset of Alzheimer's disease. *Nat Neurosci* 20:1052-1061.
27. Deming Y, Li Z, Kapoor M, Harari O, Del-Aguila JL, Black K, et al. (2017): Genome-wide association study identifies four novel loci associated with Alzheimer's endophenotypes and disease modifiers. *Acta Neuropathol* 133:839-856.
28. Nievergelt CM, Maihofer AX, Klengel T, Atkinson EG, Chen CY, Choi KW, et al. (2019): International meta-analysis of PTSD genome-wide association studies identifies sex- and ancestry-specific genetic risk loci. *Nat Commun* 10:4558.
29. Pardinas AF, Holmans P, Pocklington AJ, Escott-Price V, Ripke S, Carrera N, et al. (2018): Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat Genet* 50:381-389.
30. Liu J, Li S, Li X, Li W, Yang Y, Guo S, et al. (2021): Genome-wide association study followed by trans-ancestry meta-analysis identify 17 new risk loci for schizophrenia. *BMC Med* 19:177.
31. Hammond RK, Pahl MC, Su C, Cousminer DL, Leonard ME, Lu S, et al. (2021): Biological constraints on GWAS SNPs at suggestive significance thresholds reveal additional BMI loci. *Elife* 10.
32. Wang X, Tucker NR, Rizki G, Mills R, Krijger PH, de Wit E, et al. (2016): Discovery and validation of sub-threshold genome-wide association study loci using epigenomic signatures. *Elife* 5:e10557.
33. Andrews SJ, Fulton-Howard B, Goate A (2020): Interpretation of risk loci from genome-wide association studies of Alzheimer's disease. *Lancet Neurol* 19:326-335.
34. The 1000 Genomes Project Consortium (2015): A global reference for human genetic variation. *Nature* 526:68-74.
35. Wang K, Li M, Hakonarson H (2010): ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 38:e164.
36. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M (2019): CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res* 47:D886-D894.
37. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ (2015): Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 4:7.
38. Kim SA, Brossard M, Roshandel D, Paterson AD, Bull SB, Yoo YJ (2019): gpart: human genome partitioning and visualization of high-density SNP data by identifying haplotype blocks. *Bioinformatics* 35:4419-4421.
39. GTEx Consortium (2017): Genetic effects on gene expression across human tissues. *Nature* 550:204-213.
40. GTEx Consortium (2013): The Genotype-Tissue Expression (GTEx) project. *Nat Genet* 45:580-585.
41. Fromer M, Roussos P, Sieberts SK, Johnson JS, Kavanagh DH, Perumal TM, et al. (2016): Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nat Neurosci* 19:1442-1453.
42. Ng B, White CC, Klein HU, Sieberts SK, McCabe C, Patrick E, et al. (2017): An xQTL map integrates the genetic architecture of the human brain's transcriptome and epigenome. *Nat Neurosci* 20:1418-1426.
43. Battle A, Mostafavi S, Zhu X, Potash JB, Weissman MM, McCormick C, et al. (2014): Characterizing the genetic basis of transcriptome diversity through RNA-sequencing of 922 individuals. *Genome Res* 24:14-24.
44. Lloyd-Jones LR, Holloway A, McRae A, Yang J, Small K, Zhao J, et al. (2017): The genetic architecture of gene expression in peripheral blood. *Am J Hum Genet* 100:228-237.
45. Gamazon ER, Wheeler HE, Shah KP, Mozaffari SV, Aquino-Michaels K, Carroll RJ, et al. (2015): A gene-based association method for mapping traits using reference transcriptome data. *Nat Genet* 47:1091-1098.
46. Guenther MG, Levine SS, Boyer LA, Jaenisch R, Young RA (2007): A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* 130:77-88.
47. Shlyueva D, Stampfel G, Stark A (2014): Transcriptional enhancers: from properties to

genome-wide predictions. *Nat Rev Genet* 15:272-286.
48. Cheung P, Lau P (2005): Epigenetic regulation by histone methylation and histone variants. *Mol Endocrinol* 19:563-573.
49. Brovkina MV, Duffie R, Burtis AEC, Clowney EJ (2021): Fruitless decommissions regulatory elements to implement cell-type-specific neuronal masculinization. *PLoS Genet* 17:e1009338.
50. Bailey TL, Elkan C (1994): Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol* 2:28-36.
51. Quinlan AR, Hall IM (2010): BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26:841-842.
52. GTEx Consortium (2015): Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* 348:648-660.
53. Zhang DF, Fan Y, Xu M, Wang G, Wang D, Li J, et al. (2019): Complement C7 is a novel risk gene for Alzheimer's disease in Han Chinese. *Natl Sci Rev* 6:257-274.
54. Rustenhoven J, Smith AM, Smyth LC, Jansson D, Scotter EL, Swanson MEV, et al. (2018): PU.1 regulates Alzheimer's disease-associated genes in primary human microglia. *Mol Neurodegener* 13:44.
55. Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, et al. (2015): Proteomics. Tissue-based map of the human proteome. *Science* 347:1260419.
56. Skene NG, Grant SG (2016): Identification of vulnerable cell types in major brain disorders using single cell transcriptomes and expression weighted cell type enrichment. *Front Neurosci* 10:16.
57. Shen B, Zhang W, Zhang J, Zhou J, Wang J, Chen L, et al. (2014): Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. *Nat Methods* 11:399-402.
58. Gaudelli NM, Komor AC, Rees HA, Packer MS, Badran AH, Bryson DI, et al. (2017): Programmable base editing of A·T to G·C in genomic DNA without DNA cleavage. *Nature* 551:464-471.
59. Liao Y, Smyth GK, Shi W (2014): featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30:923-930.
60. McCarthy DJ, Campbell KR, Lun AT, Wills QF (2017): Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. *Bioinformatics* 33:1179-1186.
61. Ramirez F, Ryan DP, Gruning B, Bhardwaj V, Kilpert F, Richter AS, et al. (2016): deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* 44:W160-165.
62. Zhang DF, Li J, Wu H, Cui Y, Bi R, Zhou HJ, et al. (2016): CFH variants affect structural and functional brain changes and genetic risk of Alzheimer's disease. *Neuropsychopharmacology* 41:1034-1045.
63. Xiang Q, Bi R, Xu M, Zhang DF, Tan L, Zhang C, et al. (2017): Rare genetic variants of the transthyretin gene are associated with Alzheimer's disease in Han Chinese. *Mol Neurobiol* 54:5192-5200.
64. Yang J, Ferreira T, Morris AP, Medland SE, Genetic Investigation of ATC, Replication DIG, et al. (2012): Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet* 44:369-375, S361-363.
65. Yang J, Lee SH, Goddard ME, Visscher PM (2011): GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* 88:76-82.
66. Naj AC, Jun G, Beecham GW, Wang LS, Vardarajan BN, Buros J, et al. (2011): Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. *Nat Genet* 43:436-441.
67. Willis SN, Tellier J, Liao Y, Trezise S, Light A, O'Donnell K, et al. (2017): Environmental sensing by mature B cells is controlled by the transcription factors PU.1 and SpiB. *Nat Commun* 8:1426.
68. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. (2010): LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* 26:2336-2337.
69. Li D, Hsu S, Purushotham D, Sears RL, Wang T (2019): WashU Epigenome Browser update 2019. *Nucleic Acids Res* 47:W158-W165.
70. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, et al. (2011): Integrative genomics viewer. *Nat Biotechnol* 29:24-26.
71. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. (2003): Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 13:2498-2504.
72. Barrett JC, Fry B, Maller J, Daly MJ (2005): Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21:263-265.