Archival Report

Biological Psychiatry

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

Min Xu, Qianjin Liu, Rui Bi, Yu Li, Hongli Li, Wei-Bo Kang, Zhongjiang Yan, Quanzhen Zheng, Chunli Sun, Maosen Ye, Bo-Lin Xiang, Xiong-Jian Luo, Ming Li, Deng-Feng Zhang, and Yong-Gang Yao

ABSTRACT

BACKGROUND: Genome-wide association studies have identified dozens of genetic risk loci for Alzheimer's disease (AD), yet the underlying causal variants and biological mechanisms remain elusive, especially for loci with complex linkage disequilibrium and regulation.

METHODS: To fully untangle the causal signal at a single locus, we performed a functional genomic study of 11p11.2 (the *CELF1/SPI1* locus). Genome-wide association study signals at 11p11.2 were integrated with datasets of histone modification, open chromatin, and transcription factor binding to distill potentially functional variants (fVars). Their allelic regulatory activities were confirmed by allele imbalance, reporter assays, and base editing. Expressional quantitative trait loci and chromatin interaction data were incorporated to assign target genes to fVars. The relevance of these genes to AD was assessed by convergent functional genomics using bulk brain and single-cell transcriptomic, epigenomic, and proteomic datasets of patients with AD and control individuals, followed by cellular assays.

RESULTS: We found that 24 potential fVars, rather than a single variant, were responsible for the risk of 11p11.2. These fVars modulated transcription factor binding and regulated multiple genes by long-range chromatin interactions. Besides *SPI1*, convergent evidence indicated that 6 target genes (*MTCH2*, *ACP2*, *NDUFS3*, *PSMC3*, *C1QTNF4*, and *MADD*) of fVars were likely to be involved in AD development. Disruption of each gene led to cellular amyloid-β and phosphorylated tau changes, supporting the existence of multiple likely causal genes at 11p11.2.

CONCLUSIONS: Multiple variants and genes at 11p11.2 may contribute to AD risk. This finding provides new insights into the mechanistic and therapeutic challenges of AD.

https://doi.org/10.1016/j.biopsych.2023.05.020

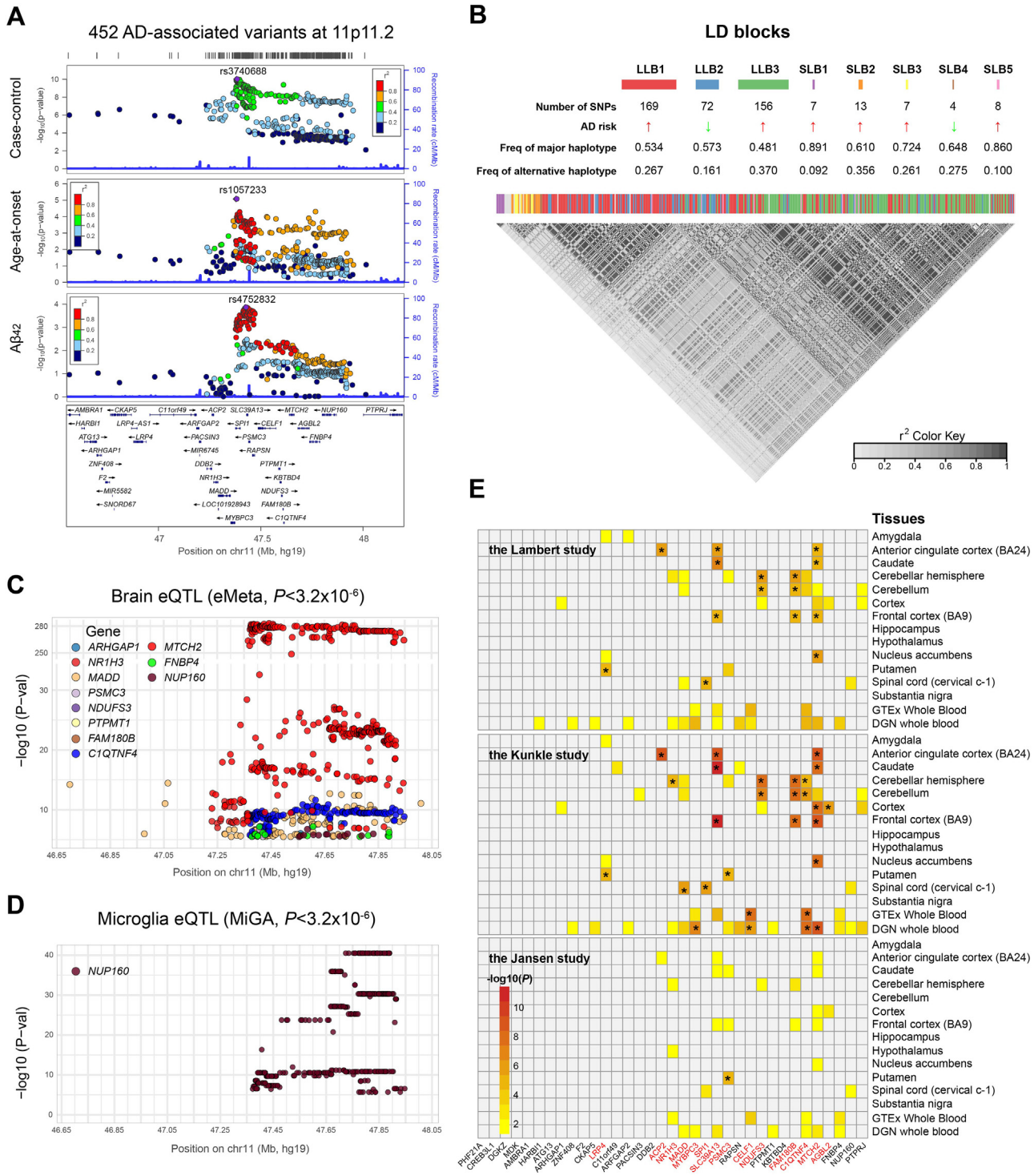
Large-cohort genome-wide association studies (GWASs) have identified dozens of risk loci for Alzheimer's disease (AD) (1–5), thereby providing biological insights and potential targets for AD. However, it has been challenging to distinguish the biologically causative variants and genes from statistical associations (6), especially for loci with complex linkage disequilibrium (LD) (1) such as 19q13.2 (the *APOE* locus), 11p11.2 (the *CELF1/SPI1* locus), and 17p13.2 (the *SCIMP* locus). Among them, the 11p11.2 locus has the most complicated LD and contains several long AD-associated haplotypes with hundreds of single nucleotide polymorphisms (SNPs) and numerous coexpressed genes (7). Previous GWASs have repeatedly linked 11p11.2 with AD risk (2–5,8,9), but the exact causal variants and genes at this locus remain controversial (2,5,10–14). Transcription factor (TF) PU.1 encoded by *SPI1* at 11p11.2 has been reported to account for AD risk by modulating the expression of immune-related genes and microglia function (9,15,16). However, transcriptomic and functional studies have suggested other potentially causal genes, such as *MADD* (10–12,17), *PSMC3* (2,11), and *MTCH2* (2,12). The highly correlated SNPs and coexpressed genes (7,11) at 11p11.2 complicate the identification of true causative variants and genes. A conceptual update with an unbiased fine-mapping strategy followed by systematic functional investigations is urgently required for the decipherment of such complex loci.

In this study, multilevel omics data, especially those not confounded by LD and gene coexpression, were integrated with AD GWASs to identify potentially functional variants (fVars) and their target genes that are likely involved in AD pathogenesis. We prioritized 24 fVars from 452 AD-associated SNPs at 11p11.2. As exemplified by base editing, fVars at 11p11.2 were able to regulate multiple nearby or distal genes via long-range chromatin contacts. In addition to the established *SPI1*, 6

SEE COMMENTARY ON PAGE 692

ISSN: 0006-3223

target genes of potential fVars were assigned as likely AD causal genes at 11p11.2. The convergent functional genomics (CFG) and cellular assays suggested that these genes could modulate AD-related molecular phenotypes, indicating that multiple genes, rather than a single gene, are responsible for the risk of 11p11.2. Our study offered novel conceptual insight and a new framework for understanding the molecular mechanisms underlying complex GWAS loci.



METHODS AND MATERIALS

Detailed data resources, analyses, materials, and methods are described in the Supplemental Materials and Methods. Briefly, we defined a regulatory fVar as a genomic variant capable of modulating gene expression by affecting the binding of certain TFs to the active regulatory element in which the variant was located. If a gene was expressionally associated (i.e., an eQTL gene of the fVar) and physically interacted with the fVar, it was defined as a target gene of this fVar. The CFG strategy (18,19) was applied to prioritize likely causal genes from target genes of fVars. Multiomics data of AD-related tissues/cells were integrated with AD GWASs (2–4) to distill potential fVars and likely causal genes. In vitro cellular assays were performed to confirm the allelic effects of potential fVars and the effects of likely causal genes on AD-related molecular phenotypes. Publicly available data utilized in this study, newly obtained data, and codes are listed in Table S1.

RESULTS

Multiple SNPs and Genes Were Associated With AD Risk at 11p11.2

From the 3 large-scale AD GWASs [Lambert *et al.* (4), Kunkle *et al.* (2), and Jansen *et al.* (3)], we retrieved 452 SNPs at 11p11.2 suggestively associated with AD risk ($p < 1 \times 10^{-5}$), with consistent direction of effects (GWAS beta) across the 3 studies (Figure 1A; Table S2). In the Kunkle *et al.* study (2), 168 SNPs reached genome-wide significance ($p < 5 \times 10^{-8}$) (Table S2). SNPs at 11p11.2 were also nominally associated with age at onset of AD (9) and amyloid-β 42 ($A\beta_{42}$) level in the cerebrospinal fluid (20) (Figure 1A), suggesting robust involvement of 11p11.2 in AD. LD-based clumping analysis using Plink version 1.9 (21) showed that these AD-associated SNPs were in 3 long LD blocks ($r^2 > 0.8$ and $n_{SNPs} > 50$) and 5 relatively short LD blocks ($r^2 > 0.8$ and $n_{SNPs} < 20$) (Figure 1B). Similar LD blocks (Figure S1) were identified using an independent algorithm, bigLD (22). Haplotypes constituted by minor alleles of SNPs in long LD block 2 and short LD block 4 were associated with reduced AD risk, while those of other LD blocks were associated with increased AD risk (Figure 1B; Table S2), suggesting that multiple signals may be responsible for AD at 11p11.2.

These AD-associated SNPs at 11p11.2 may affect the expression of multiple genes. In the bulk brain expressional quantitative trait loci (eQTL) dataset eMeta (N = 1194) (6), expression of 11 genes was significantly associated with 11p11.2 SNPs ($p < 3.2 \times 10^{-6}$) (Figure 1C), and 7 could be replicated in the psychENCODE dataset (N = 1387) (23) (Figure S2). We also explored eQTL of microglia (24) and monocytes (25,26), which were actively involved in AD (9,11). *NUP160* was the only gene associated with 11p11.2 in microglia (the MiGA [Microglia Genomic Atlas] dataset, N = 216) (24) (Figure 1D) ($p < 3.2 \times 10^{-6}$). Except for the previously reported *SPI1* (9), expression of *NUP160*, *DDB2*, *MADD*, and *MTCH2* were also associated with 11p11.2 in monocytes [$p < 3.2 \times 10^{-6}$, the Raj dataset, N = 461 (25) and the Kim-Hellmuth dataset, N = 134 (26)] (Figure S3). Different top eQTL genes in bulk brain tissues (*MTCH2*), microglia (*NUP160*), and monocytes (*SPI1*) eQTL datasets suggested cell type–specific effects of genes at 11p11.2.

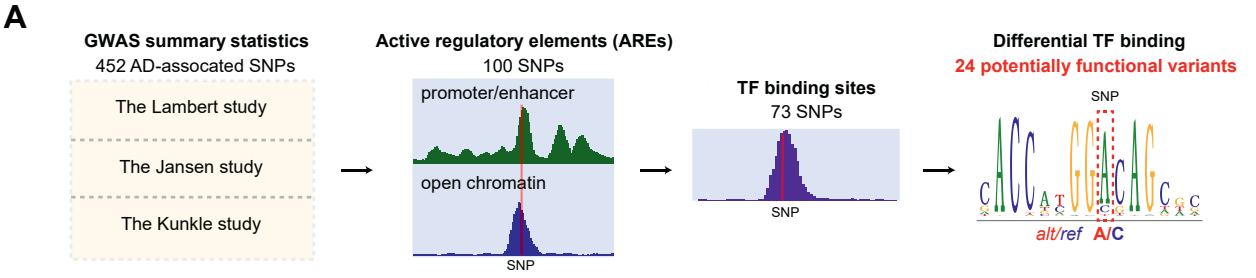
In addition to the eQTL correlation analysis, we performed a transcriptome-wide association study to infer the potentially causal genes at 11p11.2 by integrating the AD GWASs mentioned previously (2–4) with eQTL data of GTEx (Genotype-Tissue Expression) brain tissues (27,28) and peripheral blood (29,30). Two algorithms, MetaXcan (31) and the summary data–based Mendelian randomization (SMR) (32), were applied. Fourteen likely causal genes were predicted by MetaXcan (31) (Bonferroni-corrected $p < .05/35$ genes/14 tissues = 1×10^{-4}) (Figure 1E), among which *SLC39A13*, *CELF1*, and *MTCH2* were also predicted by SMR (32) ($p_{SMR} < 1 \times 10^{-4}$ and $p_{HEIDI} > .05$) (Table S3). We further retrieved fine-mapping data from a colocalization-based study (24) in which microglia, monocytes, and bulk brain eQTL datasets [(24) and references therein] were integrated with AD GWASs (2–4). Colocalization analyses with eQTLs from bulk brain tissues (*ACP2*, *C1QTNF4*, and *MTCH2*) and monocytes (*MYBPC3*, *SPI1*, and *C1QTNF4*) identified multiple genes colocalized with AD GWASs (PP.H4 > 0.7), while no genes were identified in microglia (Table S4). Collectively, integrative analyses based on different eQTL datasets or algorithms distilled different genes at 11p11.2, and 7 genes (*ACP2*, *MYBPC3*, *SPI1*, *SLC39A13*, *CELF1*, *MTCH2*, and *C1QTNF4*) were prioritized by at least 2 algorithms. It seems that prioritizing highly correlated SNPs and coexpressed genes using statistical fine-mapping approaches based on LD and association statistics can prove challenging, as suggested by the results obtained in this study and previous studies (33,34). Alternatively, these results may indicate the presence of multiple causal variants and genes at 11p11.2.

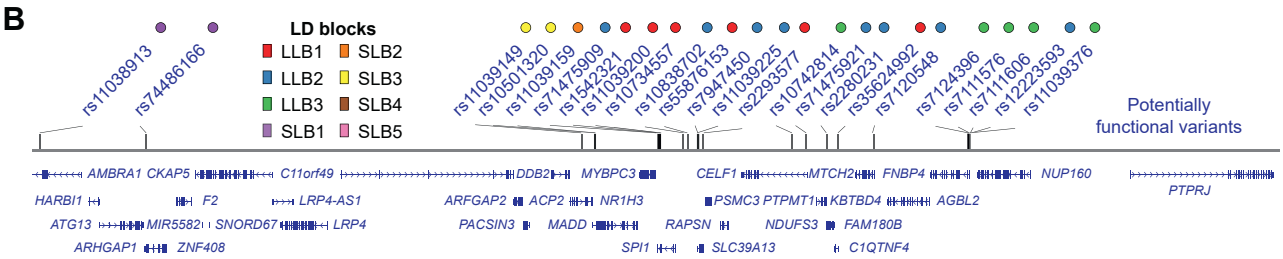
The Functional Genomic Strategy Prioritized 24 Potential fVars at 11p11.2

Functional fine mapping, rather than statistical association, may help to identify potentially causal variants with biological functions (11,35–38). Thus, we integrated bulk and single-cell multiomics data of brain tissues and data of monocytes to

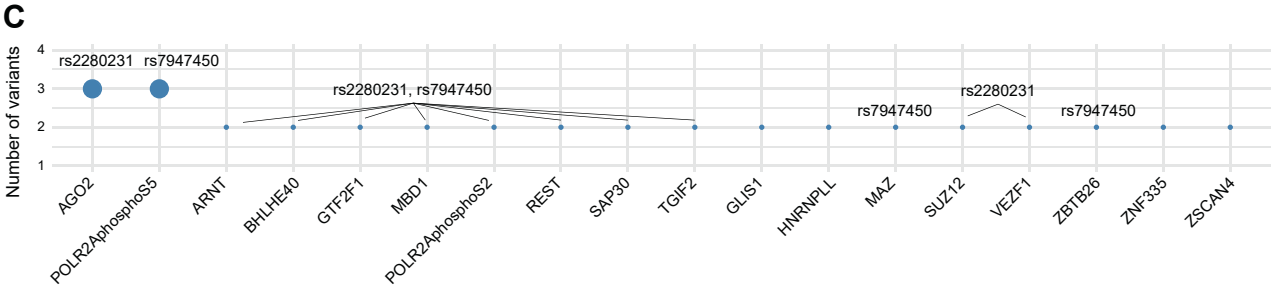
Figure 1. Complex LD structure impedes the identification of functional variants and causal genes underlying AD risk locus 11p11.2. **(A)** A large number of SNPs at 11p11.2 were associated with AD disease states, age at onset, and cerebrospinal fluid $A\beta_{42}$ level. Regional association plots were generated using genome-wide association study data from Kunkle *et al.* (2) (case-control), Huang *et al.* (9) (age at onset), and Deming *et al.* (20) ($A\beta_{42}$), respectively. **(B)** LD blocks at 11p11.2 based on genotype data of European individuals from the 1000 Genomes Project (phase 3) (44). LB, LD block ($r^2 > 0.8$); LLB ($r^2 > 0.8$ and $n_{SNPs} > 50$); SLB ($r^2 > 0.8$ and $n_{SNPs} < 20$). **(C, D)** SNPs at 11p11.2 were associated with expression of multiple genes in **(C)** brain tissues (6) and **(D)** microglia (24). Genes with $p < 3.2 \times 10^{-6}$ (Bonferroni corrected for 452 SNPs and 35 genes) are shown. **(E)** Multiple potentially causal genes at 11p11.2 inferred by MetaXcan (31). Genome-wide association study summary statistics from the Lambert *et al.* (4), the Kunkle *et al.* (2), and the Jansen *et al.* (3) studies were integrated with eQTL of 13 GTEx brain tissues and whole blood (27,28), and DGN whole blood (29), respectively. Genes with $p < 1 \times 10^{-4}$ (Bonferroni corrected) are marked by asterisks (*). Aβ, amyloid-β; AD, Alzheimer's disease; BA, Brodmann area; DGN, Depression Genes and Networks; eQTL, expressional quantitative trait loci; GTEx, Genotype-Tissue Expression; LD, linkage disequilibrium; LLB, long LD block; MiGA, Microglia Genomic Atlas; SLB, short LD block; SNP, single nucleotide polymorphism.

prioritize potentially regulatory fVars at 11p11.2 (Figure 2A). We identified a total of 100 (of the 452) AD-associated SNPs located in bulk or single-cell active regulatory elements (Table S2). Of these SNPs, 24 were defined as potential fVars because they were predicted to affect the binding affinities of certain TFs to active regulatory elements (Figure 2B; Table S2). These fVars were in 6 separate LD blocks (1–8 fVars in each block, median = 3.5) (Figure 2B; Table S5), and 13 fVars were prioritized by both the bulk-tissue and single-cell data (4 blocks, each containing 2–6 fVars, median = 2.5) (Table S2), which provided additional evidence for the presence of multiple independent signals at 11p11.2.





SNP	Pos (hg19)	Ref/Alt	Location	Promoters/enhancers						Open chromatins		TFBS	diffTF	eGenes			ASE	Reporter assay
				H3K4me3	H3K9ac	H3K4me1	H3K27ac	scH3K4me3	scH3K27ac	ATAC	scATAC			Bulk brain	Microglia	Monocyte		
rs11038913	11:46559730	T/C	*AMBRA1* enhancer	0	1	7	1	0	0	1	8	40	2	*MADD*				
rs74486166	11:46700203	G/C	*ARHGAP1* enhancer	0	2	3	9	0	1	1	0	1	1	*MADD*				
rs11039149	11:47276675	A/G	*NR1H3* enhancer	3	4	19	14	0	3	1	2	17	1	*MADD, MTCH2, NR1H3*	*NR1H3*	*ACP2, DDB2, MADD*		√
rs10501320	11:47293799	G/C	*MADD* enhancer	11	9	19	10	1	2	1	2	18	1	*LRP4, FAM180B, MADD, MTCH2, ARHGAP1, NR1H3*	*NR1H3*			
rs11039159	11:47294021	G/T	*MADD* enhancer	10	8	19	8	1	2	1	0	13	3	*MADD, MTCH2, NR1H3*	*ACP2*	*ACP2, DDB2, MADD*		
rs71475909	11:47377283	A/G	*SPI1* enhancer	17	7	3	5	3	2	1	15	44	6	*FNBP4, MADD, MTCH2*	*AC090559.1*	*SPI1, NUP160*		
rs1542321	11:47379615	A/G	*SPI1* enhancer	0	0	0	2	0	0	1	0	7	1	*PTPMT1, FNBP4, C1QTNF4, PACSIN3, FAM180B, MADD, PSMC3, MTCH2, KBTBD4, NDUFS3, ACP2*	*PSMC3*	*SPI1, DDB2*	√	√
rs11039200	11:47379646	A/C	*SPI1* enhancer	0	0	0	2	0	0	1	0	7	3	*FNBP4, C1QTNF4, FAM180B, MADD, PSMC3, MTCH2, NDUFS3, ACP2*	*NUP160*	*SPI1, DDB2, MTCH2*	√	√
rs10734557	11:47380061	C/A	*SPI1* enhancer	0	0	2	2	0	0	1	0	7	1	*FNBP4, C1QTNF4, FAM180B, MADD, PSMC3, MTCH2, NDUFS3, ACP2*	*PSMC3*	*SPI1, DDB2*	√	√
rs10838702	11:47410888	G/T	*RP11-750H9.5* enhancer	1	1	2	3	1	1	1	0	3	1	*FNBP4, MADD, SLC39A13, MTCH2, NDUFS3,*	*AC090559.1*	*SPI1, SLC39A13, NUP160*		
rs55876153	11:47416636	G/A	*RP11-750H9.5* enhancer	5	4	3	13	1	1	2	3	49	4	*FNBP4, C1QTNF4, FAM180B, MADD, PSMC3, MTCH2, NDUFS3, ACP2*	*PSMC3*	*SPI1, DDB2,ACP2, MTCH2, NUP160*		
rs7947450	11:47429904	G/A	*SLC39A13* promoter	27	12	0	18	4	4	32	27	204	27	*FNBP4, MADD, SLC39A13, MTCH2*	*AC090559.1*	*SPI1, NUP160*		√
rs11039225	11:47430599	G/A	*SLC39A13* enhancer	27	12	17	17	4	4	2	12	20	1	*FNBP4, MADD, SLC39A13, MTCH2*	*AC090559.1*	*SPI1, NUP160*	√	√
rs2293577	11:47437202	T/C	*SLC39A13* enhancer	0	2	2	4	0	1	1	1	11	1	*FNBP4, C1QTNF4, FAM180B, MADD, PSMC3, MTCH2, ACP2*	*PSMC3*	*SPI1, C1QTNF4, NUP160*	√	√
rs10742814	11:47554337	G/A	*CELF1* enhancer	1	2	14	10	0	2	8	10	2	1	*FNBP4, C1QTNF4, NUP160, MADD, SLC39A13, MTCH2*	*NUP160*	*SPI1, NUP160*		
rs71475921	11:47573342	G/A	*CELF1* promoter	18	4	8	6	3	3	1	1	5	1	*FNBP4, MADD, MTCH2*	*NUP160*	*SPI1, NUP160*		√
rs2280231	11:47600438	C/T	*KBTBD4-NDUFS3* promoter	28	12	0	18	4	4	32	27	297	34	*FNBP4, MADD, MTCH2*	*NUP160*	*MTCH2,NUP160*	√	√
rs35624992	11:47615798	G/C	*C1QTNF4* promoter	21	9	5	7	4	4	3	3	16	2	*FNBP4, C1QTNF4, FAM180B, PSMC3, MADD, MTCH2, KBTBD4, NDUFS3*	*NUP160*	*SPI1, NUP160*	√	√
rs7120548	11:47662932	T/C	*MTCH2* promoter	20	6	17	10	2	2	1	0	9	1	*FNBP4, MADD, MTCH2*	*NUP160*	*SPI1, NUP160*		√
rs7124396	11:47787434	G/T	*FNBP4* enhancer	20	6	3	4	3	0	1	0	20	2	*FNBP4, C1QTNF4, NUP160, MADD, MTCH2, NDUFS3*	*NUP160*	*SPI1, NUP160*		
rs7111576	11:47787445	G/A	*FNBP4* enhancer	21	7	3	4	3	1	1	0	22	2	*PTPMT1, FNBP4, C1QTNF4, NUP160, FAM180B, PSMC3, MADD, MTCH2, NR1H3*	*PSMC3*	*SPI1, NUP160*		
rs7111606	11:47787538	G/A	*FNBP4* enhancer	27	9	4	8	3	2	1	0	28	3	*PTPMT1, FNBP4, C1QTNF4, NUP160, FAM180B, PSMC3, MADD, MTCH2, NR1H3*	*PSMC3*	*SPI1, NUP160*		
rs12223593	11:47789082	T/G	*FNBP4* promoter	9	3	0	8	3	3	32	27	104	13	*FNBP4, MADD, SLC39A13, MTCH2*	*NUP160*	*MTCH2, NUP160*		
rs11039376	11:47789549	G/T	*FNBP4* promoter	27	12	1	18	4	4	11	19	72	7	*PTPMT1, FNBP4, C1QTNF4, NUP160, FAM180B, PSMC3, MADD, MTCH2, NR1H3*	*PSMC3*	*SPI1, NUP160*		



Among 24 potential fVars, 7 (especially rs2280231 and rs7947450) were in promoters and were predicted to affect the binding of more TFs than fVars in enhancers (Figure 2B, C). A total of 99 TFs were predicted to be affected by these 24 variants, including 18 TFs that were affected by 2 or more fVars (Figure 2C). No variants were located in repressors as revealed by H3K27me3 and H3K4me3 data (39,40) in the analyzed cells and tissues (Table S2).

Besides the above regulatory fVars, 2 missense variants (rs1064608, *MTCH2* p.P.281A; rs7941404, *AGBL2* p.R394H) were found in 452 AD-associated SNPs at 11p11.2 (Table S2). *MTCH2* p.P.281A was predicted to be deleterious by combined annotation-dependent depletion (score = 24.8) (41), but the *AGBL2* p.R394H (score = 2.997) was predicted to be benign. We focused mainly on regulatory fVars in this study.

Allele-Specific Expression Analyses and Dual-Luciferase Reporter Assays Validated the Allelic Regulatory Effects of Multiple fVars at 11p11.2

Allele-specific expression (ASE) data of 53 GTEx tissues (27,28,42) were analyzed to verify the cis-regulatory effects of the prioritized fVars. Twelve potential fVars were captured by the ASE data, of which 6 showed significant allelic imbalanced expression in all combined GTEx tissues (binomial test, Bonferroni-corrected $p < .05$) (Figure 3A), and 5 had relatively insufficient capture rate for analysis (Figure S4). Four of the 6 fVars showed significant allelic effects (Bonferroni-corrected $p < .05$) (Figure S5) when only brain tissues were analyzed. In particular, fVar rs35624992 in the promoter of *C1QTNF4*, a gene specifically expressed in the brain (43), had brain-specific ASE effects (Figure 3A; Figure S5). fVars rs10734557, rs1542321, and rs11039200 in the enhancer of *SPI1*, which was primarily expressed in microglia and monocytes (9,15,43), showed more prominent allelic imbalance in whole blood than in the brain (Bonferroni-corrected $p < .05$) (Figure S5).

We further validated the allelic regulatory effects of 7 fVars with significant ASE by using dual-luciferase reporter assays in the commonly used HEK293T and U251 cell lines. Significantly different activities were observed for 2 alleles of each fVar, and the overall differential expression tendencies were consistent with the ASE results ($p < .05$) (Figure 3B–F). Interestingly, fVars in *SPI1*, i.e., rs1542321, rs11039200, and rs10734557, which are fairly close and in strong LD [$r^2 > 0.95$, European individuals (44)] with each other (Figure S6A), all showed independent allelic-specific effects as revealed by the reporter assays (Figure 3B). Considering potential cell type–specific expression of eQTL genes at 11p11.2, we further tested effects of these *SPI1* fVars using 2 human microglia cell lines, HMC3 and HM. Independent effects were also observed for these fVars, with a more profound effect for haplotype rs1542321-G: rs11039200-C:rs10734557-A (Figure S6B, C). Interestingly, the direction of differential expression was opposite in microglia cell lines and HEK293T and U251 cells, indicating cell type–specific effects of these fVars. We also tested 4 potential fVars that were not captured by the ASE data. Similarly, these fVars showed consistent allelic effects in both HEK293T and U251 cell lines (Figure 3G–J). We compared our results with the results of a recent study that used massively parallel reporter assay (MPRA) to identify regulatory variants for AD (45). Of 34 SNPs at 11p11.2 that were captured by MPRA (Table S6) (45), 3 fVars (rs71475921, rs7120548, and rs12223593) identified in this study had significant allelic effects in MPRA ($p < .05$). Furthermore, the allelic effects of rs2280231, rs71475921, and rs7120548 in MPRA (45) were consistent with our luciferase reporter assays (Figure 3E, I, J). Taken together, these results confirmed the reliability of our fine-mapping strategy and the coexistence of multiple fVars at 11p11.2.

Integration of eQTL and Chromatin Interaction Data Assigned 17 Target Genes to Potential fVars at 11p11.2

Bulk brain, microglia, and monocyte eQTL datasets showed that most potential fVars at 11p11.2 were linked with expression of multiple genes (eGenes for each SNP, Bonferroni-corrected eQTL $p < .05/35$ genes within ± 1 Mb of 11p11.2 ≈ 0.001) (Figure 2B). We integrated available chromatin interaction data (high-throughput chromosome conformation capture [HiC] and proximity ligation–assisted chromatin immunoprecipitation sequencing data) of neurons (37,46), astrocytes (46), microglia (37), oligodendrocytes (37), and monocytes (39,40) with eQTL data (6,23–26) to confirm the regulation of eGenes by potential fVars. A total of 17 candidate target genes (eQTL $p < .001$; HiC score > 3) were assigned to 24 potential fVars (Table S7). Fifteen of these genes had no cell type–specific expression in the brain (Figure S7). *SPI1* and *DDB2* (primarily expressed in microglia) (Figure S7) were specifically assigned as target genes of fVars with the monocyte

Figure 2. Systematic functional genomic study prioritized multiple potential fVars at 11p11.2. **(A)** A fine-mapping strategy for distilling fVars. **(B)** Detailed annotation for 24 potential fVars. Numbers in the 5th–12th columns represent the number of datasets supporting each term; the TFBS column shows the number of distinct TF binding sites in which the target variant was located; the diffTF column shows the number of TFs whose binding affinities were disrupted by alternative alleles of target variants; bulk-tissue histone modification data of 8 brain regions, 6 types of neural cells, and monocytes are from the ENCODE (39,40) (H3K4me3, $n = 28$; H3K9ac, $n = 12$; H3K4me1, $n = 24$; H3K27ac, $n = 18$); single-cell H3K4me3 (scH3K4me3, $n = 4$) and H3K27ac (scH3K27ac, $n = 4$) data are from neurons, astrocytes, microglia, and oligodendrocytes isolated from resected cortical brain tissues (37); bulk-tissue ATAC sequencing data (ATAC, $n = 29$) are from 14 brain regions (91) and monocytes (92); scATAC data ($n = 31$) are from scATAC sequencing of 4 brain regions (36), ATAC sequencing of induced pluripotent stem cell–induced cortical excitatory neurons (46), induced pluripotent stem cell–derived hippocampal dentate gyrus–like neurons (46), primary astrocytes (46), neurons (37), astrocytes (37), microglia (37), and oligodendrocytes (37); TFBS data are from the ENCODE ($n_{TFs} = 623$, $n_{datasets} = 1322$); bulk brain eGenes are from the eMeta (6) and psychENCODE (23), microglia eGenes are from the Microbial Genomes Atlas (24), and monocyte eGenes are from Raj *et al.* (25) and Kim-Hellmuth *et al.* (26); fVars validated by ASE or luciferase-reporter assays are labeled by a tick. **(C)** TFs disrupted by 2 or more potential fVars. The dot size is proportional to the number of potential fVars that disrupted TF binding. AD, Alzheimer's disease; alt/ref, alternative allele/reference allele; ASE, allele-specific expression; ATAC, assay for transposase-accessible chromatin; eGenes, expressional quantitative trait loci genes of target variants; ENCODE, Encyclopedia of DNA Elements; fVar, functional variant; GWAS, genome-wide association study; LD, linkage disequilibrium; LLB, long LD block; Pos, chromosome position; scATAC, single-cell ATAC; SLB, short LD block; SNP, single nucleotide polymorphism; TF, transcription factor.

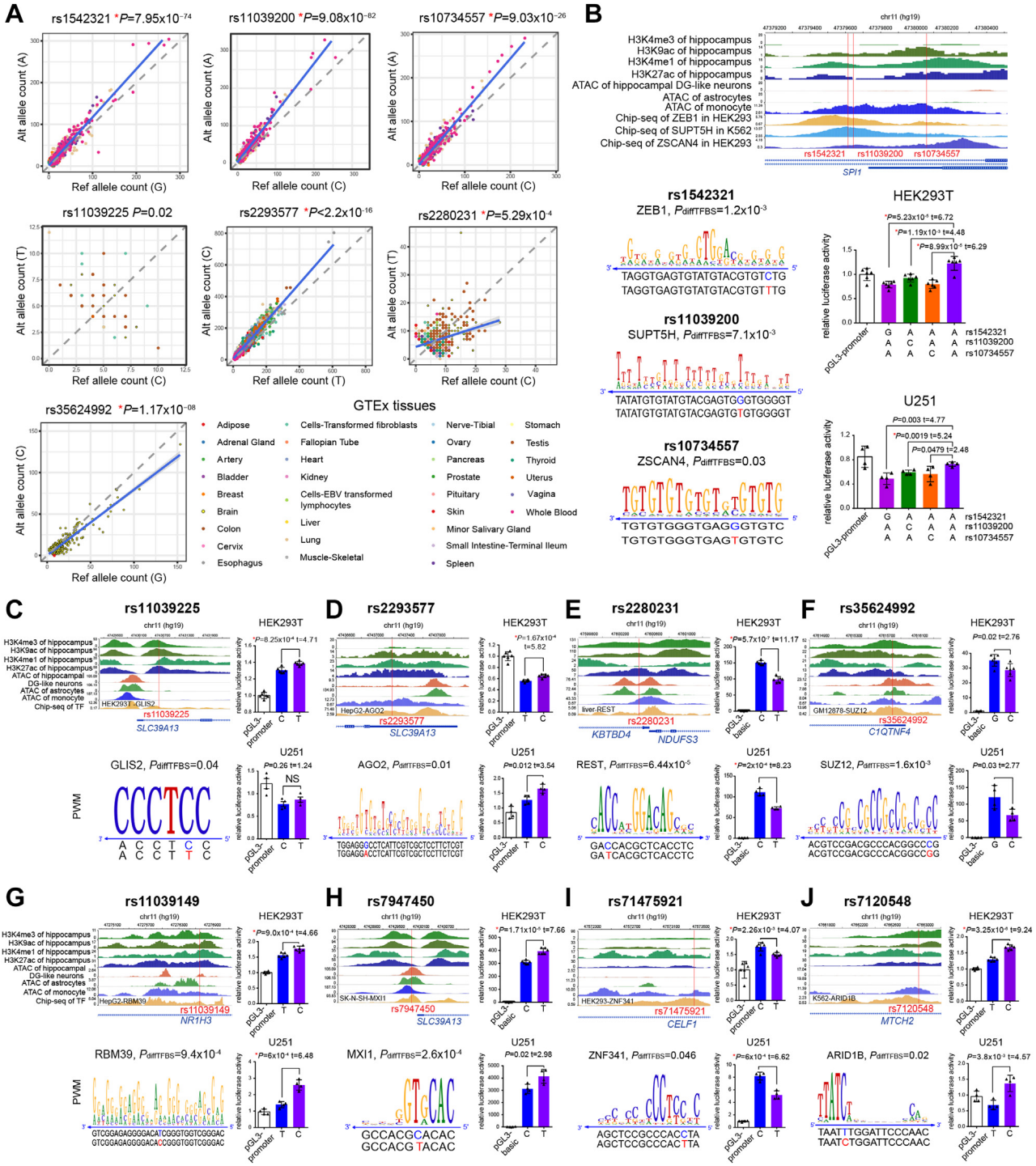


Figure 3. Allele-specific expression analyses and dual-luciferase reporter assays validated the coexistence of multiple fVars at 11p11.2. **(A)** fVars with significant allele-specific expression (allelic imbalance) in all GTEx human tissues (27,28,42). Allele counts for the Ref allele and the Alt allele were plotted for each variant. Each dot represents an individual sample and is colored by tissue. For each dot plot, the blue line represents a linear regression of all points, and the gray shade represents the 95% CI for the regression. *p* Values were measured by binomial tests. *p* Values < .007 after Bonferroni correction for the total number of comparisons (.05/7) are marked by red asterisks (*). **(B)** Functional annotation and dual-luciferase reporter assays for 3 *SPI1* fVars in strong linkage disequilibrium. (Top panel) Regulatory elements containing each fVar were represented by ChIP-seq of histone modifications of promoters (H3K4me3 and H3K9ac) and enhancers (H3K4me1 and H3K27ac) (39,40), ATAC-seq (36,37,46,91,92), and ChIP-seq of representative TFs disrupted by fVars (39,40). (Bottom left panels) PWM of TFs whose binding affinities were predicted to be disrupted by fVars. (Bottom right panels) Dual-luciferase reporter assays for fVars in HEK293T and U251 cells. **(C–J)** Functional annotation and dual-luciferase reporter assays for each fVar. Top left, regulatory elements containing each potential fVar were represented by ChIP-seq of histone modifications of

data (25,26). We found no microglia-specific target genes (probably) because of the relatively small sample size of the microglia eQTL datasets (24).

Excluding rs11038913 and rs74486166, each of the remaining 22 potential fVars had more than one target gene. In particular, *SPI1* fVars rs1542321, rs11039200, and rs10734557 were assigned the largest number of target genes, e.g., *SPI1*, *MADD*, *MTCH2*, and *PSMC3*. Among all candidate target genes, *MTCH2* and *MADD* interacted with the largest number of potential fVars (Table S7). When a more stringent threshold (eQTL $p < 3.2 \times 10^{-6}$ and HiC score > 5) was used to assign target genes to fVars, 9 genes were retained, supporting the robustness of the pattern wherein most fVars at 11p11.2 are linked to multiple eGenes (Table S7).

A Single fVar Regulated Multiple Genes Through Chromatin Interactions

To validate whether a single fVar could regulate multiple genes, we created cell clones carrying different genotypes of fVars at 11p11.2 by using base editing (47). Among 7 ASE and reporter assay–validated fVars (Figure 3A–F), rs2293577 and rs2280231 were chosen because they were located in the editing window of base editor ABE7.10 (47).

rs2293577 is in an enhancer within *SLC39A13* (Figure 3D). The effect C-allele of rs2293577 was associated with upregulated expression of *ACP2*, *MADD*, *PSMC3*, *FAM180B*, *MTCH2*, and *FNBP4* and downregulated *C1QTNF4* expression in brain tissues (Figure 4A). It was also associated with *PSMC3* in microglia and with *SPI1*, *C1QTNF4*, and *NUP160* in monocytes (Table S7). Six of these eGenes interacted with rs2293577-located enhancer by chromatin contacts (Figure 4B) (HiC score > 3). We performed precise base editing for rs2293577 in HEK293T cells and obtained single-cell–derived clones with genotypes TT (reference genotype) and TC (Figure S8). Assay for transposase-accessible chromatin sequencing and real-time quantitative polymerase chain reaction were conducted to test the regulatory effect of rs2293577 T>C on its potential target genes. The results showed that C-allele of rs2293577 did not affect the chromatin accessibility of the *SLC39A13* enhancer (Figure 4C, D). Consistent with the eQTL data (Figure 4A), all target genes of rs2293577 in cells with genotype TC had higher chromatin accessibility and messenger RNA (mRNA) levels than cells with TT, especially for *ACP2* and *MTCH2* (Figure 4C–E), supporting the regulation of multiple target genes by a single fVar, rs2293577.

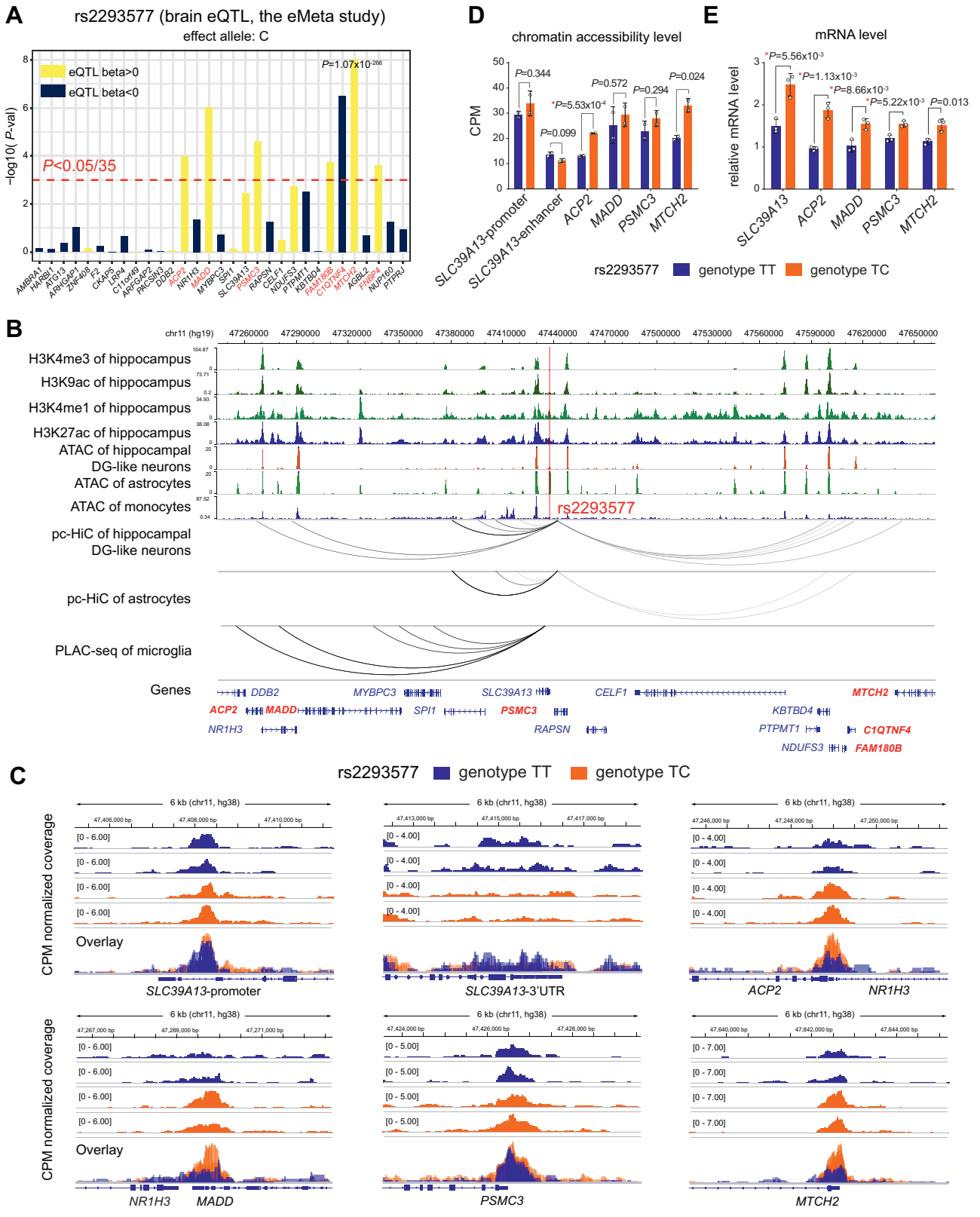
Another fVar, rs2280231, which is in the bidirectional promoter of *KBTBD4* and *NDUFS3* (Figure 3E), showed a similar effect. The effect T-allele of rs2280231 was associated with lower expression of *MADD*, *MTCH2*, and *FNBP4* in brain tissues (Figure 5A) and was associated with *NUP160* in microglia and monocytes (Table S7). *MADD*, *MTCH2*, and *FNBP4* also had significant chromatin interactions with rs2280231 (Figure 5B). *NUP160* interacted with rs2280231 in neurons and astrocytes, but not microglia or monocytes, which may be due to the insufficient capture of HiC of these 2 cells. Similarly, HEK293T cells carrying reference genotype CC and edited genotype CT of rs2280231 were obtained by precise base editing (Figure S8). The chromatin accessibility of rs2280231-located promoter was not altered (Figure 5C, D), while the *KBTBD4* mRNA level was significantly downregulated (Figure 5E). Inconsistent with eQTL results (Figure 5A), the chromatin accessibility and mRNA levels of *MADD*, *FNBP4*, and *NUP160* were significantly decreased in HEK293T cells with genotype CT compared with that in HEK293T cells with CC (Figure 5C–E).

Overall, these results provided additional evidence for the coexistence of multiple fVars at 11p11.2 and that one fVar could regulate multiple target genes.

The CFG Strategy Prioritized 6 Likely Causal Genes in AD Pathogenesis at 11p11.2

Seventeen potential target genes were assigned to fVars (Table S7); however, it is possible that only a small proportion of these genes have a role in causing AD and that others are the result of pleiotropy. Thus, we applied a CFG strategy integrating multilines of AD-related evidence to assess each gene's relevance to AD (18,48,49). Evidence from integrative analyses of AD GWASs and eQTL, single-cell (50), or bulk brain (18) transcriptomes, epigenomes (51), and proteomes (52) at different AD pathological stages were considered. Six genes had more than 3 lines of evidence supporting their involvement in AD (Table 1). Among those genes, *MTCH2* received the highest CFG score, followed by *ACP2* and *NDUFS3*. The previously established causal gene *SPI1* (9,15,16) had a relatively low CFG score of 2. This was because most of the data used for CFG analysis were collected from bulk brain tissues, which may not be suitable for accurate quantification of genes that were mainly expressed in microglia, e.g., *SPI1*. Except for *SPI1* (9), nearly all genes showed a consistently downregulated tendency in brains of patients with AD compared with control brains at the single-cell level (Figure 6A; Table S8), bulk brain mRNA level (Figure 6B; Table S9), and epigenetic (H3K27ac, Figure 6C) and protein levels (Figure 6D). These changes were consistent with the eQTL effects of fVars at 11p11.2, in which upregulation of *SPI1* but downregulation of other genes was associated with increased AD risk (Table S7). Altogether, different levels of omics data indicated that multiple genes at 11p11.2 were likely to be involved in AD pathogenesis.

promoters and enhancers (39,40), ATAC-seq (36,37,46,91,92), and ChIP-seq of TFs disrupted by fVars (39,40); bottom left, PWM of the TF whose binding affinity was predicted to be disrupted by an fVar; right, dual-luciferase reporter assays for each fVar in HEK293T and U251 cells. Functional annotations in **(B–J)** were displayed on genome assembly hg19. For each fVar **(B–J)**, the most significantly disrupted PWMs are shown. The reporter assay results in **(B–J)** are representative of 3 independent experiments with similar results; bars represent mean ± SD ($n = 6$ biological replicates for HEK293T cells and $n = 4$ biological replicates for U251 cells). p Values were calculated by two-sided Student's t test, together with the t statistics (HEK293T cells, $df = 10$; U251 cells, $df = 6$). p Values < .002 after Bonferroni correction for the total number of comparisons (.05/22) are marked by red asterisks (*). Alt, alternative; ATAC, assay for transposase-accessible chromatin; ChIP-seq, chromatin immunoprecipitation sequencing; DG, dentate gyrus; EBV, Epstein–Barr virus; fVar, functional variant; GTEx, Genotype-Tissue Expression; NS, not significant; PWM, position weight matrix; Ref, reference; TF, transcription factor.


A
rs2293577 (brain eQTL, the eMeta study)
effect allele: C
eQTL beta>0
eQTL beta<0
P=1.07x10-266
P<0.05/35
−log10(P-val)
AMBRA1 HARBI1 ATG13 ARHGAP1 ZNF408 F2 CKAP5 LRP4 C11orf49 ARFGAP2 PACSIN3 DDB2 ACP2 NR1H3 MADD MTBPC3 SPI1 SLC39A13 PSMC3 RAPSN CELF1 NDUFS3 PTPMT1 KBTBD4 FAM180B C1QTNF4 MTCH2 AGBL2 FNBP4 NUP160 PTPRJ
D
chromatin accessibility level
CPM
P=0.344
P=0.099
*P=5.53x10-4
P=0.572
P=0.294
P=0.024
SLC39A13-promoter
SLC39A13-enhancer
ACP2
MADD
PSMC3
MTCH2
E
mRNA level
relative mRNA level
*P=5.56x10-3
*P=1.13x10-3
*P=8.66x10-3
*P=5.22x10-3
P=0.013
SLC39A13
ACP2
MADD
PSMC3
MTCH2
rs2293577 genotype TT genotype TC
B
chr11 (hg19)
H3K4me3 of hippocampus
H3K9ac of hippocampus
H3K4me1 of hippocampus
H3K27ac of hippocampus
ATAC of hippocampal DG-like neurons
ATAC of astrocytes
ATAC of monocytes
rs2293577
pc-HiC of hippocampal DG-like neurons
pc-HiC of astrocytes
PLAC-seq of microglia
Genes
DDB2 MYBPC3 SLC39A13 CELF1 MTCH2
ACP2 MADD SPI1 PSMC3 KBTBD4
NR1H3 RAPSN PTPMT1 C1QTNF4
NDUFS3 FAM180B
C
rs2293577 genotype TT genotype TC
CPM normalized coverage
Overlay
SLC39A13-promoter
SLC39A13-3'UTR
ACP2
NR1H3
NR1H3
MADD
PSMC3
MTCH2

Multiple Genes at 11p11.2 Affected AD-Related Molecular Phenotypes

Likely causal genes at 11p11.2 were involved in different AD-related pathways (18, 53–56), including energy metabolism [*NDUFS3* (57), *MTCH2* (58–60), and *C1QTNF4* (61,62)], autophagy [*ACP2* (63,64)], immunity [*C1QTNF4* (61,62)], and ubiquitination [*PSMC3* (65)]. Knockout of most of these genes caused AD-related pathologies (Table S10) (63,66,67). No protein-protein interactions were observed among these genes (68), and they had no cell type–specific expression in the brain (Figure S7). Therefore, likely causal genes at 11p11.2 may be involved in AD pathogenesis via different pathways, either individually or synergistically.

Considering the fact that SNPs at 11p11.2 were nominally associated with $A\beta_{42}$ levels in the cerebrospinal fluid (Figure 1A), we knocked down or overexpressed 3 likely causal genes that were frequently highlighted by previous studies, i.e., *MADD* (10–12), *PSMC3* (2,11), and *MTCH2* (2,12) (Figure 7A). Specifically, knockdown of each of these 3 genes significantly increased $A\beta_{42}$ levels in U251-APP cells (69,70), whereas overexpression had opposite effects (Figure 7B). Expression alterations of *MADD* and *PSMC3* also affected the phosphorylated tau (pTau-396) levels (Figure 7C). It is interesting that disruption of these genes led to alterations of both Aβ and pTau levels, which could be affected by multiple processes (53,71). For example, overexpression of *PSMC3* significantly decreased $A\beta_{42}$ and pTau levels, probably due to enhanced function of proteasomes. Knockdown of *MADD* and *MTCH2* had more profound effects on $A\beta_{42}$ levels than their overexpression, suggesting different mechanisms of these genes in contributing to AD pathogenesis (Figure 7B, C). All these results supported involvement of multiple genes at 11p11.2 in AD pathogenesis. The exact mechanisms by which disruption of genes with diverse molecular functions converge upon the phenotype of increased $A\beta_{42}$ and pTau levels remain to be elucidated.

DISCUSSION

Elucidating mechanisms underlying AD GWAS risk loci is particularly challenging for loci of extended LD (33,34). Fine-mapping strategies integrating association statistics (e.g., QTL) and epigenetic datasets have paved a critical step from identifying risk loci to pinpointing likely causal variants and genes (2,3,7,9,11,37,72–74). However, the results for complex GWAS loci tend to vary across studies, mainly due to intensive LD, limited types, and sample sizes of integrated omics datasets. In this study, we focused on 11p11.2, a typically complicated AD risk signal containing numerous highly correlated SNPs and coexpressed genes. Although stepwise conditional analysis (75–78) of GWASs (2–4) showed that there might be only one statistically significant causal/association signal at 11p11.2 (Figure S9), this signal cannot usually reflect the number of biologically causal variants because it only represents a combinatory pattern of potentially causal variants (33,34). Biologically causal variants at 11p11.2 required further functional fine-mapping and experimental investigations.

We took advantage of recent large-scale genetic, epigenetic, transcriptomic, and proteomic datasets of brain tissues, neural cells, and monocytes at the bulk-tissue or single-cell level to enable the prioritization of potential fVars and genes. Multiomic functional genomic study assisted by cellular validations prioritized and validated multiple potential fVars at 11p11.2. In addition to *SPI1*, 6 likely causal genes were identified. Our results suggested a "multiple causal genes at a single locus" pattern for complex AD risk loci, which may provide a novel framework toward the mechanistic interpretation of genetic risk signals of AD and other complex diseases.

The complexity of gene regulation at 11p11.2 has also been reported in previous studies (2,7,10–13,74). For instance, Karch *et al.* (7) identified several AD-associated eQTL genes at 11p11.2. They speculated that genes may act cooperatively to modify AD risk, and a key regulator at 11p11.2 may influence the expression of many other genes (7). Huang *et al.* (9) found that downregulation of PU.1 (SPI1) by the protective haplotype at 11p11.2 was responsible for a lower AD risk, and PU.1 may be a key regulator of AD. Notably, *SPI1* was mainly expressed in microglia, but other genes at 11p11.2 had no apparent cell-type–expression specificity in the brain (Figure S7) (50). PU.1 knockout in B cells had no obvious effects on expression of eQTL genes at 11p11.2 (Figure S10) (79). These observations provided no direct support for *SPI1* as the single causal gene. Recently, based on myeloid cell omics data, Novikova *et al.* (11) found that expression of *MADD* and *NUP160* were likely to be regulated by active enhancers at 11p11.2 and were causally

Figure 4. Functional variant rs2293577 regulated expression of multiple target genes. **(A)** rs2293577 was associated with expression of multiple genes at 11p11.2. Brain eQTL data were taken from Qi *et al.* (6). Red dotted line represents the Bonferroni-corrected threshold for eGenes of each single nucleotide polymorphism ($p < .05/35$ genes), and eGenes are marked in red. **(B)** Multiple eGenes interacted with rs2293577 via chromatin interactions. Chromatin immunoprecipitation sequencing datasets of histone modifications are from the ENCODE (39,40). ATAC-seq datasets of hippocampal DG-like neurons and astrocytes (46), and monocytes (92), pc-HiC dataset of hippocampal DG-like neurons and astrocytes (46), and PLAC-seq dataset of microglia (37) were taken from public sources. Interactions with a score >3 ($p < .001$) are shown. eGenes with significant chromatin interactions with rs2293577 are marked in red. Functional annotations are displayed on genome assembly hg19. **(C)** Chromatin accessibility of active regulatory element–harboring rs2293577 and target genes of this functional variant in HEK293T cells with different rs2293577 genotypes. Coverage signals were normalized by CPM. Each genotype (TT and TC) of rs2293577 had 2 biological replicates for the ATAC-seq. Peaks were displayed on genome assembly GRCh38. **(D)** Quantification of chromatin accessibility changes in **(C)**. Each dot represents 1 biological replicate. Bars are presented as mean ± SD. *p* Values were calculated by two-sided Student's *t* test. The *t* statistics was based on $df = 2$ for all comparisons: *SLC39A13*-promoter, $t = 1.23$; *SLC39A13*-enhancer, $t = 2.95$; *APC2*, $t = 42.50$; *MADD*, $t = 0.67$; *PSMC3*, $t = 1.41$; *MTCH2*, $t = 6.30$. **(E)** mRNA levels of rs2293577-located gene and its target genes in HEK293T cells with different rs2293577 genotypes. Each genotype had 3 biological replicates. Bars are presented as mean ± SD. *p* Values were calculated by two-sided Student's *t* test. For all comparisons, $df = 4$; *t* statistics: *SLC39A13*, $t = 5.44$; *APC2*, $t = 8.34$; *MADD*, $t = 4.80$; *PSMC3*, $t = 5.53$; *MTCH2*, $t = 4.30$. *p* Values < .01 after Bonferroni correction for the total number of comparisons (.05/5) are marked by red asterisks (*). ATAC, assay for transposase-accessible chromatin; CPM, counts per million; DG, dentate gyrus; ENCODE, Encyclopedia of DNA Elements; eQTL, expressional quantitative trait loci; HiC, high-throughput chromosome conformation capture; mRNA, messenger RNA; pc-HiC, promoter capture HiC; PLAC-seq, proximity ligation–assisted chromatin immunoprecipitation sequencing; UTR, untranslated region.

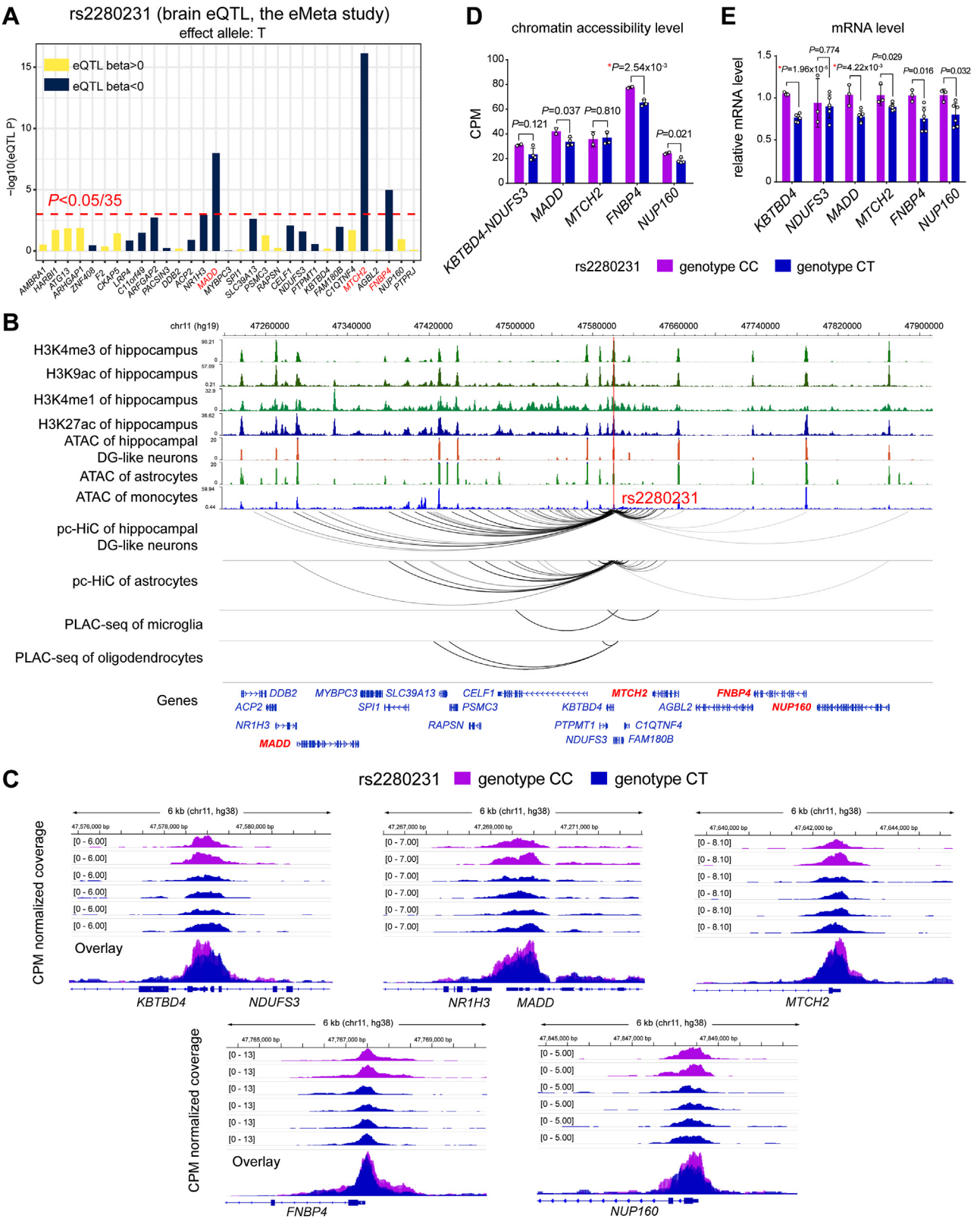

A
rs2280231 (brain eQTL, the eMeta study)
effect allele: T
eQTL beta>0
eQTL beta<0
−log10(eQTL P)
P<0.05/35
AMBRA1 HAAO BIC ATG13 ARHGAP1 ZNF408 P2 CKAP5 LRP4 C11orf49 ARFGAP2 PACSIN3 DDB2 ACP2 NR1H3 MADD MYBPC3 SPI1 SLC39A13 PSMC3 RAPSN CELF1 NDUFS3 PTPMT1 KBTBD4 FAM180B C1QTNF4 MTCH2 AGBL2 FNBP4 NUP160 PTPRJ
D
chromatin accessibility level
CPM
P=0.121
P=0.037
P=0.810
*P=2.54x10-3
P=0.021
KBTBD4-NDUFS3 MADD MTCH2 FNBP4 NUP160
E
mRNA level
relative mRNA level
*P=1.96x10-5
P=0.774
*P=4.22x10-3
P=0.029
P=0.016
P=0.032
KBTBD4 NDUFS3 MADD MTCH2 FNBP4 NUP160
rs2280231 genotype CC genotype CT
B
chr11 (hg19)
H3K4me3 of hippocampus
H3K9ac of hippocampus
H3K4me1 of hippocampus
H3K27ac of hippocampus
ATAC of hippocampal DG-like neurons
ATAC of astrocytes
ATAC of monocytes
rs2280231
pc-HiC of hippocampal DG-like neurons
pc-HiC of astrocytes
PLAC-seq of microglia
PLAC-seq of oligodendrocytes
Genes
DDB2 MYBPC3 SLC39A13 CELF1 MTCH2 FNBP4
ACP2 SPI1 PSMC3 KBTBD4 AGBL2 NUP160
NR1H3 RAPSN PTPMT1 C1QTNF4
MADD NDUFS3 FAM180B
C
rs2280231 genotype CC genotype CT
CPM normalized coverage
6 kb (chr11, hg38)
Overlay
KBTBD4 NDUFS3
NR1H3 MADD
MTCH2
FNBP4
NUP160

Table 1. Convergent Functional Genomics Ranking of Candidate Target Genes of Potential Functional Variants at 11p11.2

Genes	Integrative Analyses, GWAS + eQTL			Early-Stage mRNA Change	Late-Stage mRNA Change		Epigenetic Change (H3K27ac)	Protein Change	CFG Score
	MetaXcan	SMR	Colocalized		Single-Cell	Bulk			
MTCH2	✓	✓	✓	✓	✓	✓	✓	✓	5
ACP2	✓	–	✓	✓	✓	✓	✓	–	4
NDUFS3	✓	–	–	✓	✓	✓	–	✓	4
PSMC3	✓	–	–	✓	–	✓	✓	–	4
C1QTNF4	✓	–	✓	✓	–	✓	✓	–	4
MADD	✓	–	–	✓	–	✓	✓	–	4
LRP4	✓	–	–	–	–	✓	✓	–	3
NUP160	–	–	–	✓	✓	–	–	–	2
DDB2	–	–	–	–	–	✓	✓	–	2
SPI1	✓	–	✓	–	–	–	✓	–	2
SLC39A13	✓	✓	–	–	–	–	–	–	1
FAM180B	✓	–	–	–	–	–	–	–	1
NR1H3	✓	–	–	–	–	–	✓	–	2
PACSIN3	–	–	–	–	–	–	✓	–	1
FNBP4	–	–	–	–	–	–	–	–	0
KBTBD4	–	–	–	–	–	–	–	–	0
PTPMT1	–	–	–	–	–	–	–	–	0

Colocalized: colocalized genes (PP.H4 > 0.7) as reported in Lopes *et al.* (24). Early-stage mRNA change: genes differentially expressed in AD patients with early-stage pathology compared with control individuals based on single-cell RNA sequencing from Mathys *et al.* (50). Late-stage mRNA change: genes differentially expressed in AD patients with late-stage pathology compared with control individuals. Single-cell dataset was taken from Mathys *et al.* (50); bulk brain tissue dataset was taken from Alzdata [Xu *et al.* (18) and references therein]. Epigenetic change (H3K27ac): genes with significant H3K27ac change in patients with AD compared with control individuals based on data reported by Marzi *et al.* (51). Protein change: proteins with significant change in patients with AD compared with control individuals based on data from Johnson *et al.* (52). CFG score: 1 point was added if a gene had a significant result (✓) in each of the 5 lines of evidence (the integrative analysis, early-stage mRNA change, late-stage mRNA change, epigenetic change, and protein change). The sum of all lines of evidence resulted in a total CFG score ranging from 0 (no association) to 5 (the strongest association).

AD, Alzheimer's disease; CFG, convergent functional genomics; eQTL, expressional quantitative trait loci; GWAS, genome-wide association study; mRNA, messenger RNA; SMR, summary-data-based Mendelian randomization.

linked to AD risk. Chen *et al.* (13) found that AD risk SNPs may coregulate several nearby genes rather than only the nearest genes. These reports, together with our current findings, indicate that multiple fVars and causal genes cofunction at 11p11.2. However, systematic functional assessments and experimental validations are lacking to rule out the effects of LD and gene coexpression. In this study, we validated the coexistence of multiple fVars by functional genomics, reporter assays, and base-editing experiments. We proved that each fVar has a regulatory effect. One variant could regulate multiple genes by chromatin contacts, as exemplified by base editing of rs2293577 and rs2280231. Furthermore, CFG scoring and gene-level assays confirmed that multiple genes at 11p11.2 were able to affect AD-related molecules. Thus, our study has shown that multiple likely causal genes regulated by fVars, rather than a single key regulator, accounted for AD risk at 11p11.2.

This multiple-causal-variant, multiple-causal-gene pattern at a single locus was not restricted to 11p11.2 and may also be applicable to other AD loci and complex diseases. For example, recent studies reported several AD susceptible coding and noncoding variants at the *APOE* locus (80–82). Kikuchi *et al.* (73) found an enhancer variant at AD risk locus 7q22.1 connecting many eQTL genes via the CTCF-mediated chromatin loops. For the hypertension risk locus *Agtrap-Plod1*, knocking out each of the 5 genes at this locus affected

Figure 5. Functional variant rs2280231 regulated expression of multiple target genes. **(A)** rs2280231 was associated with expression of multiple genes at 11p11.2. **(B)** Multiple eGenes of rs2280231 interacted with rs2280231 via chromatin interactions. **(C)** Chromatin accessibility of active regulatory element–harboring rs2280231 and target genes of this functional variant in HEK293T cells with different rs2280231 genotypes. **(D)** Quantification of chromatin accessibility changes in **(C)**. HEK293T cells with genotype CC had 2 biological replicates, and genotype CT had 4 biological replicates. The *df* = 4 for all comparisons to calculate Student *t* test *p* values; *t* statistics: *KBTBD4-NDUFS3*, *t* = 1.96; *MADD*, *t* = 3.08; *MTCH2*, *t* = 0.26; *FNBP4*, *t* = 6.73; *NUP160*, *t* = 3.68. *p* Values < .01 after Bonferroni correction for the total number of comparisons (.05/5) are marked by red asterisks (*). **(E)** mRNA levels of rs2280231-located genes and target genes in HEK293T cells with different rs2280231 genotypes. Cells with genotype CC had 3 biological replicates, and cells with genotype CT had 6 biological replicates. The *df* = 7 for all comparisons to calculate Student *t* test *p* values; *t* statistics: *KBTBD4*, *t* = 10.13; *NDUFS3*, *t* = 0.30; *MADD*, *t* = 4.16; *MTCH2*, *t* = 2.73; *FNBP4*, *t* = 3.17; *NUP160*, *t* = 2.66. *p* Values < .008 after Bonferroni correction for the total number of comparisons (.05/6) are marked by red asterisks (*). Additional information can be found in the legend for Figure 4. CPM, counts per million; eQTL, expressional quantitative trait loci; HiC, high-throughput chromosome conformation capture; mRNA, messenger RNA; pc-HiC, promoter capture HiC; PLAC-seq, proximity ligation–assisted chromatin immunoprecipitation sequencing.

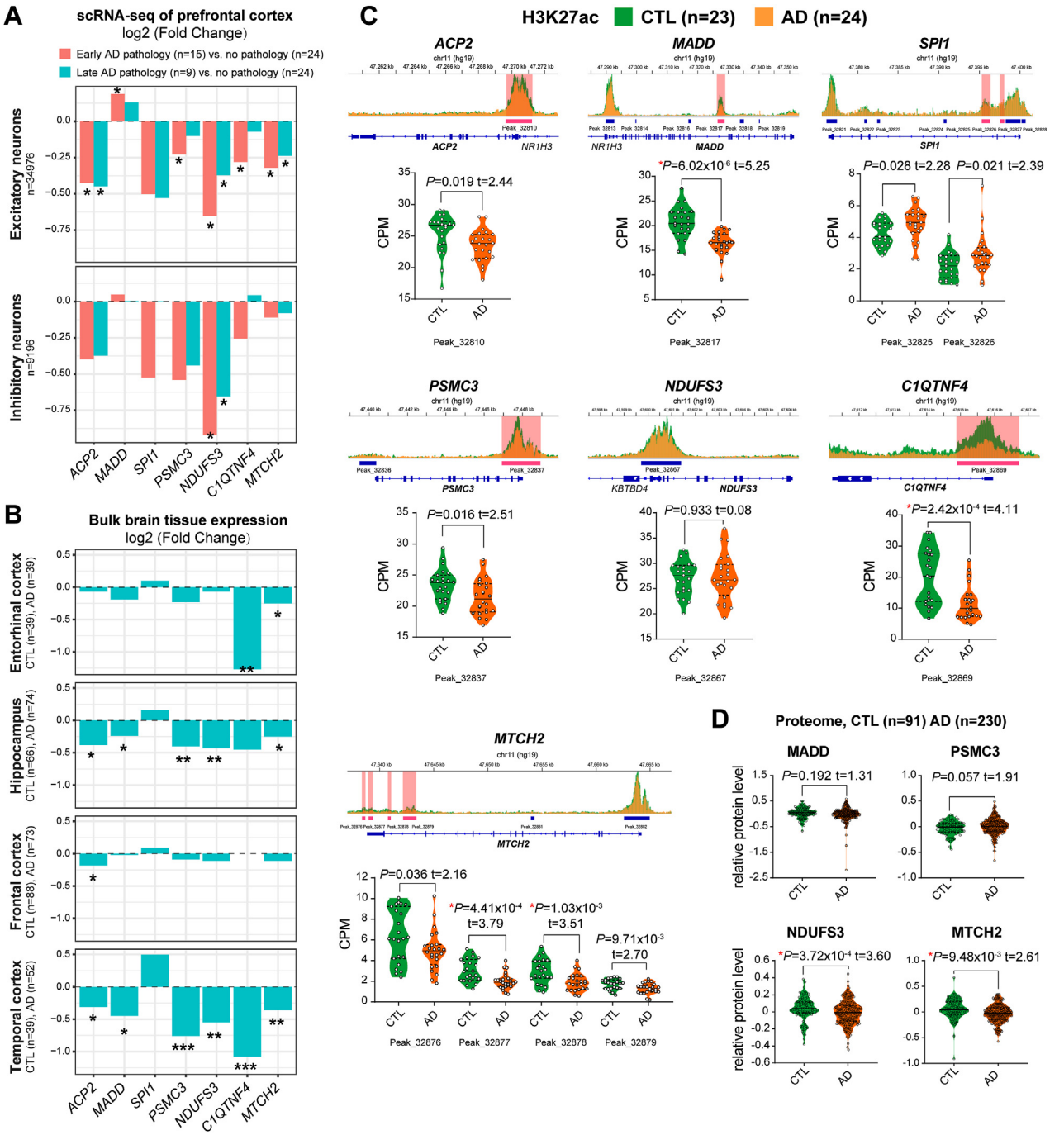
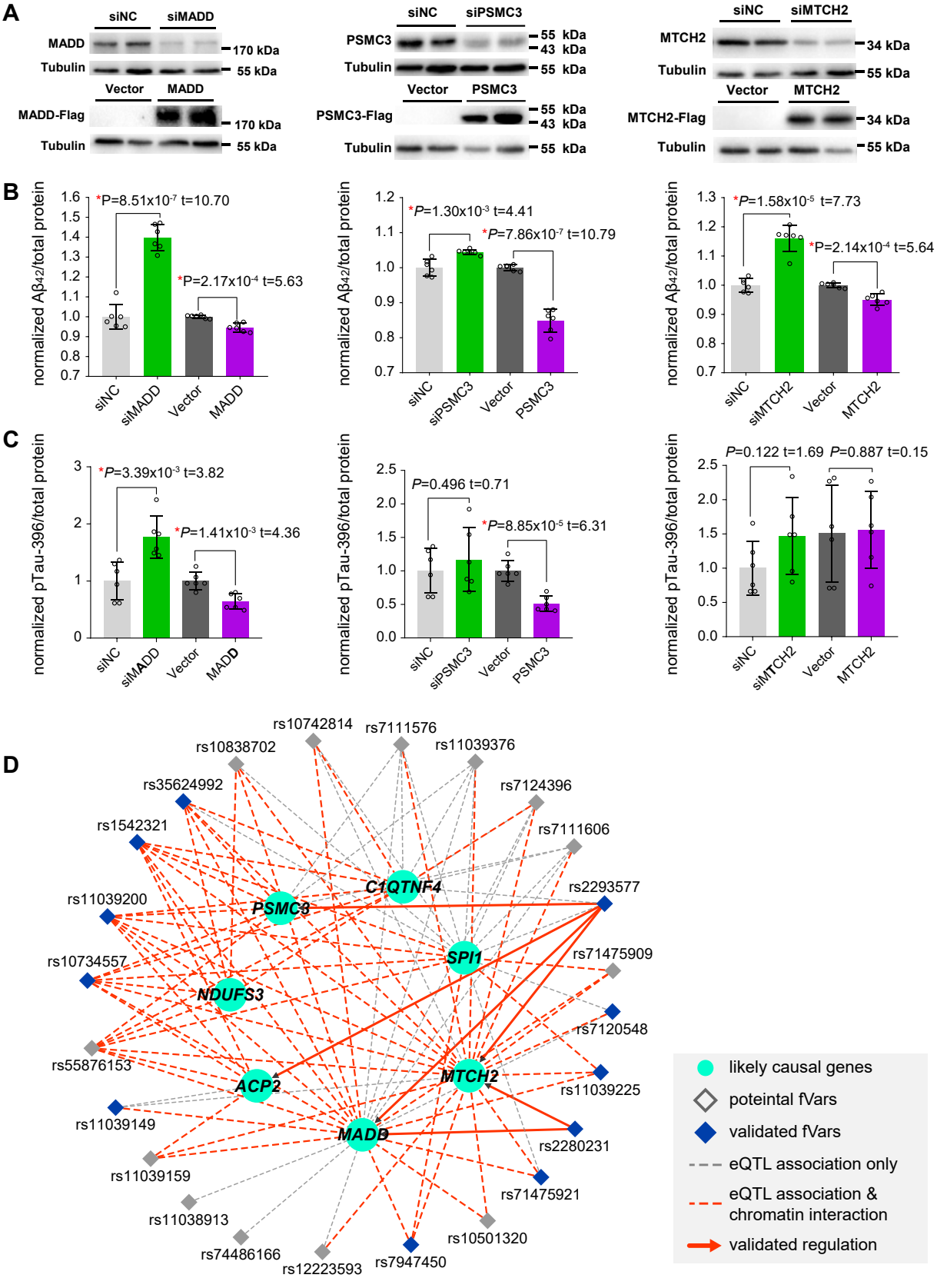


Figure 6. Multilevel omics data supported an involvement of 6 genes at 11p11.2 in AD pathogenesis. **(A, B)** Differential expression of genes likely to be involved in AD pathogenesis at 11p11.2. The single-cell **(A)** and bulk brain tissue **(B)** datasets are from public sources. For scRNA-seq (50), genes with false discovery rate < .05 for both the two-sided Wilcoxon rank-sum test and the Poisson mixed-model test were regarded as statistically significant. For bulk brain tissues (18), genes with false discovery rate < .05 based on the empirical Bayes method (*t* tests) are highlighted. *$p < .05$, **$p < .01$, ***$p < .001$. Original *p* values in **(A, B)** are listed in Tables S8 and S9. **(C)** Differential acetylation (H3K27ac) of active regulatory elements of potentially causal genes in postmortem entorhinal cortex tissues. AD, $n = 24$; CTL, $n = 23$. Peak and peak counts were based on a public dataset (51). For genes with multiple H3K27ac peaks, peaks with $p < .05$ are shadowed in pink (top), and peak counts are presented in violin plots (bottom). Peaks are displayed on human genome assembly hg19. **(D)** Alterations of protein level of 4 captured proteins in dorsolateral prefrontal cortex tissues. AD, $n = 230$; CTL, $n = 91$. The original data are from (52) and are plotted. Lines in each violin plot in **(C, D)** represent the first, the median, and the third quartile, respectively. *p* Values in **(C, D)** were calculated by Student's *t* test (two-tailed), with *t* statistics in each figure section. A red asterisk (*) was placed before the *p* value if it was less than the Bonferroni-correction threshold: **(C)** $df = 45$ for all comparisons, $p < .005$ (.05/11); for **(D)**, $df = 319$ for all comparisons, $p < .0125$ (.05/4). AD, Alzheimer's disease; CPM, counts per million; CTL, control; scRNA-seq, single-cell RNA sequencing.


A
siNC
siMADD
MADD
170 kDa
Tubulin
55 kDa
Vector
MADD
MADD-Flag
170 kDa
Tubulin
55 kDa
siNC
siPSMC3
PSMC3
55 kDa
43 kDa
Tubulin
55 kDa
Vector
PSMC3
PSMC3-Flag
55 kDa
43 kDa
Tubulin
55 kDa
siNC
siMTCH2
MTCH2
34 kDa
Tubulin
55 kDa
Vector
MTCH2
MTCH2-Flag
34 kDa
Tubulin
55 kDa
B
normalized Aβ42/total protein
*P=8.51x10-7 t=10.70
*P=2.17x10-4 t=5.63
siNC
siMADD
Vector
MADD
normalized Aβ42/total protein
*P=1.30x10-3 t=4.41
*P=7.86x10-7 t=10.79
siNC
siPSMC3
Vector
PSMC3
normalized Aβ42/total protein
*P=1.58x10-5 t=7.73
*P=2.14x10-4 t=5.64
siNC
siMTCH2
Vector
MTCH2
C
normalized pTau-396/total protein
*P=3.39x10-3 t=3.82
*P=1.41x10-3 t=4.36
siNC
siMADD
Vector
MADD
normalized pTau-396/total protein
P=0.496 t=0.71
*P=8.85x10-5 t=6.31
siNC
siPSMC3
Vector
PSMC3
normalized pTau-396/total protein
P=0.122 t=1.69
P=0.887 t=0.15
siNC
siMTCH2
Vector
MTCH2
D
rs10742814
rs7111576
rs10838702
rs11039376
rs35624992
rs7124396
rs1542321
rs7111606
C1QTNF4
rs2293577
rs11039200
PSMC3
SPI1
rs71475909
rs10734557
NDUFS3
rs7120548
rs55876153
MTCH2
ACP2
rs11039225
rs11039149
MADD
rs2280231
rs11039159
rs71475921
rs11038913
rs10501320
rs74486166
rs12223593
rs7947450
likely causal genes
poteintal fVars
validated fVars
eQTL association only
eQTL association & chromatin interaction
validated regulation

hypertension-related endophenotypes (83). Similar results were also observed for locus 3p21.1, which conferred risk for schizophrenia and bipolar disorder (84). In the recent MPRA study, Abell *et al.* (85) found multiple fVars in tight LD, and most haplotypes showed additive effects. These studies represent only the tip of the iceberg in terms of the complexity of complex diseases, thereby dampening the enthusiasm for finding a common target for intervention.

Conclusions

In summary, we identified 24 potential fVars and 6 genes that are likely to be involved in AD pathogenesis at 11p11.2 (Figure 7D). Although multilevel omics data were integrated, especially for epigenetic data that are relatively insusceptible to LD and coexpression, gene-level analyses in the CFG may be confounded by the strong gene coexpression due to extended LD (86). Therefore, further in vitro and in vivo systematic functional evaluations (87–90) are necessary to confirm the true causal variants and genes. Except for *SPI1*, the other 6 prioritized genes at 11p11.2 were ubiquitously expressed in all brain cell types (Figure S7). Many unanswered questions, such as whether these genes work together in the same cell type, whether they each function within a specific cell type, or whether they regulate each other through cell-cell interactions, require further investigations. The deciphering of in-depth regulatory and pathogenic mechanisms of multiple causal variants and genes is needed to fully understand the underpinnings of AD development.

ACKNOWLEDGMENTS AND DISCLOSURES

This work was supported by the National Natural Science Foundation of China (Grant No. 32230021 [to Y-GY], Grant No. 82022017 [to D-FZ], and Grant No. 31970560 [to RB]); the Strategic Priority Research Program (B) of Chinese Academy of Sciences (Grant No. XDB02020003 [to Y-GY]); the Ministry of Science and Technology of China (Grant No. STI2030-Major Projects-2022ZD0213500 [to D-FZ]); the Youth Innovation Promotion Association of Chinese Academy of Sciences (to D-FZ, MX, QL, and RB); and Yunnan Province (Grant No. 202101AT070285 [to MX] and Grant No. 202003AD150009 [to RB]). Data collection was supported through funding by National Institute on Aging (Grant Nos. P30AG10161, R01AG15819, R01AG17917, R01AG30146, R01AG36836, U01AG32984, and U01AG46152); the Illinois Department of Public Health; and the Translational Genomics Research Institute.

We thank Ian Logan for language editing and 3 anonymous reviewers for helpful comments about this manuscript. We thank Professor Xingxu Huang for sharing the pCMV-ABE7.10 vector and the pGL3-U6-sgRNA-PGK-puromycin vector. We appreciate the ENCODE (Encyclopedia of DNA Elements) Consortium and the ENCODE production laboratories for the sharing of epigenetic data. The single-cell expression data of patients with AD and control individuals used in this study were provided by the Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, Illinois.

All authors report no biomedical financial interests or potential conflicts of interests.

ARTICLE INFORMATION

From the Key Laboratory of Animal Models and Human Disease Mechanisms of the Chinese Academy of Sciences & Yunnan Province and KIZ/CUHK Joint Laboratory of Bioresources and Molecular Research in Common Diseases, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, China (MX, QL, RB, YL, HL, W-BK, ZY, QZ, CS, MY, B-LX, X-JL, ML, D-FZ, Y-GY); Kunming College of Life Science, University of Chinese Academy of Sciences, Kunming, China (MX, QL, RB, YL, HL, W-BK, ZY, QZ, CS, MY, B-LX, ML, D-FZ, Y-GY); National Resource Center for Non-Human Primates, National Research Facility for Phenotypic & Genetic Analysis of Model Animals (Primate Facility), Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, China (RB, D-FZ, Y-GY); and Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming, China (RB, X-JL, D-FZ, Y-GY).

MX, QL, and RB contributed equally to this work.

Address correspondence to Yong-Gang Yao, Ph.D., at yaoyg@mail.kiz.ac.cn, or Deng-Feng Zhang, Ph.D., at zhangdengfeng@mail.kiz.ac.cn.

Received Aug 29, 2022; revised May 25, 2023; accepted May 26, 2023.

Supplementary material cited in this article is available online at https://doi.org/10.1016/j.biopsych.2023.05.020.

REFERENCES

1. 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.
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 Aβ, 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. 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.
5. Schwartzentruber J, Cooper S, Liu JZ, Barrio-Hernandez I, Bello E, Kumasaka N, *et al.* (2021): Genome-wide meta-analysis, fine-mapping and integrative prioritization implicate new Alzheimer's disease risk genes. Nat Genet 53:392–402.
6. 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.
7. Karch CM, Ezerskiy LA, Bertelsen S, Alzheimer's Disease Genetics Consortium (ADGC), Goate AM (2016): Alzheimer's disease risk polymorphisms regulate gene expression in the ZCWPW1 and the CELF1 Loci. PLoS One 11:e0148717.

Figure 7. Expressional alterations of 3 likely causal genes affected the levels of Aβ and pTau in U251-APP cells. **(A)** Western blot for U251-APP cell lysates showing the efficiency of knockdown or overexpression of *MADD*, *PSMC3*, and *MTCH2*. siRNA for each gene and siNC were used for knockdown. Empty vector (Vector) and expression vector for each gene tagged by Flag were used for overexpression. Tubulin was used as the loading control. **(B, C)** Alterations of $A\beta_{42}$ level in culture supernatant **(B)** and pTau-396 level in cell lysate **(C)** of U251-APP cells with knockdown or overexpression of *MADD*, *PSMC3*, and *MTCH2*. The $A\beta_{42}$ and pTau-396 levels were measured by enzyme-linked immunosorbent assay and were normalized to the total amount of protein in each sample. Each group had 6 biological replicates. Bars are presented as mean ± SD. *p* Values (two-tailed Student's *t* test, *df* = 10 for all comparisons) and the *t* statistics are labeled on the top of each bar. *p* Values $< .004$ after Bonferroni correction for the total number of comparisons (.05/12) are marked by red asterisks (*). **(D)** A summary of the regulatory relationships between 24 potential fVars and 6 likely causal genes at 11p11.2. The previously established causal gene *SPI1* (9,15,16) was also included. Experimentally validated regulatory relationships are represented by solid lines with arrows and others by dotted lines. Aβ, amyloid-β; eQTL, expressional quantitative trait loci; fVar, functional variant; pTau, phosphorylated tau; siNC, siRNA negative control; siRNA, small interfering RNA.

8. Wightman DP, Jansen IE, Savage JE, Shadrin AA, Bahrami S, Holland D, *et al.* (2021): A genome-wide association study with 1,126, 563 individuals identifies new risk loci for Alzheimer's disease. Nat Genet 53:1276–1282.
9. 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.
10. Hu Y, Li M, Lu Q, Weng H, Wang J, Zekavat SM, *et al.* (2019): A statistical framework for cross-tissue transcriptome-wide association analysis. Nat Genet 51:568–576.
11. Novikova G, Kapoor M, Tcw J, Abud EM, Efthymiou AG, Chen SX, *et al.* (2021): Integration of Alzheimer's disease genetics and myeloid genomics identifies disease risk regulatory elements and genes. Nat Commun 12:1610.
12. Gockley J, Montgomery KS, Poehlman WL, Wiley JC, Liu Y, Gerasimov E, *et al.* (2021): Multi-tissue neocortical transcriptome-wide association study implicates 8 genes across 6 genomic loci in Alzheimer's disease. Genome Med 13:76.
13. Chen HH, Petty LE, Sha J, Zhao Y, Kuzma A, Valladares O, *et al.* (2021): Genetically regulated expression in late-onset Alzheimer's disease implicates risk genes within known and novel loci. Transl Psychiatry 11:618.
14. Patel D, Zhang X, Farrell JJ, Chung J, Stein TD, Lunetta KL, Farrer LA (2021): Cell-type-specific expression quantitative trait loci associated with Alzheimer disease in blood and brain tissue. Transl Psychiatry 11:250.
15. 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.
16. Pimenova AA, Herbinet M, Gupta I, Machlovi SI, Bowles KR, Marcora E, Goate AM (2021): Alzheimer's-associated PU.1 expression levels regulate microglial inflammatory response. Neurobiol Dis 148: 105217.
17. Del Villar K, Miller CA (2004): Down-regulation of DENN/MADD, a TNF receptor binding protein, correlates with neuronal cell death in Alzheimer's disease brain and hippocampal neurons. Proc Natl Acad Sci USA 101:4210–4215.
18. 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.
19. 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.
20. 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.
21. 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.
22. 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.
23. 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.
24. 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.
25. 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.
26. Kim-Hellmuth S, Bechheim M, Pütz B, Mohammadi P, Nédélec Y, Giangreco N, *et al.* (2017): Genetic regulatory effects modified by immune activation contribute to autoimmune disease associations. Nat Commun 8:266.
27. GTEx Consortium (2013): The Genotype-Tissue Expression (GTEx) project. Nat Genet 45:580–585.
28. GTEx Consortium (2017): Genetic effects on gene expression across human tissues. Nature 550:204–213.
29. 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.
30. 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.
31. 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.
32. 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.
33. Wang QS, Huang H (2022): Methods for statistical fine-mapping and their applications to auto-immune diseases. Semin Immunopathol 44:101–113.
34. Uffelmann E, Huang QQ, Munung NS, de Vries J, Okada Y, Martin AR, *et al.* (2021): Genome-wide association studies. Nat Rev Methods Primers 1:59.
35. Nativio R, Lan Y, Donahue G, Sidoli S, Berson A, Srinivasan AR, *et al.* (2020): An integrated multi-omics approach identifies epigenetic alterations associated with Alzheimer's disease. Nat Genet 52:1024–1035.
36. Corces MR, Shcherbina A, Kundu S, Gloudemans MJ, Frésard 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.
37. 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.
38. Huo Y, Li S, Liu J, Li X, Luo XJ (2019): Functional genomics reveal gene regulatory mechanisms underlying schizophrenia risk. Nat Commun 10:670.
39. ENCODE Project Consortium (2012): An integrated encyclopedia of DNA elements in the human genome. Nature 489:57–74.
40. 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.
41. 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.
42. GTEx Consortium (2015): Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. Science 348:648–660.
43. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, *et al.* (2015): Tissue-based map of the human proteome. Science 347:1260419.
44. 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, *et al.* (2015): A global reference for human genetic variation. Nature 526:68–74.
45. Cooper YA, Teyssier N, Dräger NM, Guo Q, Davis JE, Sattler SM, *et al.* (2022): Functional regulatory variants implicate distinct transcriptional networks in dementia. Science 377:eabi8654.
46. 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.
47. Gaudelli NM, Komor AC, Rees HA, Packer MS, Badran AH, Bryson DI, Liu DR (2017): Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. Nature 551:464–471.
48. Niculescu AB 3rd, Segal DS, Kuczenski R, Barrett T, Hauger RL, Kelsoe JR (2000): Identifying a series of candidate genes for mania and

psychosis: A convergent functional genomics approach. Physiol Genomics 4:83–91.
49. Wu Y, Yao YG, Luo XJ (2017): SZDB: A database for schizophrenia genetic research. Schizophr Bull 43:459–471.
50. 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.
51. 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.
52. 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.
53. Sims R, Hill M, Williams J (2020): The multiplex model of the genetics of Alzheimer's disease. Nat Neurosci 23:311–322.
54. Hardy J, Bogdanovic N, Winblad B, Portelius E, Andreasen N, Cedazo-Minguez A, Zetterberg H (2014): Pathways to Alzheimer's disease. J Intern Med 275:296–303.
55. Sheng M, Sabatini BL, Südhof TC (2012): Synapses and Alzheimer's disease. Cold Spring Harb Perspect Biol 4:a005777.
56. Bartels T, De Schepper S, Hong S (2020): Microglia modulate neurodegeneration in Alzheimer's and Parkinson's diseases. Science 370:66–69.
57. Murray J, Zhang B, Taylor SW, Oglesbee D, Fahy E, Marusich MF, *et al.* (2003): The subunit composition of the human NADH dehydrogenase obtained by rapid one-step immunopurification. J Biol Chem 278:13619–13622.
58. Buzaglo-Azriel L, Kuperman Y, Tsoory M, Zaltsman Y, Shachnai L, Zaidman SL, *et al.* (2017): Loss of muscle MTCH2 increases whole-body energy utilization and protects from diet-induced obesity. Cell Rep 18:1335–1336.
59. Guna A, Stevens TA, Inglis AJ, Replogle JM, Esantsi TK, Muthukumar G, *et al.* (2022): MTCH2 is a mitochondrial outer membrane protein insertase. Science 378:317–322.
60. Bar-Lev Y, Moshitch-Moshkovitz S, Tsarfaty G, Kaufman D, Horev J, Resau JH, Tsarfaty I (2016): Mimp/Mtch2, an obesity susceptibility gene, induces alteration of fatty acid metabolism in transgenic mice. PLoS One 11:e0157850.
61. Li Q, Wang L, Tan W, Peng Z, Luo Y, Zhang Y, *et al.* (2011): Identification of C1qTNF-related protein 4 as a potential cytokine that stimulates the STAT3 and NF-κB pathways and promotes cell survival in human cancer cells. Cancer Lett 308:203–214.
62. Byerly MS, Petersen PS, Ramamurthy S, Seldin MM, Lei X, Provost E, *et al.* (2014): C1q/TNF-related protein 4 (CTRP4) is a unique secreted protein with two tandem C1q domains that functions in the hypothalamus to modulate food intake and body weight. J Biol Chem 289:4055–4069.
63. Bailey K, Rahimi Balaei M, Mehdizadeh M, Marzban H (2013): Spatial and temporal expression of lysosomal acid phosphatase 2 (ACP2) reveals dynamic patterning of the mouse cerebellar cortex. Cerebellum 12:870–881.
64. Saftig P, Hartmann D, Lüllmann-Rauch R, Wolff J, Evers M, Köster A, *et al.* (1997): Mice deficient in lysosomal acid phosphatase develop lysosomal storage in the kidney and central nervous system. J Biol Chem 272:18628–18635.
65. Livneh I, Cohen-Kaplan V, Cohen-Rosenzweig C, Avni N, Ciechanover A (2016): The life cycle of the 26S proteasome: From birth, through regulation and function, and onto its death. Cell Res 26:869–885.
66. Bult CJ, Blake JA, Smith CL, Kadin JA, Richardson JE, Mouse Genome Database Group (2019): Mouse genome database (MGD) 2019. Nucleic Acids Res 47:D801–D806.
67. Ruggiero A, Aloni E, Korkotian E, Zaltsman Y, Oni-Biton E, Kuperman Y, *et al.* (2017): Loss of forebrain MTCH2 decreases mitochondria motility and calcium handling and impairs hippocampal-dependent cognitive functions. Sci Rep 7:44401.
68. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, *et al.* (2019): STRING v11: Protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res 47:D607–D613.
69. 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.
70. 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.
71. Sanchez CG, Acker CM, Gray A, Varadarajan M, Song C, Cochran NR, *et al.* (2021): Genome-wide CRISPR screen identifies protein pathways modulating Tau protein levels in neurons. Commun Biol 4:736.
72. Amlie-Wolf A, Tang M, Way J, Dombroski B, Jiang M, Vrettos N, *et al.* (2019): Inferring the molecular mechanisms of noncoding Alzheimer's disease-associated genetic variants. J Alzheimers Dis 72:301–318.
73. Kikuchi M, Hara N, Hasegawa M, Miyashita A, Kuwano R, Ikeuchi T, Nakaya A (2019): Enhancer variants associated with Alzheimer's disease affect gene expression via chromatin looping. BMC Med Genomics 12:128.
74. Lutz MW, Sprague D, Chiba-Falek O (2019): Bioinformatics strategy to advance the interpretation of Alzheimer's disease GWAS discoveries: The roads from association to causation. Alzheimers Dement 15:1048–1058.
75. Yang J, Ferreira T, Morris AP, Medland SE, Genetic Investigation of ANthropometric Traits (GIANT) Consortium, DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium, *et al.* (2012): Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. Nat Genet 44:S369–S375.
76. 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.
77. Knight J, Spain SL, Capon F, Hayday A, Nestle FO, Clop A, *et al.* (2012): Conditional analysis identifies three novel major histocompatibility complex loci associated with psoriasis. Hum Mol Genet 21:5185–5192.
78. Chen JA, Chen Z, Won H, Huang AY, Lowe JK, Wojta K, *et al.* (2018): Joint genome-wide association study of progressive supranuclear palsy identifies novel susceptibility loci and genetic correlation to neurodegenerative diseases. Mol Neurodegener 13:41.
79. 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.
80. Rasmussen KL, Tybjaerg-Hansen A, Nordestgaard BG, Frikke-Schmidt R (2020): APOE and dementia – Resequencing and genotyping in 105,597 individuals. Alzheimers Dement 16:1624–1637.
81. Zhou X, Chen Y, Mok KY, Kwok TCY, Mok VCT, Guo Q, *et al.* (2019): Non-coding variability at the APOE locus contributes to the Alzheimer's risk. Nat Commun 10:3310.
82. 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.
83. Flister MJ, Tsaih SW, O'Meara CC, Endres B, Hoffman MJ, Geurts AM, *et al.* (2013): Identifying multiple causative genes at a single GWAS locus. Genome Res 23:1996–2002.
84. Yang Z, Zhou D, Li H, Cai X, Liu W, Wang L, *et al.* (2020): The genome-wide risk alleles for psychiatric disorders at 3p21.1 show convergent effects on mRNA expression, cognitive function, and mushroom dendritic spine. Mol Psychiatry 25:48–66.
85. Abell NS, DeGorter MK, Gloudemans MJ, Greenwald E, Smith KS, He Z, Montgomery SB (2022): Multiple causal variants underlie genetic associations in humans. Science 375:1247–1254.

86. Wainberg M, Sinnott-Armstrong N, Mancuso N, Barbeira AN, Knowles DA, Golan D, *et al.* (2019): Opportunities and challenges for transcriptome-wide association studies. Nat Genet 51:592–599.
87. Mulvey B, Lagunas T Jr, Dougherty JD (2021): Massively parallel reporter assays: Defining functional psychiatric genetic variants across biological contexts. Biol Psychiatry 89:76–89.
88. Tewhey R, Kotliar D, Park DS, Liu B, Winnicki S, Reilly SK, *et al.* (2018): Direct identification of hundreds of expression-modulating variants using a multiplexed reporter assay. Cell 172:1132–1134.
89. Kampmann M (2020): CRISPR-based functional genomics for neurological disease. Nat Rev Neurol 16:465–480.
90. Chen ZY, Zhang Y (2022): Animal models of Alzheimer's disease: Applications, evaluation, and perspectives. Zool Res 43:1026–1040.
91. 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.
92. Novakovic B, Habibi E, Wang SY, Arts RJW, Davar R, Megchelenbrink W, *et al.* (2016): β-glucan reverses the epigenetic state of LPS-induced immunological tolerance. Cell 167:1354–1368.e14.

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

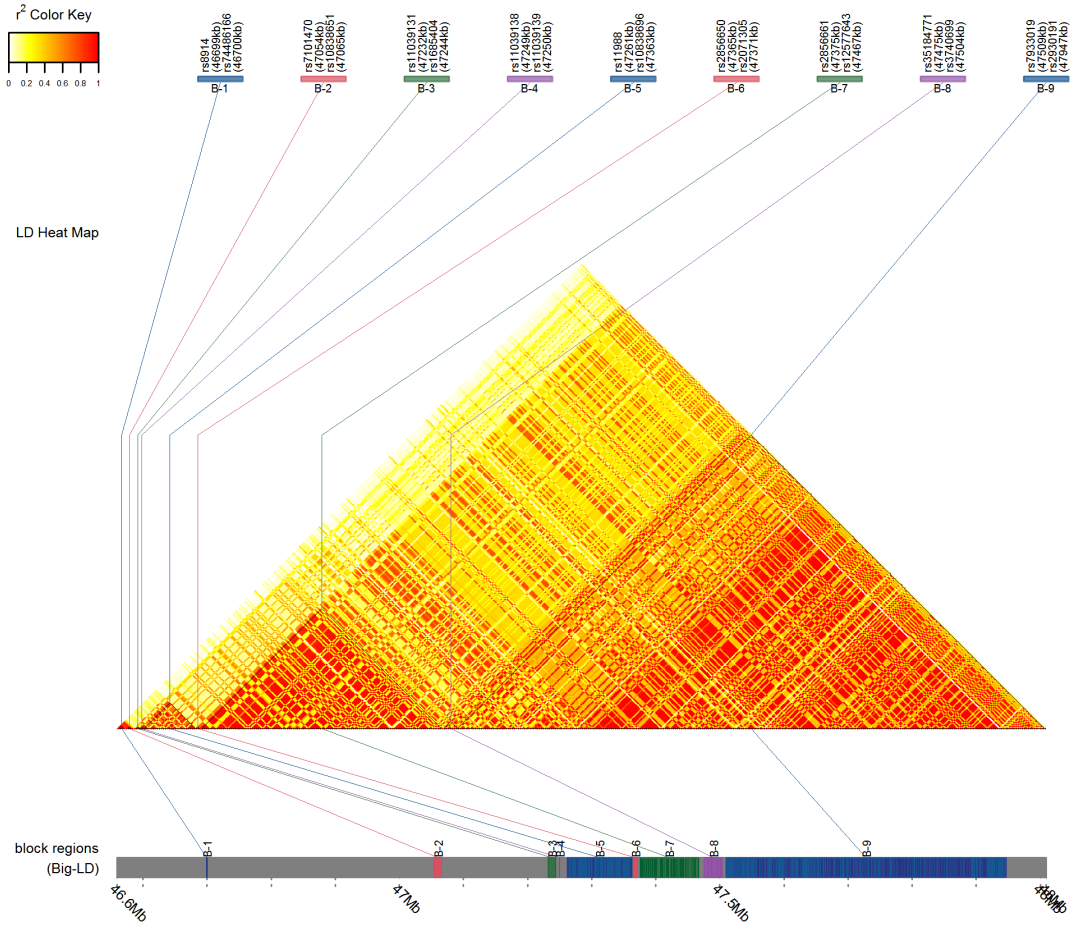


Figure S1. Linkage disequilibrium (LD) blocks at 11p11.2 detected by the bigLD. LD detection was performed by the bigLD (38) based on genotype data of 503 European individuals from the 1000 Genomes project (phase 3) (34). Each LD block (from B-1 to B-9) was represented by a different color.

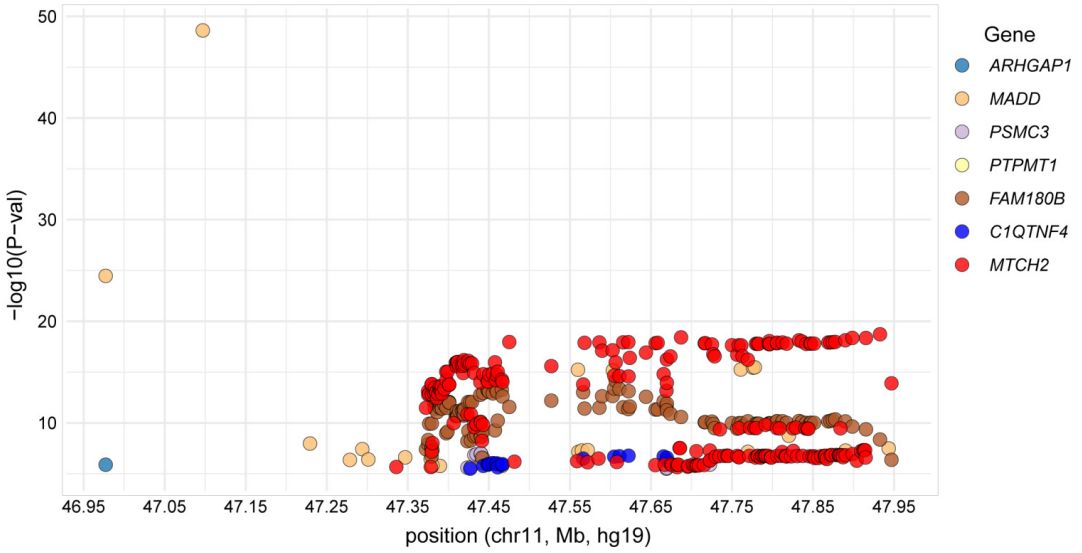


Figure S2. AD-associated SNPs at 11p11.2 were associated with expression of multiple genes in prefrontal cortex. The eQTL data were taken from the psychENCODE dataset (13). eQTLs with $P < 3.2 \times 10^{-6}$ (Bonferroni-corrected) were shown.

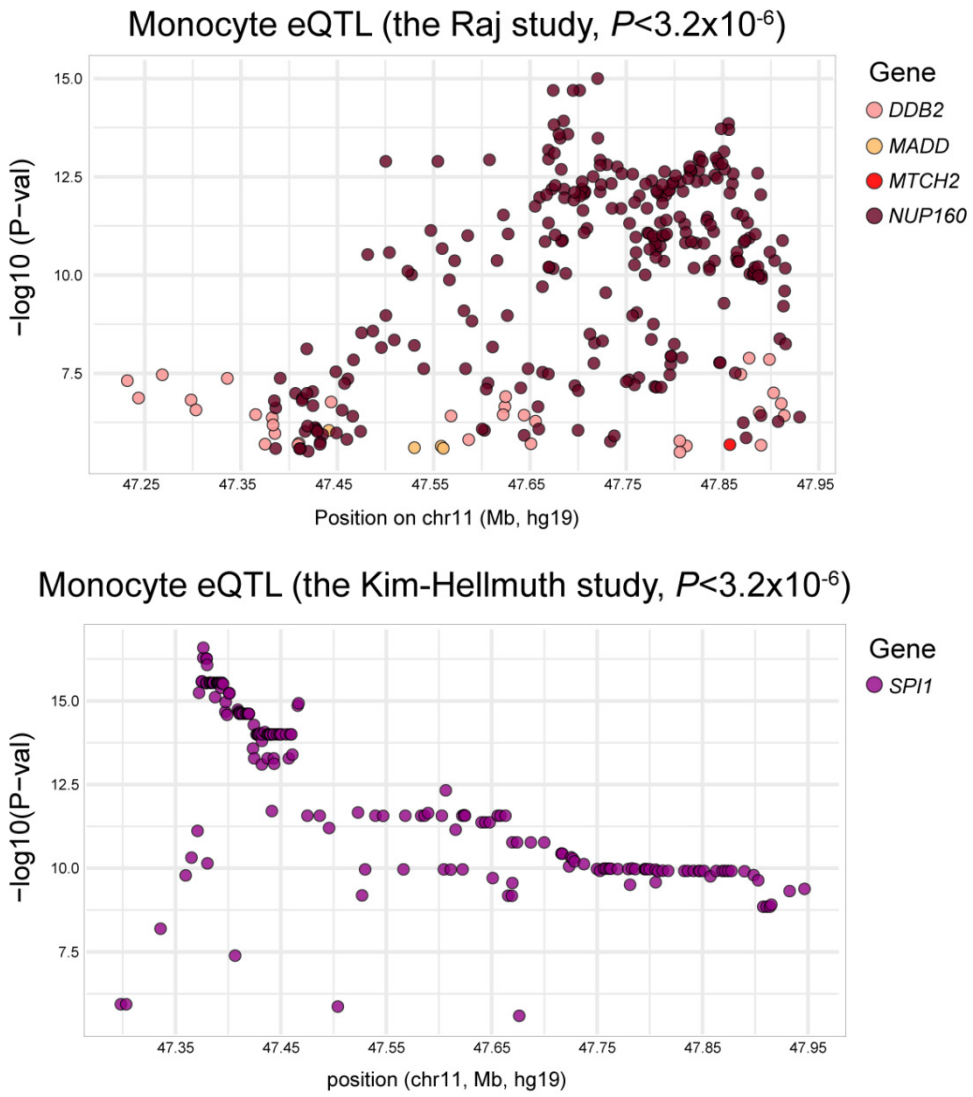


Figure S3. AD-associated SNPs at 11p11.2 were associated with expression of multiple genes in monocytes. The Raj study, monocyte eQTL dataset from Raj et al. (16); The Kim-Hellmuth study, monocyte eQTL dataset from Kim-Hellmuth et al. (15); eQTLs with $P < 3.2 \times 10^{-6}$ (Bonferroni-corrected) were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

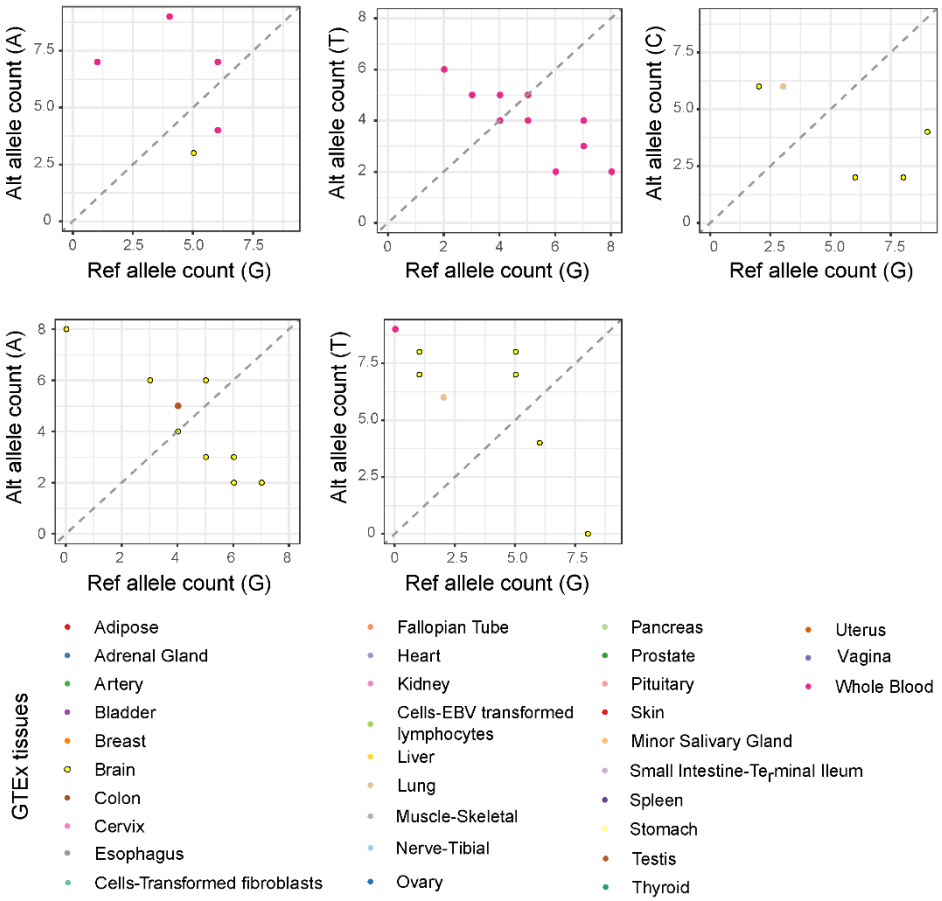


Figure S4. Potentially functional variants with a low capture rate by the GTEx ASE data (39, 40, 52)**.** Allele counts for the reference (Ref) allele and the alternative (Alt) allele were plotted for each fVar. Each dot represented an individual sample, and was colored by tissue. As the capture rates 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.

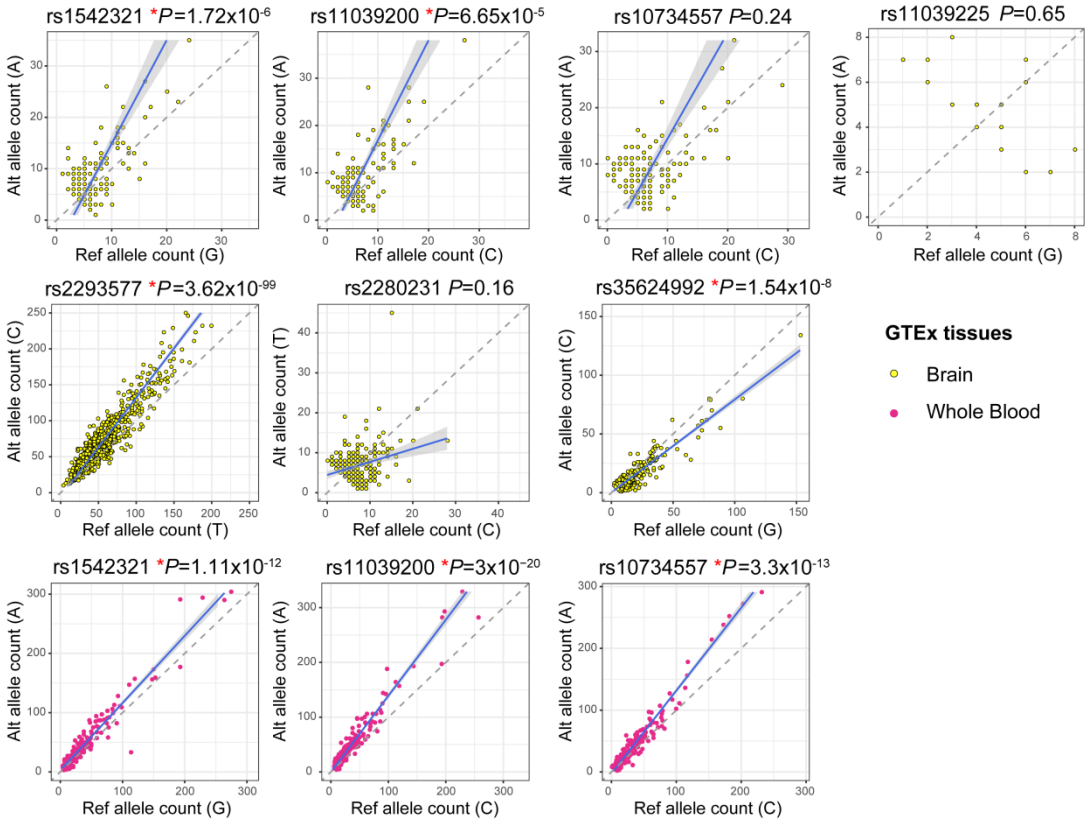


Figure S5. ASE of potentially functional variants in GTEx brain tissues and blood (39, 40, 52)**.** Allele counts for the reference (Ref) allele and the alternative (Alt) allele were plotted for each fVar. Each dot represented an individual sample, and was colored by tissue. *P* values were measured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

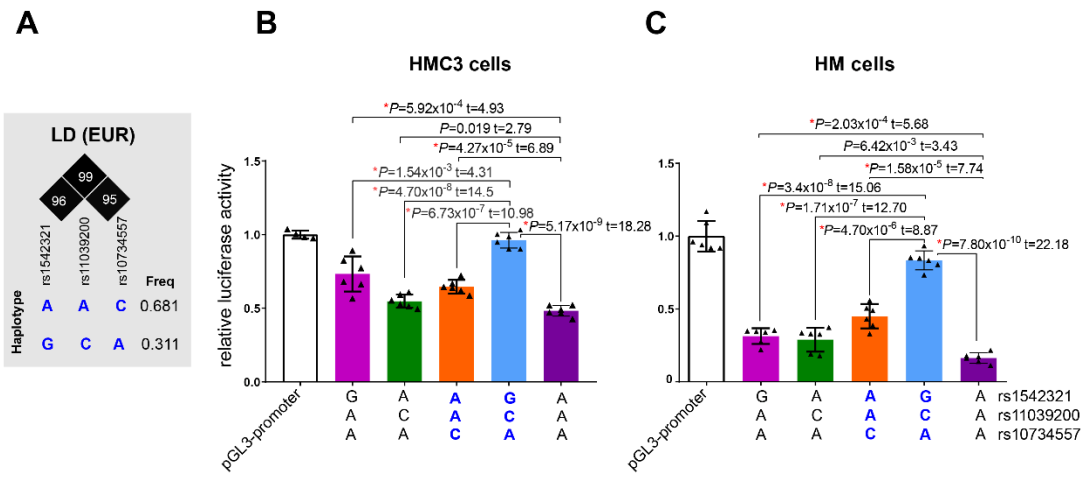


Figure S6. Three *SPI1* potential fVars showed individual and addictive effects in human microglia cell lines. (A) Linkage disequilibrium (LD) of three fVars in *SPI1* and their haplotypes. Result was performed by Haploview 4.1 (72) based on genotype data of 503 European individuals (EUR) from the 1000 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 (*).

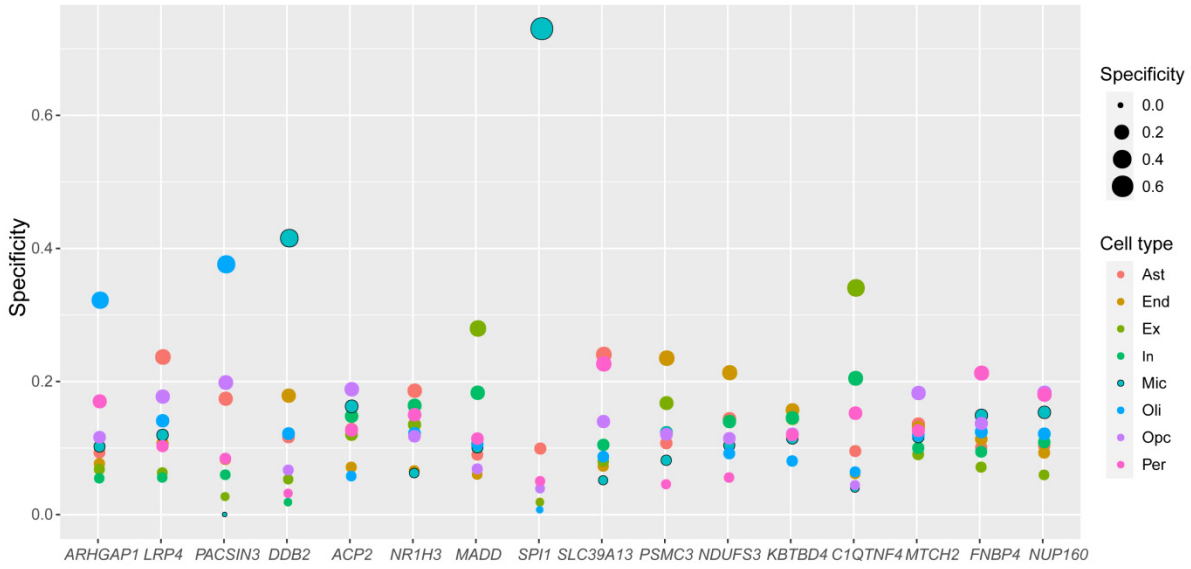


Figure S7. Cell-type expression specificity of eGenes of potentially functional variants at 11p11.2. Cell-type specificity for each gene was calculated using R package EWCE (56), based on single cell RNA-seq data from Mathys et al. (23) (frontal cortex, n = 48). Ast: astrocytes; End: endothelial cells;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

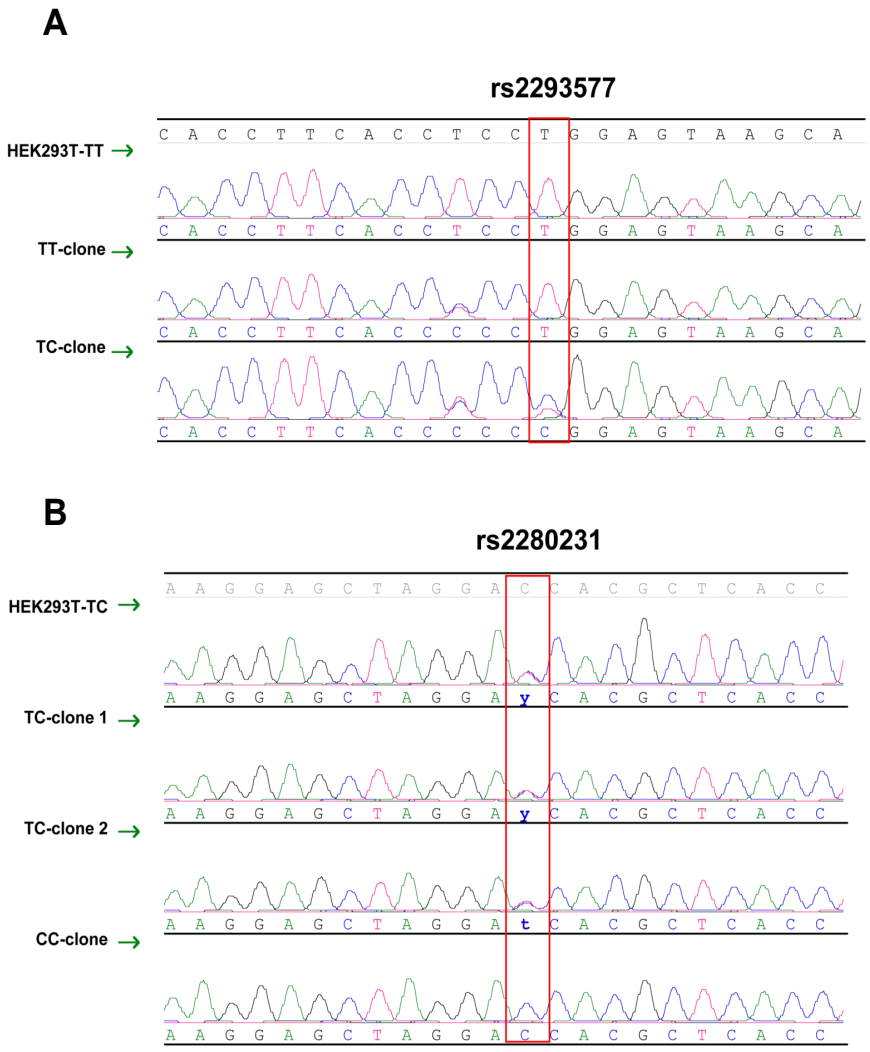


Figure S8. Sanger sequencing validated a successful base-editing of rs2293577 (A) and rs2280231 (B) in HEK293T cells. The original HEK293T cell line has a genotype TT for rs2293577 (HEK293T-TT) and TC for rs2280231 (HEK293T-TC). We obtained HEK293T cell clones with genotypes TT (TT-clone) and TC (TC-clone)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

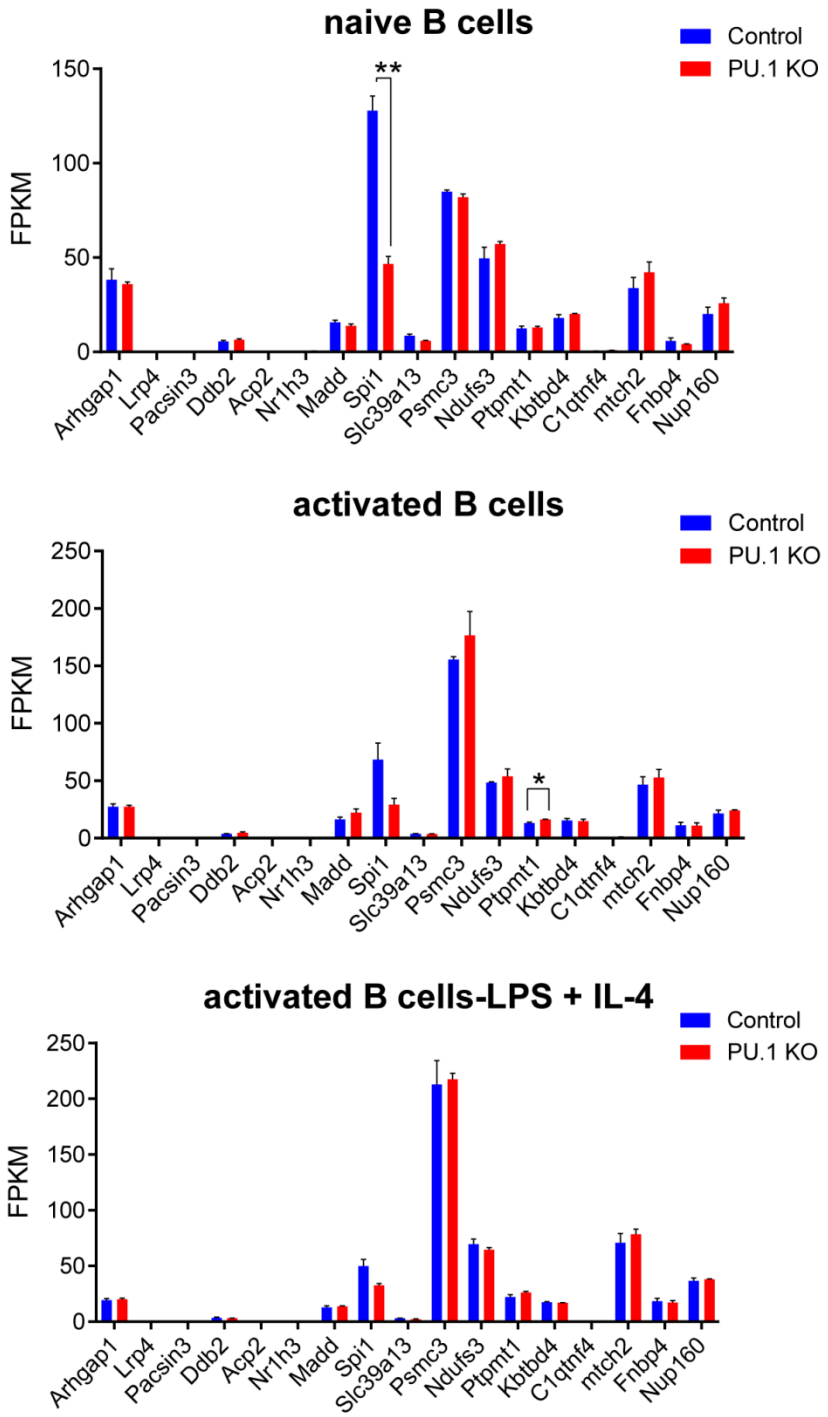


Figure S10. Expressional changes of eGenes at 11p11.2 in B cells isolated from PU.1 (*SPI1*) knockout mice (67). The *P* value was calculated by the Student's *t* test (two-tailed) to evaluate expressional difference of each gene between control and PU.1 KO cells. FPKM, fragments per kilobase of transcript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

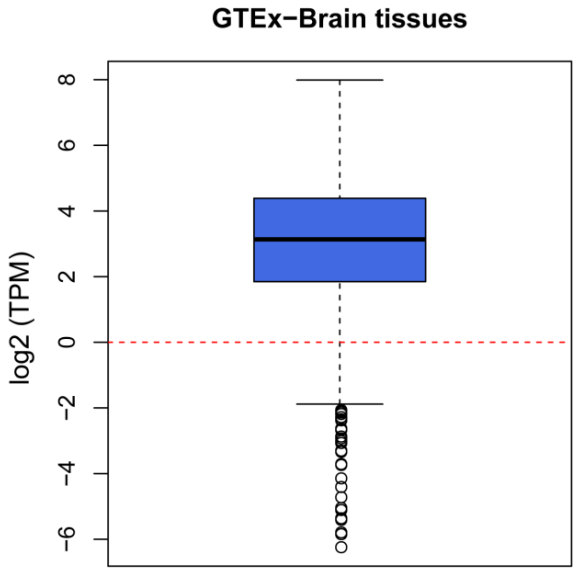


Figure S11. Mean expression of TFs in GTEx brain tissues. Original RNA-seq data of bulk brain tissues were obtained from the GTEx (39, 40), and were normalized by transcription per million (TPM).

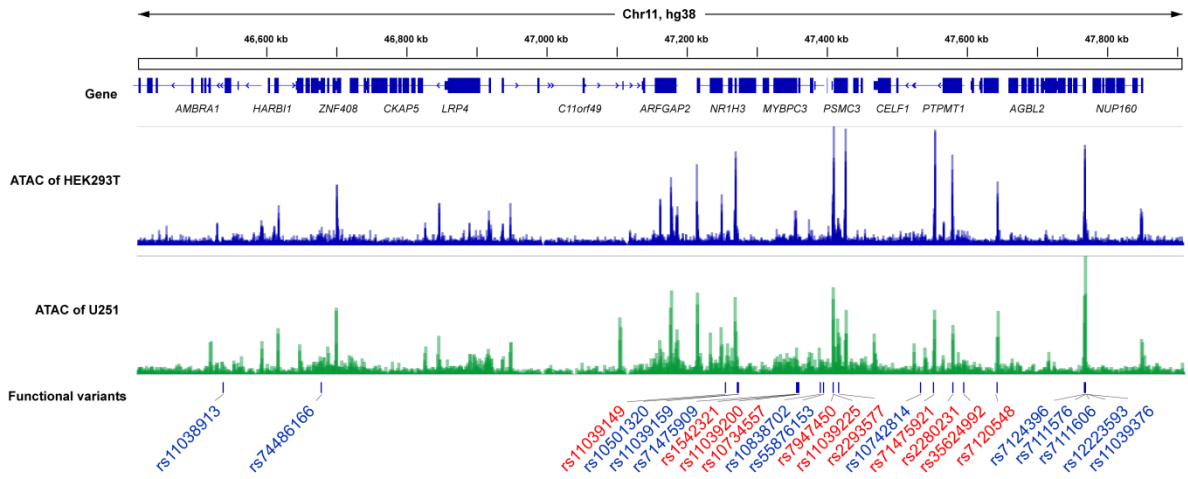


Figure S12. Open chromatin peaks in HEK293T and U251 cells at 11p11.2. Potentially functional variants (fVars) tested by the dual-luciferase reporter assays were marked in red.

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.