RESEARCH ARTICLE SUMMARY

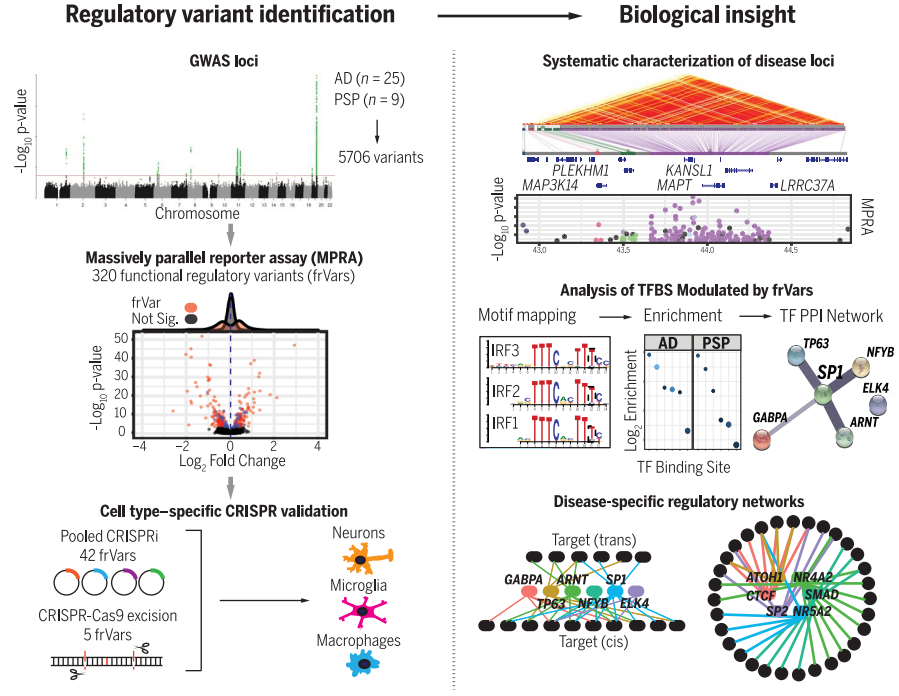
NEUROGENETICS

Functional regulatory variants implicate distinct transcriptional networks in dementia

Yonatan A. Cooper, Noam Teyssier, Nina M. Dräger, Qiuyu Guo, Jessica E. Davis, Sydney M. Sattler, Zhongan Yang, Abdulsamie Patel, Sarah Wu, Sriram Kosuri, Giovanni Coppola, Martin Kampmann, Daniel H. Geschwind*

INTRODUCTION: The widespread adoption of genome-wide association studies (GWASs) has revolutionized the detection of genetic loci associated with complex traits. However, identification of the causal variants and mechanisms underlying genotype–phenotype associations poses an enormous challenge because most common susceptibility loci reside in noncoding genomic regions and are composed of many correlated polymorphisms owing to linkage disequilibrium (LD). Massively parallel reporter assays (MPRAs) permit the high-throughput functional characterization of noncoding genetic variation, yet they have not been systematically applied to neurodegenerative disease. In this study, we used MPRA coupled with CRISPR-based validation to identify likely causal genetic variants underlying two neurodegenerative conditions that are neuropathologically linked by intracellular tau protein aggregation—Alzheimer's disease (AD) and progressive supranuclear palsy (PSP).

RATIONALE: Neurodegenerative dementias such as AD and PSP are a major cause of morbidity and mortality worldwide. There are no disease-modifying therapeutics for either disorder, motivating major efforts to characterize disease pathogenesis. Identifying causal genetic factors and downstream associated risk genes are fundamental initial steps in developing a mechanistic understanding that would enable therapeutic development.



Massively parallel functional genomics identifies genes and mechanisms in dementia. We used MPRA to screen thousands of common genetic variants associated with either AD or PSP to identify 300 likely functional regulatory variants. When coupled with CRISPR editing, this approach permitted the dissection of multiple risk loci and identification of several risk genes. The ability to view the aggregate activity of genetic risk shows convergence on disease- and cell type–specific transcriptional networks. TF, transcription factor; PPI, protein–protein interaction.

RESULTS: We tested the transcriptional regulatory activity of 5706 noncoding single-nucleotide variants, representing 25 genome-wide significant loci associated with AD and nine loci associated with PSP, by MPRA, using the neuroepithelial-like human embryonic kidney 293T (HEK293T) cell line. We identified 320 different functional regulatory variants (frVars) that affect gene expression within 27 of these loci. AD frVars were enriched within microglial enhancers, whereas PSP frVars were enriched within neuronal enhancers and, to a lesser extent, oligodendrocytes, consistent with differential cellular impact in each disease.

The majority of frVars (94%) overlapped two or more known functional annotations in human brain tissue or blood, nearly two-thirds of which were predicted to disrupt transcription factor binding, indicating their potential relevance in human disease. Forty-two high-confidence regulatory variants distributed across 15 AD loci and three PSP loci were selected for validation, using either CRISPR droplet sequencing (CROP-seq) or direct CRISPR excision in induced pluripotent stem cell–derived neurons, microglia, and astrocytes, enabling validation of 19 functional variants, implicating 20 risk genes across 11 loci.

Our data provide evidence implicating *C4A*, *PVRL2*, and *APOC1* as risk genes in AD and *PLEKHM1* and *KANSL1* as risk genes in PSP, as well as additional validation for more than a dozen other genes. MPRA-defined functional variants preferentially disrupt predicted transcription factor binding sites (TFBSs) that converge on enhancers with differential cell type–specific activity in PSP and AD, implicating a neuronal *SP1*-driven regulatory network in PSP pathogenesis.

CONCLUSION: We provide systematic characterization of common variants underlying disease risk for two distinct neurodegenerative disorders, AD and PSP. Our work illustrates the utility of integrating multiplexed reporter and CRISPR assays to efficiently characterize noncoding disease-associated variation, thereby permitting the identification of risk genes, even in complex loci with high LD. These analyses support a mechanism underlying noncoding genetic risk, whereby common genetic variants drive disease risk in aggregate through polygenic cell type–specific regulatory effects on specific gene networks. ■

The list of author affiliations is available in the full article online.
*Corresponding author. Email: dhg@mednet.ucla.edu
Cite this article as Y. A. Cooper *et al.*, *Science* **377**, eabi8654 (2022). DOI: 10.1126/science.abi8654

RESEARCH ARTICLE

NEUROGENETICS

Functional regulatory variants implicate distinct transcriptional networks in dementia

Yonatan A. Cooper[1,2,3], Noam Teyssier[4], Nina M. Dräger[4], Qiuyu Guo[3], Jessica E. Davis[5], Sydney M. Sattler[4], Zhongan Yang[3], Abdulsamie Patel[3], Sarah Wu[3], Sriram Kosuri[5], Giovanni Coppola[3,6], Martin Kampmann[4,7], Daniel H. Geschwind[1,8,9,10]*

Predicting the function of noncoding variation is a major challenge in modern genetics. In this study, we used massively parallel reporter assays to screen 5706 variants identified from genome-wide association studies for both Alzheimer's disease (AD) and progressive supranuclear palsy (PSP), identifying 320 functional regulatory variants (frVars) across 27 loci, including the complex 17q21.31 region. We identified and validated multiple risk loci using CRISPR interference or excision, including complement 4 (*C4A*) and *APOC1* in AD and *PLEKHM1* and *KANSL1* in PSP. Functional variants disrupt transcription factor binding sites converging on enhancers with cell type–specific activity in PSP and AD, implicating a neuronal SP1-driven regulatory network in PSP pathogenesis. These analyses suggest that noncoding genetic risk is driven by common genetic variants through their aggregate activity on specific transcriptional programs.

Neurodegenerative disorders such as Alzheimer's disease (AD) and progressive supranuclear palsy (PSP) are a major cause of morbidity and mortality worldwide (*1*). Given the paucity of disease-modifying therapeutics, there is a considerable need for investigation of causal disease mechanisms. Sporadic AD and PSP, both known as tauopathies because of the pathological deposition of tau in the brains of affected individuals (*2*), are complex polygenic traits with heritability estimates of between 40 and 80% (*3*, *4*) and numerous susceptibility loci identified by genome-wide association studies (GWASs) (*5*–*8*). Most of these loci overlap noncoding regions and encompass hundreds of variants that are difficult to study owing to linkage disequilibrium (LD), which hampers the identification of underlying regulatory variants and associated risk genes (*9*). Salient examples include regions harboring multiple independent association signals, such as the H1 pan-neurodegenerative risk haplotype at 17q21.31 that includes the *MAPT* locus (*10*) and the AD risk locus at 19q13.32 that harbors *APOE* (*11*, *12*).

Computational methods alone have limited predictive utility, especially in noncoding regions or regions of high LD (*13*–*15*). It is increasingly recognized that functional methods are necessary to identify true causal variants within most loci, but the sheer numbers of variants challenge most experimental approaches. Massively parallel reporter assays (MPRAs) provide a solution, enabling high-throughput experimental characterization of the transcriptional-regulatory potential of noncoding DNA elements. In an MPRA, various regulatory elements are cloned into an expression vector harboring a reporter gene and a unique DNA barcode to create an expression library that is assayed by means of high-throughput sequencing (*16*, *17*). MPRAs have prioritized functional common variation for multiple traits and disorders (*18*, *19*) but have yet to be systematically applied to neurodegenerative disease.

Results

MPRA to identify candidate regulatory GWAS variants

We conducted a staged analysis to identify regulatory variants underlying GWAS loci for two neurodegenerative tauopathies—Alzheimer's disease and progressive supranuclear palsy (Fig. 1A). First, we performed two MPRAs to screen both alleles of 5706 noncoding variants in linkage disequilibrium [squared correlation coefficient (R^2) > 0.8] with all genome-wide significant polymorphisms from two recent AD GWASs (*5*, *6*) and a PSP GWAS (*8*). In the first MPRA (MPRA 1), we tested 5223 variants encompassing 14 AD and five PSP GWAS loci. In the second MPRA (MPRA 2), we replicated 326 variants from MPRA 1 to determine assay reproducibility and screened another 483 variants encompassing 11 additional AD loci and four newly identified suggestive loci for PSP (Fig. 1, B and C; fig. S1; and Table 1) (*20*).

The performance of MPRAs to detect allelic skew depends on high library transfection efficiency, favoring the use of easy-to-transfect immortalized cell lines, such as human embryonic kidney 293T (HEK293T) cells (*21*). However, we recognized that gene regulatory elements may differ between primary and immortalized cell lines, as well as between different primary cell types. Indeed, AD-associated variants fall within open chromatin across several different neuronal and glial cell types, which are largely nonoverlapping (*6*). We explored this further using ENCODE deoxyribonuclease (DNAse) hypersensitive site (DHS) data (*22*), confirming poor DNA accessibility overlap between different brain cell types, such as astrocytes and neural progenitors (mean Jaccard index = 0.14). Notably, when we compared this overlap to that of the neuroepithelial-derived HEK293T cell line, we found that HEK293T cells had the highest mean pairwise Jaccard index (0.22; Fig. 1D) compared with active regulatory regions identified across all brain tissues. Moreover, we found that AD and PSP GWAS variants that fell within open chromatin in any brain cell type were also likely to fall within open chromatin in HEK293T cells (mean = 60%; Fig. 1E). These data indicated that HEK293T cells would provide a technically tractable model for efficient screening within a single line.

We performed both MPRAs in six independent technical replicates, obtaining activity measurements from at least five unique barcodes for both alleles of 5340 of 5706 different variants (93.6%; data S1) with median library complexities of 40 and 54 barcodes per allele, respectively (fig. S2). Consistency within (fig. S2) and between the separate MPRA experiments was high, highlighting the MPRA's reproducibility [activity scores, Pearson's correlation coefficient (r) = 0.98, $P < 2 \times 10^{-16}$; effect sizes, r = 0.94, $P < 2 \times 10^{-16}$; Fig. 2A]. Overall, we observed that ~19% of library elements were transcriptionally active (fig. S3A), in concordance with previous estimates (*18*, *19*). Relative to nonactive elements, transcriptionally active elements were enriched within genomic annotations (HEK293T cells), including DHS (open and active chromatin), H3K27Ac (active enhancers and promoters), H3K4Me3 (active promotors), and chromatin immunoprecipitation sequencing (ChIP-seq) peaks

[1]Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, CA 90095, USA. [2]Medical Scientist Training Program, David Geffen School of Medicine, University of California, Los Angeles, CA 90095, USA. [3]Center for Neurobehavioral Genetics, Jane and Terry Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, CA 90095, USA. [4]Institute for Neurodegenerative Diseases, University of California, San Francisco, CA 94158, USA. [5]Department of Chemistry and Biochemistry, University of California, Los Angeles, CA 90095, USA. [6]Department of Psychiatry and Biobehavioral Sciences, David Geffen School of Medicine, University of California, Los Angeles, CA 90095, USA. [7]Department of Biochemistry and Biophysics, University of California, San Francisco, CA 94143, USA. [8]Program in Neurogenetics, Department of Neurology, University of California, Los Angeles, CA 90095, USA. [9]Center for Autism Research and Treatment, Jane and Terry Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, CA 90095, USA. [10]Institute of Precision Health, University of California, Los Angeles, CA 90095, USA.
*Corresponding author. Email: dhg@mednet.ucla.edu

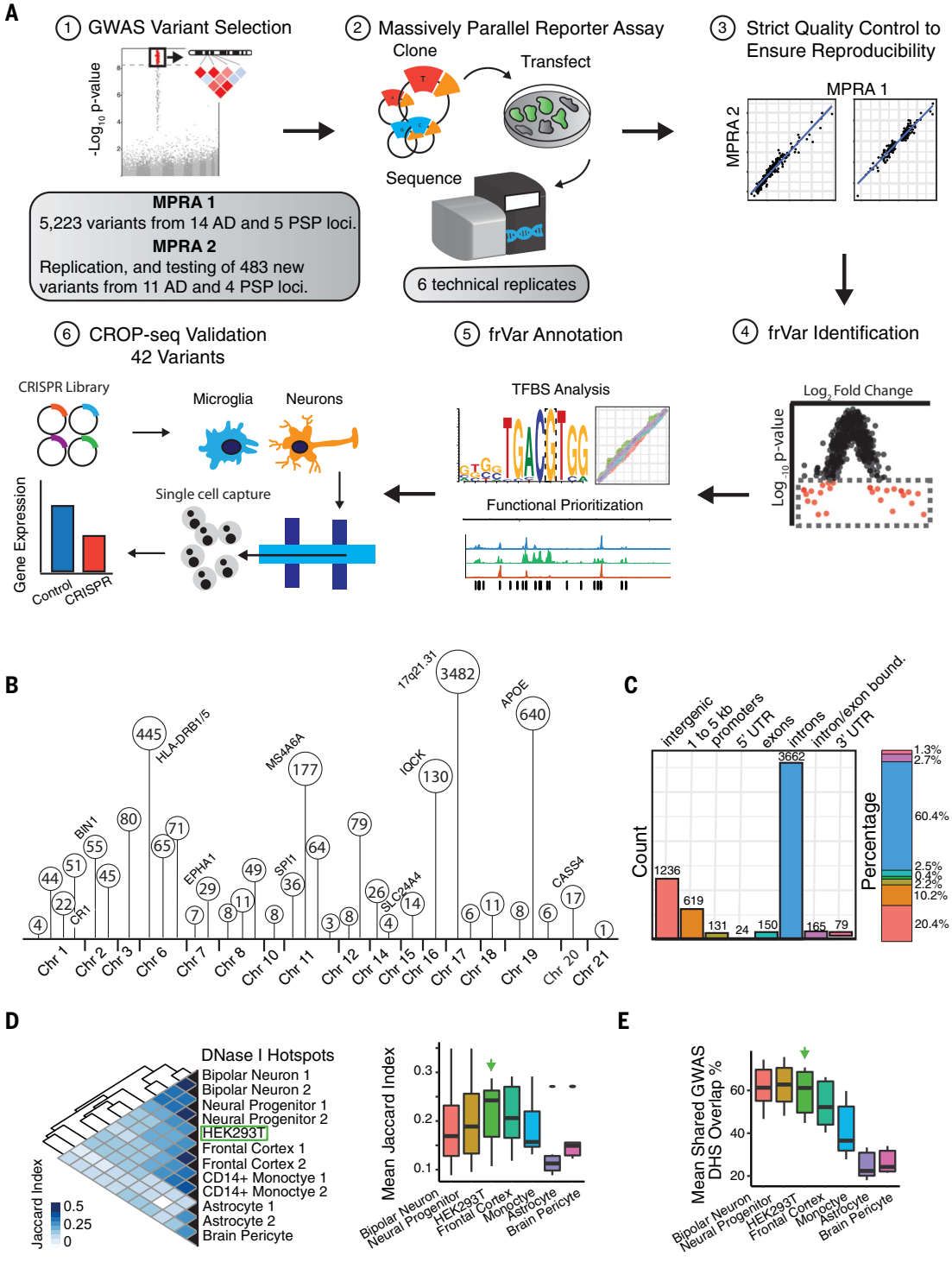


Fig. 1. Project overview. (**A**) Workflow: (1) 5223 genome-wide significant variants and LD partners encompassing 14 AD and five PSP GWAS loci were selected in MPRA 1. For MPRA 2, select variants identified in MPRA 1 were replicated. An additional 483 variants from 11 AD and four PSP loci were also tested. (2) Both alleles of each variant were barcoded and cloned into an expression library that was transfected into HEK293T cells. Allele expression was quantified by next-generation sequencing of associated barcodes. (3) Strict quality control was performed to confirm assay reproducibility. (4) Identification of variants with significant allele-specific transcriptional skew (frVars). (5) FrVars were further prioritized using brain-specific genomic annotations, and (6) 42 prioritized frVars were validated using a pooled CRISPRi screen (CROP-seq) in iPSC-derived neurons and microglia in addition to five loci validated by direct CRISPR excision. (**B**) The number of variants screened per GWAS locus and their chromosomal location. (**C**) The bar charts (counts and percentages) show genomic annotations for 5223 variants tested in MPRA 1. (**D**) (Left) Heatmap quantifying overlap (Jaccard index) of DNase hotspots between primary brain cell types and HEK293T cells. (Right) Boxplot displaying the mean pairwise Jaccard index for each cell type. (**E**) Tested GWAS variants were overlapped with DNase hotspots from each cell type. The boxplot displays the mean pairwise probability that a GWAS variant within open chromatin in a given cell type also overlaps open chromatin in another cell type. Green arrows highlight HEK293T. Error bars, SEM.

for specific transcription factors (fig. S3B), consistent with their likely functionality. We also identified significant enrichment of transcription factor binding sites (TFBSs) within active elements, including *SP/KLF*, *ETS*, and *AP-1* family members, and did not find major differences in TFBS enrichment when stratifying our results by disorder (fig. S3C).

We applied a stringent statistical threshold to define functional regulatory variants (frVars)—those with significantly different transcriptional efficacy between alleles [two-sided Mann-Whitney *U* test, false discovery rate (FDR) $q < 0.01$, Benjamini-Hochberg (BH)], identifying 320 different frVars distributed across 17 chromosomes (6.0%; Fig. 2B) within 27 of 34 tested GWAS loci, with a median of two frVars per locus (Fig. 2C and Table 1). As expected, effect sizes were generally modest (mean absolute $\log_2$ fold change = 0.53; Fig. 2D). FrVars were highly enriched within library elements that were transcriptionally active [odds ratio (OR) = 6.2; $P < 2 \times 10^{-16}$] (*20*) and within open chromatin from major

Table 1. AD and PSP GWAS loci. Description of GWAS loci and variants tested in this study. Shown are the locus lead SNPs and annotated genes, as described in the listed GWASs. For each locus, the table shows: the MPRA stage in which it was tested, the number of variants tested per locus, and the number of functional regulatory variants with significant allelic skew (FDR $q < 0.01$; frVars) ultimately identified. Chr, chromosome; Pos, position.

Chr	Pos/cytoband	Lead SNP	Annotated gene	GWAS*	MPRA stage	No. of SNPs tested	frVars
1	85603051	rs114573015	*WDR63*	PSP	2	4	0
1	180993146	rs57113693	*STX6*	PSP	1	44	2
1	207692049	rs6656401	*CR1*	AD 1	1	22	2
1	221995092	rs12125383	*DUSP10*	PSP	2	51	3
2	127892810	rs6733839	*BIN1*	AD 1	1	55	7
2	234068476	rs35349669	*INPP5D*	AD 1	2	45	2
3	39510287	rs10675541	*MOBP*	PSP	1	80	2
6	32578530	rs9271192	*HLA-DRB5– HLA-DRB1*	AD 1	1	445	13
6	45499614	rs35740963	*RUNX2*	PSP	1	65	1
6	47487762	rs10948363	*CD2AP*	AD 1	1	71	3
7	100004446; 100091795	rs1476679; rs12539172	*ZCWPW1; NYAP1*	AD 1 and 2	2	7	0
7	143110762	rs11771145	*EPHA1*	AD 1	1	29	3
8	27195121	rs28834970	*PTK2B*	AD 1	1	8	2
8	27467686	rs9331896	*CLU*	AD 1	1	11	1
8	131075859	rs2045091	*ASAP1*	PSP	2	49	2
10	11720308	rs7920721	*ECHDC3*	AD 2	2	8	1
11	47557871; 47380340	rs10838725; rs3740688†	*CELF1/ SPI1-PU.1*	AD 1 and 2	2	36	5
11	59923508	rs983392	*MS4A6A*	AD 1	1	177	4
11	85867875	rs10792832	*PICALM*	AD 1	1	64	0
11	121435587	rs11218343	*SORL1*	AD 1	1	3	0
12	21314281	rs7966334	SLCO1A2	PSP	1	8	0
12	53788003	rs147124286	*SP1*	PSP	2	79	5
14	53400629; 53391680	rs17125944; rs17125924	*FERMT2*	AD 1 and 2	2	26	2
14	92926952; 92932828	rs10498633; rs12881735	*SLC24A4/RIN3*	AD 1 and 2	1	4	1
15	59045774	rs593742	*ADAM10*	AD 2	2	14	3
16	19808163	rs7185636	*IQCK*	AD 2	2	130	17
17	17q21.31	NA	*MAPT*	PSP	1	3482	194
17	61538148	rs138190086	*ACE*	AD 2	2	6	1
18	29088958	rs8093731	*DSG2*	AD 1	1	11	2
19	1063443	rs4147929	*ABCA7*	AD 1	1	8	0
19	19q13.32	NA	*APOE/TOMM40*	AD 1	1	640	37
19	51727962	rs3865444	*CD33*	AD 1	2	6	0
20	55018260; 54997568	rs7274581; rs6024870	*CASS4*	AD 1 and 2	2	17	4
21	28156856	rs2830500	*ADAMTS1*	AD 2	2	1	1

*Source GWAS for tested loci and variants: PSP, (*8*); AD 1, (*5*); AD 2, (*6*). †Locus remapped in AD GWAS 2. Lead SNPs are not in LD. Both lead SNPs were tested here.

brain cell types, as well as monocytes, which derive from the same lineage as microglia (Fig. 2E). However, when we separated frVars derived from AD versus PSP GWAS loci and identified those that were found to overlap enhancer marks from recent work in purified nuclei from postmortem human brain (*23*), we observed that the cell types affected by each disorder were distinct. A plurality of PSP frVars fell within neuronal enhancers, which was significantly above chance ($P = 8.3 \times 10^{-5}$; χ^2 test of equal proportion) (*20*). In contrast, a plurality of AD frVars fell within microglial enhancers ($P = 0.08$), in concordance with recent estimates of cell type–specific enrichment in single-nucleotide polymorphism (SNP)–based heritability for these two disorders (*24*) (Fig. 2F). Across both disorders, 55/78 (71%) of these variants overlapped with cell type–specific enhancer annotations identified in postmortem human brain (Fig. 2G), further supporting the correspondence of these findings with regulatory elements active in human brain tissue. Of note, we found functional predictions from four popular computational algorithms to have poor concordance with regulatory annotations identified in this study and another published MPRA dataset (fig. S4) (*18*, *20*).

Refinement of frVar annotations for high-confidence prediction of regulatory variants

We next sought to determine the extent to which empirical predictions derived from HEK293T cells translate to the regulatory landscape of the human brain. We characterized frVars using functional annotations from postmortem human brain tissue (*23*, *25*, *26*), brain and microglial- or monocyte-specific expression quantitative trait loci (eQTLs) (*27*–*29*), predicted disruption of TFBSs (*30*, *31*), and brain-specific chromatin interaction data (*28*) (data S2). Blood eQTLs (*27*) were also assessed as a more powered approximation of microglia eQTLs, for which study sample sizes are relatively small.

Of 320 frVars, all except one (319/320, 99.7%) had at least one functional regulatory annotation in brain, and the majority (301, 94%) overlapped two or more functional annotations (Fig. 2, H and I, and data S2): 270 (84%) were brain, microglial, or blood eQTLs; 238 (74%) overlapped transcriptionally active regions in human brain; 88 (28%) fell within a promoter or enhancer in at least one brain cell type; and 200 (63%) significantly (all $P < 0.05$) (*20*) altered transcription factor binding. This indicates that most of the regulatory variants identified by our assay likely have an

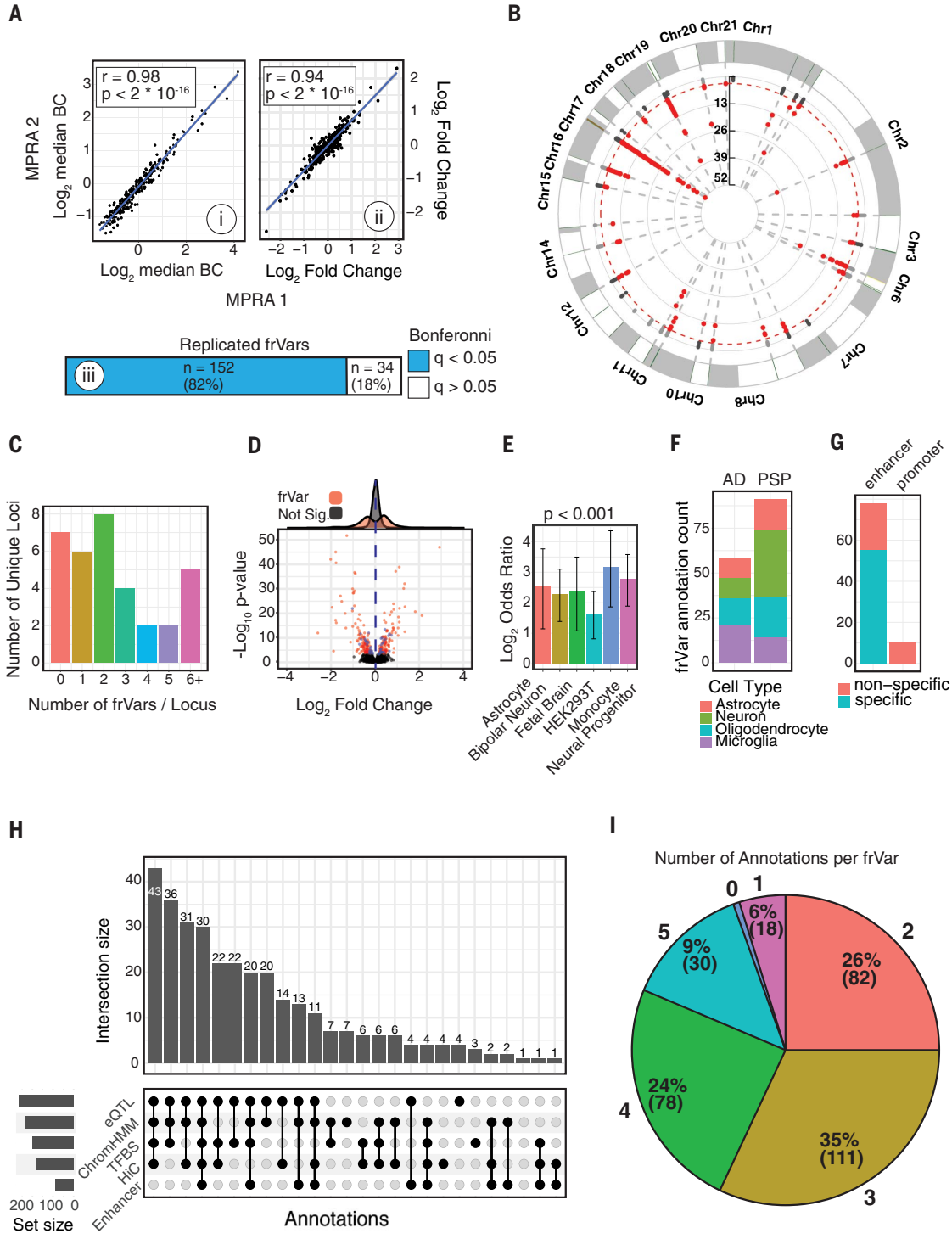


Fig. 2. Identification of MPRA functional regulatory variants. (**A**) Reproducibility of MPRA across experimental stages. A total of 326 variants, including 186 frVars identified from the preliminary assay (MPRA 1), were retested in a follow-up assay (MPRA 2). (i) Reference allele transcriptional activity and (ii) $\log_2$ effect sizes (alternative divided by reference allele) show strong correlation (Pearson's r = 0.98, 0.94; $P < 2 \times 10^{-16}$) between experiments. BC, barcode count. (iii) One hundred fifty-two of 186 assessed frVars from MPRA 1 were replicated in MPRA 2 (Bonferroni $P < 0.05$). (**B**) Manhattan plot of 5340 different variants assayed across both experiments. Red indicates frVars at FDR-adjusted $q < 0.01$ (BH method). (**C**) Histogram of the number of frVars identified per GWAS locus LD block (median = 2). (**D**) Volcano plot shows $\log_2$ allelic skew effect sizes and $-\log_{10}$ P values for 5340 different variants tested by MPRA. FrVars for MPRA stage 1 (red) and 2 (blue) highlighted. Blue dashed line indicates median effect size. (**E**) Enrichment $\log_2$ odds ratios (Fisher's exact test) of frVars within DHSs of various brain cell types (ENCODE project). Error bars, 95% confidence interval (CI); all FDR-adjusted $q < 0.001$. (**F** and **G**) FrVars were separated into those derived from AD or PSP GWAS and annotated for overlap with enhancer marks from sorted brain tissue (*23*). (F) Bar plot shows cell type–specific enhancer annotation counts for overlapping frVars separated by disease. A plurality of AD frVars fell within microglial enhancers, whereas PSP frVars fell within neuronal enhancers. (G) Most frVars overlap enhancers present in only one cell type. (**H**) Upset plot shows the number of frVars (bars) overlapping combinations (dots and lines below bars) of different functional genomic annotations. Marginal probability of each specific annotation shown in the graphic (lower left). (**I**) Pie chart depicts frVars binned by total number of overlapping genomic annotations (percentages and counts).

active regulatory role in human brain tissue. We used these annotations to further prioritize the highest-likelihood regulatory variants and nominate downstream cognate target genes.

Experimental validation of frVars

On the basis of these functional annotations, we selected 42 high-likelihood regulatory variants for additional validation, choosing those that reside within regulatory regions active in brain tissue [Table 2; selection rationale (*20*)]. These variants were distributed across 15 distinct AD loci (*APOE*, *BIN1*, *CASS4*, *CELF1/SPI1*, *CLU*, *CR1*, *ECHDC3*, *EPHA1*, *FERMT2*, *HLA-DRB1/5*, *INPP5D*, *IQCK*, *MS4A*, *PTK2B*, and

Table 2. CRISPR validation of 42 frVars. Summarized results for the 42 frVars validated using CRISPR-Cas9–mediated editing. MPRA frVars are listed with log_2 fold changes. Predicted disruption of transcription factor binding sites (TFBSx) using the union of two algorithms (*30*, *31*) is shown. The method used to validate each variant is shown: (i) pooled CRISPRi screen (CROP-seq), (ii) CRISPR enhancer excision assay (CRISPR-ex), or (iii) "None" if the variant failed validation. FDR-adjusted (Benjamini-Hochberg method) *q*-values for CRISPR experiments shown. All CRISPR $q < 0.05$.

rsID	chr:pos_A0_A1	Locus	Disorder	MPRA log_2 FC	TFBSx	Validation method	Gene_cell type	CRISPR *q*-value
rs6701713	chr1:207786289_A_G	*CR1*	AD	1.33	CTCF, ATOH1	CROP-seq	*CR1*_Microglia	0.02
rs7920721	chr10:11720308_A_G	*ECHDC3*	AD	0.65	–	None	–	–
rs10838726	chr11:47568344_C_G	*CELF1/SPI1*	AD	0.17	–	CROP-seq	*CELF1*_Microglia; *SPI1*_Microglia	1.5×10^{-4}; 0.02
rs12223593	chr11:47789082_T_G	*CELF1/SPI1*	AD	−0.46	CEBP, NFY, YBX1	CROP-seq	*FNBP4*_Neuron	9.00×10^{-3}
rs667897	chr11:59936979_A_G	*MS4A*	AD	−1.01	NFE2, MAFK	CROP-seq	*MS4A4E*_Microglia	3.00×10^{-3}
rs636317	chr11:60019150_C_T	*MS4A*	AD	−1.95	CTCF, ATOH1	CROP-seq/ CRISPR-ex	*MS4A4E*_Microglia, *MS4A6A*_THP1	0.02; 0.04
rs10876450	chr12:53811034_T_C	*SP1*	PSP	−0.40	ESR1, GATA5	None	–	–
rs7144029	chr14:53333616_C_A	*FERMT2*	AD	0.27	PRRX2, DLX3	None	–	–
rs17125924	chr14:53391680_A_G	*FERMT2*	AD	0.61	ELF3, ETV7	None	–	–
rs36026988	chr14:92938382_T_C	*SLC24A4/RIN3*	AD	0.40		CRISPR-ex	*RIN3*_Microglia	0.01
rs9935063	chr16:19710702_A_G	*IQCK*	AD	0.18	ARNT2	CROP-seq	*KNOP1*_Microglia	0.01
rs11646891	chr16:19711351_C_T	*IQCK*	AD	−0.24		CROP-seq	*GPRC5B*_Neuron; *IQCK*_Neuron; *KNOP1*_Neuron	1.6×10^{-89}; 1.9×10^{-14}; 9×10^{-3}
rs3829539	chr16:19722366_T_C	*IQCK*	AD	−0.29	–	None	–	–
rs4782272	chr16:19729016_G_C	*IQCK*	AD	−0.43	STAT3	None	–	–
rs13331873	chr16:19749709_A_G	*IQCK*	AD	0.24	GLIS3, PLAG1	None	–	–
rs4792846	chr17:43367152_G_A	*17q21.31*	PSP	0.45	–	None	–	–
rs112275793	chr17:43563304_A_C	*17q21.31*	PSP	0.60	–	None	–	–
rs111392251	chr17:43568928_T_A	*17q21.31*	PSP	−0.39	IRF1-3	CROP-seq	*PLEKHM1*_Neuron	0.01
rs76594404	chr17:43971457_G_C	*17q21.31*	PSP	0.16	EGR1, USF	CROP-seq	*MAPT*_Neuron; *MAPT*_Microglia	9.7×10^{-18}; 6.8×10^{-4}
rs76324150	chr17:43973233_C_T	*17q21.31*	PSP	0.55	ZFX	CROP-seq	*MAPT*_Neuron	1.50×10^{-3}
rs80346216	chr17:43974476_G_T	*17q21.31*	PSP	0.79	–	CROP-seq	*MAPT*_Neuron	0.04
rs7212857	chr17:44132301_C_A	*17q21.31*	PSP	−1.79	SP1	None	–	–
rs2532404	chr17:44302881_C_T	*17q21.31*	PSP	0.49	IRF4	CROP-seq	*KANSL1*_Neuron	1.30×10^{-3}
rs2732652	chr17:44345114_C_T	*17q21.31*	PSP	−0.46	PAX5	None	–	–
rs9912530	chr17:44836302_T_C	*17q21.31*	PSP	−1.76	–	None	–	–
rs2927437	chr19:45241638_A_G	*APOE*	AD	0.47	MAZ, TP73	None	–	–
rs11666329	chr19:45354296_A_G	*APOE*	AD	−0.33	SP2	None	–	–
rs412776	chr19:45379516_G_A	*APOE*	AD	1.05	GLI3	None	–	–
rs12972156	chr19:45387459_C_G	*APOE*	AD	−0.42	TFAP	None	–	–
rs10423208	chr19:45453656_G_A	*APOE*	AD	−0.24	HNF4	CROP-seq	*APOC1*_Neuron	0.04
rs117316645	chr19:45458212_G_A	*APOE*	AD	−0.93	–	CROP-seq	*APOC1*_Neuron	1.50×10^{-6}
rs13025717	chr2:127886158_C_T	*BIN1*	AD	−1.77	KLF1, KLF4, SP1	CRISPR-ex	*BIN1*_Microglia	<0.0001
rs36181881	chr2:234055276_A_G	*INPP5D*	AD	0.27		None	–	–
rs6064392	chr20:54984768_G_T	*CASS4*	AD	−1.03	ATF6	None	–	–
rs9271171	chr6:32577907_C_T	*HLA-DRB1/5*	AD	−0.35	–	CRISPR-ex	*C4A*_Astrocyte; *C4A*_THP1	0.04; 0.003
rs2790095	chr6:45439307_C_G	*RUNX2*	PSP	0.22	TCF3	None	–	–
rs75045569	chr7:143109208_T_G	*EPHA1*	AD	−0.76	RUNX	None	–	–
rs10224310	chr7:143126994_A_C	*EPHA1*	AD	−0.35	SPI1	CRISPR-ex	*EPHA1*_Microglia; *TAS2R60*_Microglia	<0.0001; 0.02
rs755951	chr8:27226790_A_C	*PTK2B*	AD	0.76	CTCF	None	–	–
rs17057051	chr8:27227554_A_G	*PTK2B*	AD	0.48	HBP1	None	–	–
rs1532277	chr8:27466181_T_C	*CLU*	AD	0.26	ARNT	CROP-seq	*CLU*_Neuron	2.30×10^{-3}
rs2045091	chr8:131075859_C_T	*ASAP1*	PSP	−0.25	NFE2, MAFG	None	–	–

SLC24A4/RIN3) and three distinct PSP loci (17q21.31, *ASAP1*, and *RUNX2*). We used two CRISPR-Cas9–mediated strategies to assay regulatory regions in their native genomic context within neurons, microglia, and astrocytes derived from human induced pluripotent stem cells (iPSCs). The first leveraged a pooled CRISPR interference (CRISPRi) platform in neurons (*32*) and microglia (*33*) using CRISPR droplet sequencing (CROP-seq) (*34*) to target all 42 variants. Recognizing that the pooled screening method using previously untested guides would not effectively edit all of the targeted regions, we additionally performed direct enhancer excision on a subset (five) of

these 42 variants that were highly significant MPRA hits and whose potential regulatory role was supported by functional genomic data, such as eQTL and high-throughput chromosome conformation capture (Hi-C)—rs13025717, rs636317, rs9271171, rs36026988, and rs10224310—assaying downstream effects on gene expression using quantitative polymerase chain reaction. We identified the candidate target genes for a given variant using eQTL or chromatin interaction data based on Hi-C from the relevant neural tissue (Table 2).

For CROP-seq, each variant was targeted using two different guide RNAs, which were introduced into iPSCs that were subsequently differentiated into neurons or microglia. We captured and sequenced 174,983 single neurons and 130,137 microglia corresponding to a median of 1030 distinct neurons and 739 microglia per guide (fig. S5, A and B). Single-cell clustering and expression of marker genes confirmed effective differentiation of both cell types (fig. S5C). CROP-seq results for all surveyed variant-gene pairs for both cell types are summarized in fig. S6. In total, we were able to validate 19 variants that regulated at least one predicted gene in at least one target cell type (Table 2 and data S3), representing a validation efficiency of 45%. This validation rate was higher than the 11% enhancer validation probability defined in a previous CROP-seq study (*35*). We verified the activity of frVars at 11 separate GWAS loci, identifying regulatory relationships with multiple previously unidentified risk genes outside of the *MAPT* and *APOE* loci, including *MS4A4E*, *C4A*, *APOC1*, and *GPRC5B* for AD and *PLEKHM1* for PSP (Fig. 3, fig. S6, and Table 2).

Several canonical risk genes, including *CR1*, *CLU*, and *BIN1*, have been consistently associated with AD risk across multiple GWASs (*36*), but the causal variants and relevant cell types underlying these associations remain to be fully clarified. We were able to identify and validate functional variants regulating each of these three established risk genes. At the Clusterin (*CLU*) locus, we validated intronic variant rs1532277 as regulating *CLU* expression within neurons (Fig. 3, A and B). Although previous studies proposed additional risk genes at this locus based solely on chromatin interaction assays (*23*), we did not detect any effects of rs1532277 on the expression of distal genes *CCDC25* or *CHRNA2* (Fig. 3B).

At the complement receptor 1 locus (*CR1*), we validated the distal intronic variant rs6701713 as regulating *CR1* expression specifically within microglia, with the disease-associated allele (A) predicted to disrupt binding of *CTCF* and increase expression of *CR1* [Genotype-Tissue Expression (GTEx)] (Fig. 3, C and D; Table 2; and data S2). Within the *BIN1* locus, previous work had rigorously established rs6733839 as a causal variant acting specifically in microglia (*23*). However, two other studies proposed rs13025717 as an additional independent risk variant at the locus on the basis of allele-specific differential chromatin accessibility and chromatin interaction data (*37*, *38*). MPRA identified rs13025717 as a highly significant frVar ($P = 2.6 \times 10^{-38}$; Fig. 3E), and we experimentally confirmed that this variant regulates *BIN1* expression in microglia using CRISPR-mediated enhancer excision (Fig. 3F). We also experimentally validated regulatory variants at other canonical loci including rs10224310 and rs36026988, which we show regulate *EPHA1* and *RIN3* expression, respectively, in microglia (Fig. 3, G to J; fig. S6; and Table 2).

We also identified variants with pleiotropic regulatory activity, allowing us to identify candidate risk genes underlying AD risk. The membrane-spanning 4-domains (*MS4A*) cluster on chromosome 11 contains more than 16 genes and harbors numerous variants conferring risk for AD (*37*, *38*). Previous work at this locus implicated rs636317 as a likely causal variant and nominated *MS4A6A* as the cognate risk gene (*37*, *38*). In this study, we confirmed the regulatory function of rs636317, as well as another variant, rs667897, within the MS4A locus (Fig. 3K). Notably, we found that both variants regulate expression of *MS4A4E*, but not *MS4A4A*, specifically within microglia, with rs667897 also suggestive for regulation of *MS4A6A* (Fig. 3, L to O). Additionally, using enhancer excision in THP-1–derived macrophages, we found that rs636317 also regulates *MS4A6A* expression (fig. S7A), although this was not significant in the CRISPRi screen (Fig. 3M). The disease-associated (T) allele of rs636317 disrupts binding of transcriptional repressor *CTCF*, thereby increasing gene expression (*38*). Conversely, the disease-associated (A) allele of rs667897 creates an *NFE2-MAFG* TFBS, increasing *MS4A6A* and *MS4A4E* expression in microglia to confer AD risk (Table 2).

We next explored the *HLA-DRB1/5* AD GWAS locus, which is a highly polymorphic region of extended LD with numerous AD-associated variants and potential risk genes. MPRA identified 13 significant variants in the locus falling within the intergenic region between *HLA-DRB1* and *HLA-DQA1*. We further validated rs9271171 using enhancer excision in iPSC-derived astrocytes, because brain chromatin conformation data from PsychENCODE predicted that rs9271171 regulates the distal gene complement 4 (*C4A*), which is most highly expressed in astrocytes (*39*). Indeed, we observed that excision of this variant significantly reduced *C4A* expression in both astrocytes ($P = 0.04$; Fig. 3, P and Q) and THP-1 macrophages ($P = 0.003$; fig. S7B), highlighting *C4A* as a risk gene for AD. Other examples of loci wherein we identified variants regulating multiple downstream target genes include *IQCK* (rs9935063, rs11646891; *GPRC5B*, *IQCK*, *KNOP1*) and *SPI1* (rs10876450, rs12223593; *SPI1*, *CELF1*, *FNBP4*) (fig. S6 and Table 2). Whether all, or a subset, of the genes regulated at these loci contribute to disease risk will require individual functional validation in model systems.

Characterization of 17q21.31 and 19q13.32

The chromosome 17q21.31 locus is noteworthy for harboring the tau-encoding *MAPT* gene within a common 900-kb inversion polymorphism and is a major risk locus for PSP (H1 haplotype OR = 4 to 5) as well as AD, Parkinson's disease, and corticobasal degeneration (*10*). 17q21.31 contains complex haplotypic substructural variation and extensive LD, hampering interrogation with traditional statistical genetics approaches (*13*). We leveraged the ability to functionally dissect this region with MPRA, testing 3482 variants within 17q21.31 in strong LD with lead SNPs from a PSP GWAS (*8*), which comprise ~24% of the more than 14,000 common variants in the region, identifying a total of 194 frVars, or 5.6% of those within this region.

We next clustered these frVars according to LD (minimum $R^2 = 0.5$), identifying seven distinct LD clusters (Fig. 4A), which suggests multiple distinct loci within this region. The largest LD cluster includes 42 frVars within the core *MAPT* gene itself [−5 kb upstream of the transcriptional start site (TSS) to the 3′ untranslated region (3′UTR], of which 13 are variants within annotated enhancers predicted to disrupt TFBSs (Fig. 4B) and are in strong LD ($R^2 > 0.9$) with the H1 marker rs8070723 (data S4). We also highlight a cluster of eight contiguous frVars overlapping a large oligodendrocyte enhancer within a distal regulatory element of the *MAPT* first intron (Fig. 4B). These variants were of interest because tau pathology within oligodendrocytes is a defining pathological feature of PSP (*40*). All eight of these regulatory variants are in low LD ($D' = 1$, $R^2 = 0.18$) with rs242557, a marker SNP for the H1C subhaplotype known to confer risk for PSP (*10*) (data S4 shows pairwise LD for rs242557).

A core promoter beginning −300 bp from the TSS (*41*) has been previously characterized in a variety of cell types (*41*, *42*). As basal *MAPT* expression (in addition to splicing) is thought to be a key molecular driver of tauopathy pathogenesis (*8*), we used the functional data from MPRA to further clarify the cis-regulatory landscape of the *MAPT* gene. The MPRA element corresponding to positions −226 to −63 of the *MAPT* gene was highly transcriptionally active, whereas elements within the region −349 to −186 drove only modest expression (data S1), suggesting a core promoter starting −226 bp upstream of the *MAPT* TSS (Fig. 4C).

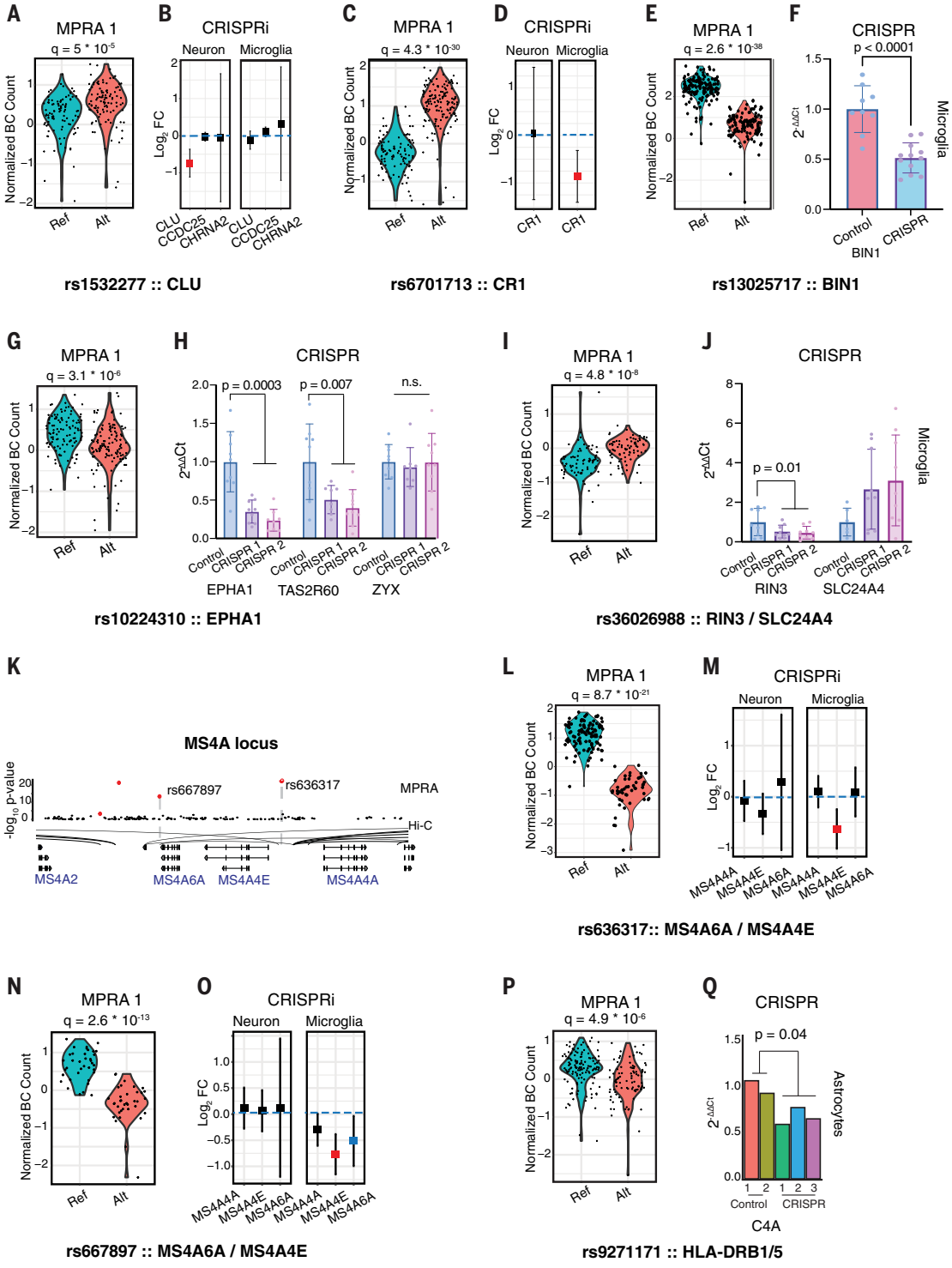


Fig. 3. CRISPR validation of eight MPRA frVars across seven AD loci. (**A** and **B**) rs1532277 in the *CLU* locus. (A) Violin plot shows MPRA normalized barcode (BC) distributions for each allele. FDR-adjusted *q*-values displayed. (B) CRISPRi targeting suppresses *CLU* (*z*-score = −3.8, $q = 2.3 \times 10^{-3}$) but not *CCDC25* or *CHRNA2* expression specifically in iPSC-derived neurons. Error bars, 95% CI. FC, fold change. (**C** and **D**) rs6701713 in the *CR1* locus. (C) MPRA data as in (A). (D) CRISPRi targeting suppresses *CR1* expression (*z*-score = −2.8, *q* = 0.04) specifically in microglia. Error bars, 95% CI. (**E** and **F**) rs13025717 in the *BIN1* locus. (E) MPRA data. (F) CRISPR-mediated excision of genomic region containing rs13025717 in microglia reduces *BIN1* expression compared with gRNA-scramble controls [control $n = 9$, CRISPR $n = 12$, $t_{19} = 5.8$; $P < 0.0001$]. Error bars, SEM. (**G** and **H**) rs10224310 in the *EPHA1* locus. (G) MPRA data. (H) CRISPR deletion of genomic region containing rs10224310 in microglia reduces *EPHA1* [control $n = 9$, CRISPR $n = 17$, $t_{24} = 6.6$; $P < 0.0001$] and intronic gene *TAS2R60* but not *ZYX* expression. Error bars, SEM. (**I** and **J**) rs36026988 in the *SLC24A4/RIN3* locus. (I) MPRA data. (J) CRISPR deletion of genomic region containing rs36026988 in microglia reduces *RIN3* [control $n = 9$, CRISPR $n = 19$, $t_{35} = 1.7$; $P = 0.05$] but not *SLC24A4* expression. (**K**) *MS4A* locus variants, highlighting rs667897 and rs636317. (**L**) rs636317 MPRA data. (**M**) CRISPRi targeting of rs636317 reduces *MS4A4E* (*z*-score = −3.0, *q* = 0.02) expression. (**N**) rs667897 MPRA data. (**O**) CRISPRi targeting of rs667897 significantly reduces *MS4A4E* (*z*-score = −3.7, *q* = 0.003) and shows a trend toward reducing *MS4A6A* expression (*z*-score = −2.2, uncorrected *P* = 0.01) in microglia. Error bars, 95% CI. (**P** and **Q**) rs9271171 in the *HLA-DRB1* locus. (P) MPRA data. (Q) CRISPR deletion of rs9271171 significantly reduces *C4A* expression in iPSC-derived astrocytes [CRISPR $n = 3$, control $n = 2$; $t_3 = 3.5$; $P = 0.04$]. All CRISPR excision assessed using two-tailed Student's *t* test comparing the averages of the control and the CRISPR conditions.

We next sought to identify functional variants likely to control *MAPT* gene expression. We chose three frVars for further experimental validation prioritized by regulatory annotations in brain and chromatin looping to the *MAPT* TSS (data S2): rs76594404, within the proximal promoter (predicted to disrupt *EGR1/USF* binding); rs76324150, which falls within the proximal first *MAPT* intron (disrupts *ZFX*); and rs80346216, within the distal first intron (Fig. 4C). These loci were confirmed using CRISPRi to regulate *MAPT* expression within iPSC-derived neurons (Fig. 4, D and E; fig. S6; and Table 2). Additionally, we identified a frVar just distal to the *MAPT* gene

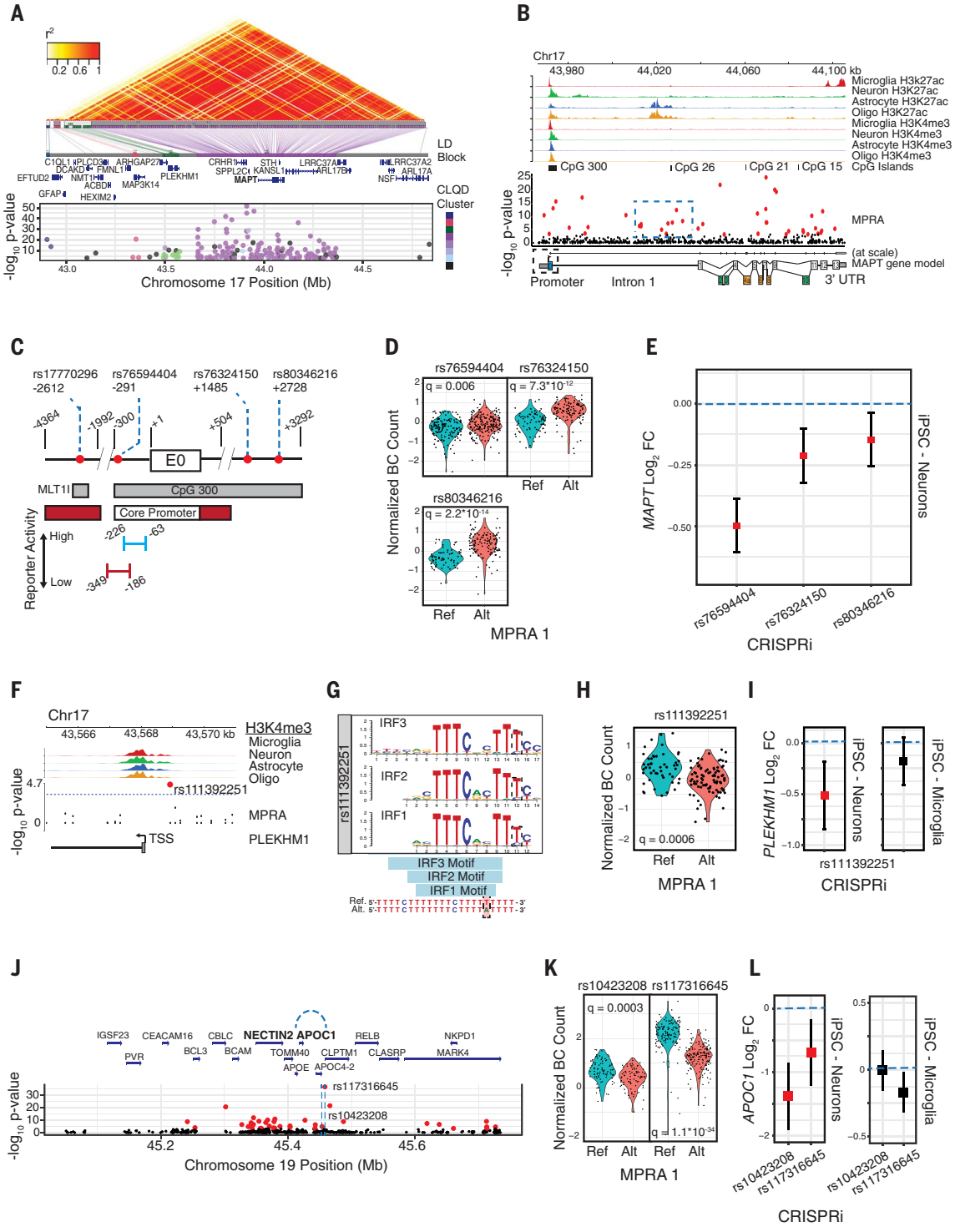


Fig. 4. Systematic characterization of 17q21.31. (**A**) (Top) 17q21.31 LD plot (1000 Genomes, CEU) above MPRA frVars clustered by LD. (Bottom) frVars plotted by position and MPRA significance ($-\log_{10} P$ values; colors indicate clusters, black indicates unclustered). Most variants fall within LD cluster 4 (purple) centering on *MAPT*. (**B**) Chromatin annotations and frVars across *MAPT*. Tracks (1 through 8) show H3K27ac (active) and H3K4me3 (promoter) ChIP-seq peaks for microglia (red), neurons (green), astrocytes (blue), and oligodendrocytes (orange) from (*23*); CpG islands (track 9); and tested variants plotted by $-\log_{10} P$ value (track 10). FrVars in red. *MAPT* gene model (track 11). Exons: blue, untranslated; white, constitutive; green, alternative neuronal; and yellow, rarely expressed in brain. Blue square shows frVars within an oligodendrocyte enhancer. (**C**) Annotation of the *MAPT* promoter [−4364 to +3292 from TSS, features from (*42*)] with frVars (red dots). Red, repressor regions. (Bottom) MPRA relative reporter activity for two elements. (**D** and **E**) Three frVars regulating *MAPT* expression. (D) MPRA violin plots show normalized allelic barcode distributions. (E) CRISPRi targeting suppresses *MAPT* expression for rs76594404 (z-score = −9.1, $q = 9.7 \times 10^{-18}$), rs76324150 (z-score = −4.0, $q = 1.5 \times 10^{-3}$), and rs80346216 (z-score = −2.8, q = 0.04) in iPSC-neurons. (**F** to **I**) rs111392251: frVar in *PLEKHM1* promoter (±3 kb). (F) Tracks (1 through 4) show H3K4me3 peaks

[as in (B)]. (G) The alternate allele of rs111392251 disrupts *IRF1-3* TFBSs. (H) MPRA data as in (D). (I) CRISPRi targeting of rs111392251 suppresses *PLEKHM1* expression (*z*-score = −3.1, q = 0.01) in neurons but not microglia. Error bars, 95% CI. (**J**) Variants within the *APOE* locus plotted by coordinate and MPRA $-\log_{10} P$ value. FrVars in red, blue dashes loop to *APOC1* promoter. (**K**) MPRA data and (**L**) CRISPRi targeting of rs10423208 (*z*-score = −5.5, $q = 1.5 \times 10^{-6}$) and rs117316645 (*z*-score = −2.8, q = 0.04) suppresses *APOC1* expression in neurons but not microglia. Error bars, 95% CI.

body, rs2532404, that we confirmed strongly regulates *KANSL1* expression in neurons, but not microglia (fig. S6).

We also identified three separate LD clusters at 17q21.31 harboring variants predicted to regulate genes besides *MAPT*. Putative risk genes overlapping regulatory variants in independent LD blocks implicated here include *MAP3K14*, *PLEKHM1*, and *LRRC37A4P* (table S1). One such cluster harbors rs111392251, a high-confidence regulatory variant located in the promoter of *PLEKHM1* that is predicted to disrupt binding of *IRF*-family TFs (Fig. 4, F to H). We used CRISPRi to validate that this locus regulates *PLEKHM1* expression within neurons but not microglia (Fig. 4I). *PLEKHM1* regulates autophagosome-lysosome formation (*43*) and has been previously suggested as a PSP risk gene on the basis of proximity (*7*), but has yet to be functionally characterized in a disease context. These results indicate that *PLEKHM1* dysregulation within human neurons contributes to PSP risk, independent of *MAPT* expression.

The *APOE* locus on 19q13.32 harbors the strongest common genetic association with late-onset AD, tagging the well-characterized *APOE4* risk haplotype (*5*, *6*). However, the extensive LD in the region has resulted in identification of hundreds of additional disease-associated variants (*11*, *12*). Recent work has uncovered evidence for *APOE*- independent risk for AD in the locus, implicating *PVRL2/NECTIN2* and *APOC1* (*11*), although some have argued that *APOE* coding variants mediate the entire association signal (*12*). We reasoned that identification of functional variants at 19q13.32 would inform our understanding of this complex regulatory architecture. We tested 640 variants in LD with the 538 genome-wide significant SNPs (*5*) at this locus and identified 37 frVars (5.8%; Fig. 4J). Through integration with whole brain and blood eQTLs, we show that 10 of these were predicted to regulate *PVRL2* expression, identifying *PVRL2* as a strong candidate risk gene at this locus (table S2). With respect to *APOC1*, we examined a distal segment of open chromatin within an intergenic region downstream from the *APOC2-APOC4* gene carrying two frVars, rs10423208 and rs117316645 (Fig. 4J), neither of which are in LD with the *APOE* haplotype tag SNPs rs7412 and rs429358 (all $R^2 < 0.01$, EUR ancestry). CRISPRi targeting of both loci robustly down-regulated *APOC1* expression specifically within neurons (Fig. 4, K and L, and Table 2). Our results provide support for *APOC1* as an AD risk gene, warranting its further study.

MPRA frVars enrich for TFBS disruption

We next hypothesized that functional variants defined by our frVars distributed across the genome would be enriched for variants that disrupt TF binding as a class, which was what we observed [enrichment OR = 1.4, P = 0.003; SNP2TFBS (*30*)]. We observed a similar magnitude of enrichment of TFBS-disrupting variants among frVars in another published dataset (*18*) (OR = 1.9, $P = 8.7 \times 10^{-8}$). Both enrichments were replicated using an alternative TFBS scoring method [motifbreakR (*31*)]. Furthermore, the magnitude and directionality of predicted TFBS disruption correlated with MPRA effect sizes for frVars in both our dataset (Spearman's rho = 0.44, $P = 2.3 \times 10^{-4}$) and the other dataset (*18*) (rho = 0.52, $P < 8.7 \times 10^{-9}$; Fig. 5, A and B). In general, we did not find that TFBS disruption alone predicts MPRA allelic skew (positive predictive value = 0.14) (*20*), although disruption of a subset of specific TFs such as *ELK4*, *ETS1*, *GABPA*, and *SP*-family members did modestly predict allelic skew captured by MPRA in these data (fig. S8).

Enrichment of functional risk variants within disease-specific transcriptional networks

We next assessed whether frVars were enriched within binding sites for specific transcription factors, and we found that TFBSs disrupted by risk variants differed by disease. In AD, CTCF ($\log_2$ OR = 4.5, $P = 3 \times 10^{-5}$), *NR4A2* ($\log_2$ OR = 4.9, P = 0.002), *NR5A2* ($\log_2$ OR = 3.6, P = 0.01), *ATOH1* ($\log_2$ OR = 3.8, P = 0.009), *SP2* ($\log_2$ OR = 2.3, P = 0.008), and *SMAD*-family (*SMAD2,3,4* heterotrimer, $\log_2$ OR = 3.7, P = 0.0002) binding sites were enriched for disrupting risk variants, which is consistent with previous work identifying enrichment of AD risk variation in microglial enhancers containing *CTCF*, *SMAD*, and *SP* binding motifs (*38*). In contrast, PSP showed a different pattern of TFBS enrichment, in which five of the six TFs predicted to be enriched for binding site disruption physically interact with the transcription factor SP1 (including SP1 itself; OR= 1.8, P = 0.007; Fig. 5C). Of note, the set of PSP-enriched TFs forms a significant protein–protein interaction (PPI) network with SP1 (STRING $P = 1.7 \times 10^{-5}$; permutation P = 0.05; Fig. 5C) (*20*). This finding is consistent with SP1's multimerization capabilities and its activity as a core component of a broad array of gene regulatory complexes that regulate tissue-specific gene expression (*44*). Notably, 11.1% of annotated PSP frVars are predicted to participate specifically within this network. We also assessed enrichments at a lower statistical threshold (TFBS enrichment $q < 0.1$), finding additional *ETS*-family TFs and SP1 interactors to be enriched within functional PSP risk variants (fig. S9). Although these TFs are highly expressed in HEK293T cells, we did not find a general relationship between relative TF abundance and MPRA allelic skew ($P > 0.05$; Spearman's correlation) (*20*), indicating that expression levels do not explain our results. Additionally, we did not identify a significant enrichment for disrupted TFBS motifs among equivalently sized random collections of GWAS variants (*20*), demonstrating that our results are not explained by the initial SNP selection.

Given these findings, we sought to verify whether our computational predictions of TFBS disruption were robust by performing electrophoretic mobility shift assays (EMSAs). We focused on SP1 because it was the core hub gene in the PPI network, testing the relative SP1 binding affinity for each of the 14 frVars predicted to disrupt SP1. binding. We identified significant differential allele-specific binding affinities in 10 of 14 variants ($q < 0.1$; Fig. 5D and fig. S10) and found a strong correlation between EMSA results and MPRA effect sizes (Fig. 5E), confirming the bioinformatic regulatory predictions of SP1 activity.

We next reasoned that we could leverage expression data from postmortem human brain tissue to garner additional evidence for dysregulation of SP1 network TFs in disease. Indeed, gene expression analysis confirmed that a majority of these network TFs, including SP1, were differentially expressed within the temporal cortex of individuals with PSP and matched controls (Fig. 5F and fig. S11A). Of note, the coding region for *SP1* itself falls within a suggestive PSP risk locus that did not reach genome-wide significance (locus combined $P = 4.1 \times 10^{-7}$) (*8*). Taken together, these data provide converging evidence for broad dysregulation of an SP1-centered transcriptional network in PSP pathophysiology.

To explore this further, we generated a two-layer directed network composed of these six PSP TFs and their likely targets, defined as genes within the cis-regulatory window of TFBS-disrupting frVars, as well as genes linked by brain Hi-C (Fig. 5G; fig. S11B for AD network) (*20*). These TFBS-disrupting variants were then annotated using cell type–specific

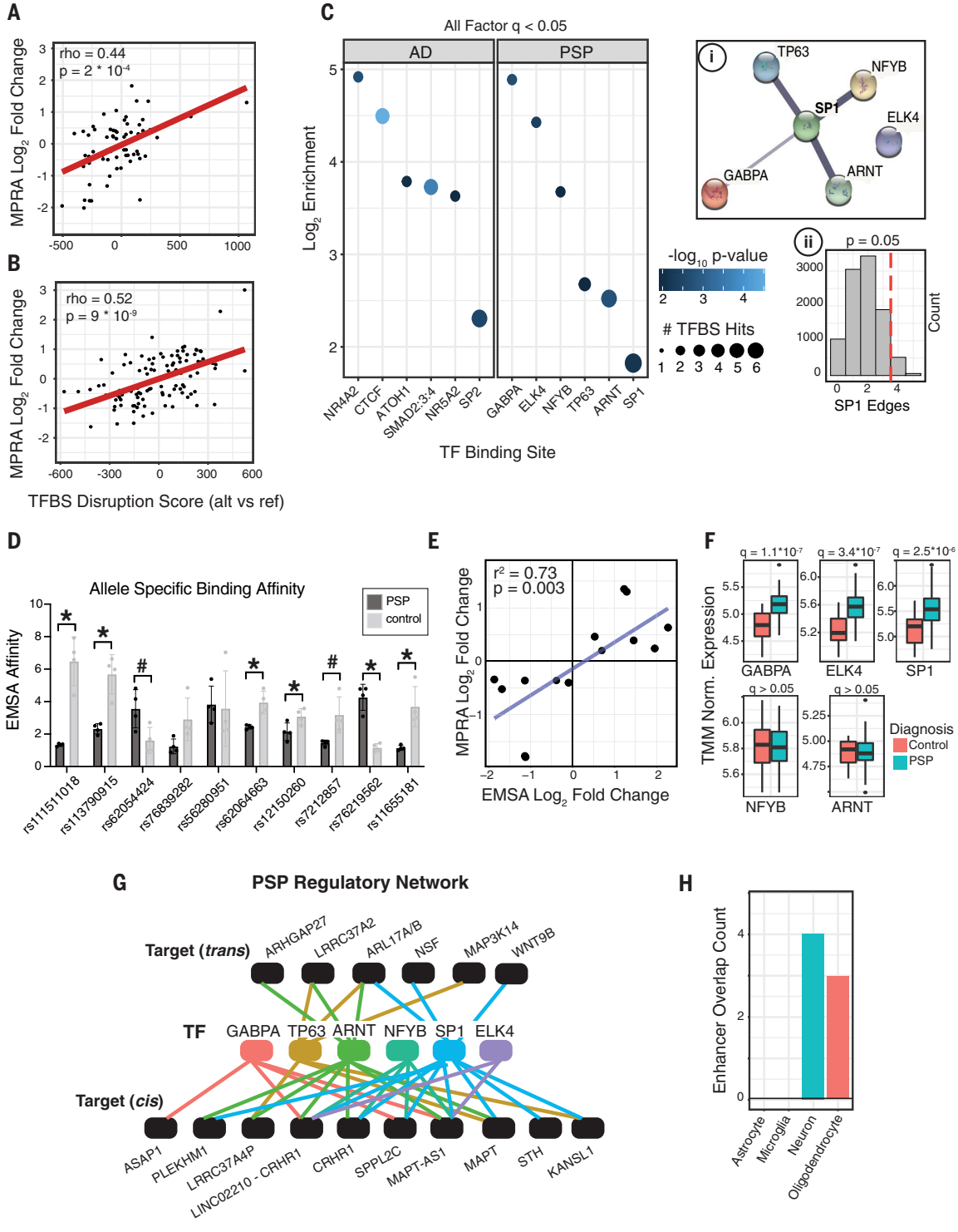


Fig. 5. FrVars disrupt disease-specific TF networks. (**A** and **B**) MPRA variant effect sizes correlate with predicted TFBS disruption scores for our dataset (A) and a published dataset (*18*) (B). Red line, OLS regression, Spearman's rho. (**C**) Shows the $\log_2$ enrichments for significantly disrupted TFBSs (FDR $q < 0.05$, BH method), with frVars from AD and PSP analyzed separately (color indicates $-\log_{10}$ enrichment *P* values, dot size indicates the number of disrupted TFBSs). (Inset) (i) PPI network from the STRING (*55*) database for significantly disrupted PSP TFs. Network hub is the transcription factor SP1. Line thickness indicates the strength of evidence. (ii) Empirical distribution for expected PPI network SP1 connectivity generated by resampling. Red dashed line indicates the observed number of PSP TF network edges. (**D**) Normalized binding affinities to SP1 for both alleles of 10 PSP-associated SNPs predicted to alter SP1 TFBSs tested by EMSA ($n = 4$ replicates). Dark gray bars, PSP risk allele; light gray bars, protective allele. Two-tailed Student's *t* test with FDR correction (BH method). *$q < 0.05$, #$q < 0.1$. Error bars, SEM. (**E**) Correlation between EMSA and MPRA effect sizes for 14 SNPs. Blue line, OLS regression, Pearson's *r*. (**F**) Boxplots of trimmed mean of *M*-values (TMM) normalized and covariate-regressed gene expression from PSP affected (teal; $n = 81$) and unaffected (red; $n = 24$) individuals for five PSP network TFs from (C). Significance assessed using linear models (FDR-corrected *q*-values shown, BH method). (**G**) Network of TFs disrupted in PSP and their target genes (cis: genes within 10 kb of disrupted TFBS; trans: distal and linked by brain Hi-C). (**H**) Bar plot counts overlap between each SNP from (G) with cell type–specific enhancers from (*23*).

enhancers derived from human brain (*23*) to identify the cell types in which target gene expression is most likely dysregulated. Variants mostly overlapped enhancers in neurons (Fig. 5H) and a few in oligodendrocytes, suggesting that this SP1 regulatory network primarily operates in neurons and, to a lesser extent, oligodendrocytes, consistent with the observation of overall neuronal enrichment of PSP risk variants identified in GWAS (*45*).

Discussion

Predicting functionality of noncoding variation is one of the major challenges in modern genetics. Here, we provide a systematic characterization of common variants underlying

disease risk for two neurodegenerative disorders, AD and PSP, identifying regulatory variants at 27 of 34 tested genomic loci. We use functional genomic annotations from human brain tissue to prioritize these variants, experimentally validating the activity of 19 variants and their cognate target genes using direct genome manipulation. Our work illustrates the utility of integrating multiplexed reporter and CRISPR assays to efficiently characterize noncoding disease-associated variation, suggesting a framework for future work in other complex traits. In agreement with previous studies (*14*), we found that four existing computational prediction algorithms for a priori variant prioritization were not highly concordant among themselves, nor were they predictive of functional variants identified by our MPRA or in a previously published dataset (*18*) [fig. S4; see (*20*)]. This emphasizes the necessity for further development and utilization of high-throughput experimental methods to prioritize noncoding regulatory variants from GWAS, especially in regions with extensive LD. In this study, we used MPRA and CRISPR to identify regulatory variants within two such regions, 17q21.31 and 19q13.32, risk factors for PSP and AD, respectively, enabling identification of risk genes within these complex loci, including *PLEKHM1* and *APOC1*.

A potential limitation of this study was the technical challenge of performing MPRA within human brain cell types, with the specific trans-regulatory environment of HEK293T cells likely influencing the generalizability of our screen (*19*). Recognizing that HEK293T cells are not primary neural cells, we directly addressed the question of cell type specificity through integration with orthogonal genomics annotations from the human brain (which reassuringly showed high overlap), followed by CRISPR validation of 42 variants within their native genomic context in neurons and microglia. In doing so, we provide strong evidence for multiple risk genes, implicating *PLEKHM1* and *KANSL1* in PSP and *MS4A4E*, *APOC1*, and *C4A* in AD. Of note, our analysis did not address regulatory function within neurovasculature and epithelial cells, which may also contribute to AD pathogenesis (*46*). Additionally, some associated loci likely operate by means of mechanisms independent of direct transcriptional regulation (i.e., splicing, posttranslational modification), requiring other approaches to delineate causal variants and their mechanisms. However, this work takes an important step forward by providing large-scale functional annotation of these dementia risk loci.

SNP-based heritability is known to enrich in regulatory regions within disease-relevant tissues (*47*). For AD, this enrichment affects networks of genes downstream of several TFs, including *CTCF* and *SMAD*, which have been described as acting in multiple brain and peripheral cell types to affect risk for AD (*38*, *48*). For PSP, our analysis identified an enrichment of TFBS-disrupting frVars in a PPI network with the transcription factor SP1, which regulates a broad array of cellular processes, including chromatin remodeling, apoptosis, immune regulation, and response to oxidative stress in neurons (*49*). Although SP1 network dysregulation has been identified in the AD brain (*50*), it does not appear to harbor AD genetic risk and therefore is likely to play a reactive or secondary role in AD. Overall, our data are consistent with PSP risk primarily affecting neurons and, to a lesser extent, oligodendrocytes (*24*, *45*). This includes an oligodendrocyte enhancer within the tau locus, which we speculate may be related to the distinctive glial pathology in PSP.

These observations suggest a refined model for understanding the combinatorial effects of common genetic risk. Signals from intergenic GWAS loci are typically interpreted as deriving from causal regulatory variants that influence downstream expression of specific cognate risk genes. Our results, implicating a TF network converging on SP1 in PSP, are consistent with a model whereby common genetic variants function in aggregate across multiple TFBSs to disrupt key cell type–specific transcriptional programs. We speculate that this genetic mechanism may manifest itself only upon induction of the relevant transcriptional network, which may occur specifically within a disease context. Our data show that disease-relevant transcriptional networks regulate a large number of cell type–enriched genes, providing a mechanism whereby genetic risk is expressed not by affecting a few core genes (*51*) but by means of polygenic cell type–specific regulatory effects on networks of genes (*52*).

Materials and methods summary

A full description of the materials and methods is provided in the supplementary materials (*20*). An abbreviated version is provided below.

Massively parallel reporter assay

Variant selection and MPRA were performed as described (*20*). To identify functional regulatory variants with significant allelic skew (frVars), normalized barcode counts were combined across replicates and significance determined using a two-way Mann-Whitney *U* test comparing barcode counts between each allele. FrVars were defined at an FDR threshold $q < 0.01$ (Benjamini-Hochberg method). MPRA $\log_2$ effect sizes were defined as the median summed normalized barcode count for the alternate allele minus the reference allele.

Genome engineering

For CRISPR excision experiments, a single line from a healthy male donor (#1205-4) was derived and cultured as previously documented at the University of California, Los Angeles (*53*). IPSC differentiation into mature astrocytes or microglia, as well as further description of the CRISPR excision experiments, is described here (*20*).

For the CROP-seq (pooled CRISPRi) screen, two iPSC lines were maintained at the University of California, San Francisco. For microglial differentiation, a transcription factor–based protocol (iTF-microglia) was used as described by Dräger *et al.* (*33*). Similarly, neurons for the CROP-seq screen were generated using the transcription factor–based i^3N approach as described by Tian *et al.* (*32*). For CROP-seq validation, we picked 42 frVars that were also annotated as lying within transcriptionally active regulatory regions defined as overlapping either brain enhancer and promoter marks from (*23*) or overlapping "active" brain ChromHMM marks from the NIH Roadmap Epigenomics Mapping Consortium (*25*), and analysis performed as described (*20*).

Bioinformatic analysis

Variant annotation

FrVars were annotated for TFBS disruption and overlapped with functional brain annotations to create a high-confidence list of likely regulatory variants (data S2). We calculated TFBS disruption using both the motifbreakR package (*31*) and the SNP2TFBS webtool (*30*). We then identified frVar overlap with enhancer and promoter annotations from sorted microglia, neurons, astrocytes, and oligodendrocytes (*23*) as described (*20*).

TFBS analysis

Predicted TFBS enrichment within "active" MPRA elements was calculated using the HOMER (4.11) software suite (*54*) against a background set of all other oligos (*20*). Predicted TFBS disruption was also scored for all variants using the SNP2TFBS webtool (*30*), and an enrichment OR for frVars was calculated using Fisher's exact test. We also partitioned frVars into those derived from AD or PSP GWAS (*20*) and ran these sets through the SNP2TFBS algorithm, filtering for TFBSs with at least two predicted disruptions. Only TFBSs significantly ($q < 0.05$ or $q < 0.1$) enriched for disruption by frVars were plotted and considered further. For these significantly disrupted TFs, we created PPI networks using the STRING (v11) webtool and standard parameters (*55*), identifying target genes and cell type enrichment as described using functional genomic data (*20*).

Additional analyses

Pairwise Jaccard index calculations, open chromatin enrichment analysis, bulk RNA-seq analysis, LD analysis within the 17q21.31 locus,

and other analyses are further described in the supplemental methods (*20*).

Statistical reporting

Statistical analysis was performed using the stats package in R. All hypothesis testing was two-sided, and all enrichment was determined using a Fisher's exact test, except when explicitly stated. For MPRA allelic skew multiple testing correction, unique variants from MPRA 1 and 2 were combined (5340 total) and Mann-Whitney *U P* values were FDR-adjusted (Benjamini-Hochberg method). FrVars were called at a threshold of $q < 0.01$. For CRISPR excision experiments, a two-tailed Student's *t* test was performed by comparing the average value of controls against the average value of enhancer-targeting conditions. For CRISPRi experiments, SNP:gene *P* values were computed and FDR-adjusted as described (*20*).

REFERENCES AND NOTES

1. M. G. Erkkinen, M. O. Kim, M. D. Geschwind, Clinical neurology and epidemiology of the major neurodegenerative diseases. *Cold Spring Harb. Perspect. Biol.* **10**, a033118 (2018). doi: 10.1101/cshperspect.a033118; pmid: 28716886
2. M. G. Spillantini, M. Goedert, Tau pathology and neurodegeneration. *Lancet Neurol.* **12**, 609–622 (2013). doi: 10.1016/S1474-4422(13)70090-5; pmid: 23684085
3. R. Sims, M. Hill, J. Williams, The multiplex model of the genetics of Alzheimer's disease. *Nat. Neurosci.* **23**, 311–322 (2020). doi: 10.1038/s41593-020-0599-5; pmid: 32112059
4. S. L. Forrest *et al*., Heritability in frontotemporal tauopathies. *Alzheimers Dement. (Amst.)* **11**, 115–124 (2019). doi: 10.1016/j.dadm.2018.12.001; pmid: 30723775
5. J. C. Lambert *et al*., Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat. Genet.* **45**, 1452–1458 (2013). doi: 10.1038/ng.2802; pmid: 24162737
6. B. W. Kunkle *et al*., 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 (2019). doi: 10.1038/s41588-019-0358-2; pmid: 30820047
7. G. U. Höglinger *et al*., Identification of common variants influencing risk of the tauopathy progressive supranuclear palsy. *Nat. Genet.* **43**, 699–705 (2011). doi: 10.1038/ng.859; pmid: 21685912
8. J. A. Chen *et al*., Joint genome-wide association study of progressive supranuclear palsy identifies novel susceptibility loci and genetic correlation to neurodegenerative diseases. *Mol. Neurodegener.* **13**, 41 (2018). doi: 10.1186/s13024-018-0270-8; pmid: 30089514
9. L. D. Ward, M. Kellis, Interpreting noncoding genetic variation in complex traits and human disease. *Nat. Biotechnol.* **30**, 1095–1106 (2012). doi: 10.1038/nbt.2422; pmid: 23138309
10. K. R. Bowles *et al*., 17q21.31 sub-haplotypes underlying H1-associated risk for Parkinson's disease are associated with LRRC37A/2 expression in astrocytes. *Mol. Neurodegener.* **17**, 48 (2022). doi: 10.1186/s13024-022-00551-x; pmid: 35841044
11. X. Zhou *et al*., Non-coding variability at the *APOE* locus contributes to the Alzheimer's risk. *Nat. Commun.* **10**, 3310 (2019). doi: 10.1038/s41467-019-10945-z; pmid: 31346172
12. G. Jun *et al*., Comprehensive search for Alzheimer disease susceptibility loci in the *APOE* region. *Arch. Neurol.* **69**, 1270–1279 (2012). doi: 10.1001/archneurol.2012.2052; pmid: 22869155
13. D. J. Schaid, W. Chen, N. B. Larson, From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nat. Rev. Genet.* **19**, 491–504 (2018). doi: 10.1038/s41576-018-0016-z; pmid: 29844615
14. L. Liu *et al*., Biological relevance of computationally predicted pathogenicity of noncoding variants. *Nat. Commun.* **10**, 330 (2019). doi: 10.1038/s41467-018-08270-y; pmid: 30659175
15. K. Eilbeck, A. Quinlan, M. Yandell, Settling the score: Variant prioritization and Mendelian disease. *Nat. Rev. Genet.* **18**, 599–612 (2017). doi: 10.1038/nrg.2017.52; pmid: 28804138
16. A. Melnikov *et al*., Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. *Nat. Biotechnol.* **30**, 271–277 (2012). doi: 10.1038/nbt.2137; pmid: 22371084
17. R. P. Patwardhan *et al*., Massively parallel functional dissection of mammalian enhancers in vivo. *Nat. Biotechnol.* **30**, 265–270 (2012). doi: 10.1038/nbt.2136; pmid: 22371081
18. R. Tewhey *et al*., Direct identification of hundreds of expression-modulating variants using a multiplexed reporter assay. *Cell* **165**, 1519–1529 (2016). doi: 10.1016/j.cell.2016.04.027; pmid: 27259153
19. J. C. Ulirsch *et al*., Systematic functional dissection of common genetic variation affecting red blood cell traits. *Cell* **165**, 1530–1545 (2016). doi: 10.1016/j.cell.2016.04.048; pmid: 27259154
20. See supplementary materials.
21. A. Melnikov, X. Zhang, P. Rogov, L. Wang, T. S. Mikkelsen, Massively parallel reporter assays in cultured mammalian cells. *J. Vis. Exp.* **2014**, 51719 (2014). doi: 10.3791/51719; pmid: 25177895
22. ENCODE Project Consortium, An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012). doi: 10.1038/nature11247; pmid: 22955616
23. A. Nott *et al*., Brain cell type–specific enhancer–promoter interactome maps and disease-risk association. *Science* **366**, 1134–1139 (2019). doi: 10.1126/science.aay0793; pmid: 31727856
24. V. Swarup *et al*., Identification of conserved proteomic networks in neurodegenerative dementia. *Cell Rep.* **31**, 107807 (2020). doi: 10.1016/j.celrep.2020.107807; pmid: 32579933
25. B. E. Bernstein *et al*., The NIH Roadmap Epigenomics Mapping Consortium. *Nat. Biotechnol.* **28**, 1045–1048 (2010). doi: 10.1038/nbt1010-1045; pmid: 20944595
26. J. Ernst, M. Kellis, ChromHMM: Automating chromatin-state discovery and characterization. *Nat. Methods* **9**, 215–216 (2012). doi: 10.1038/nmeth.1906; pmid: 22373907
27. GTEx Consortium, The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020). doi: 10.1126/science.aaz1776; pmid: 32913098
28. D. Wang *et al*., Comprehensive functional genomic resource and integrative model for the human brain. *Science* **362**, eaat8464 (2018). doi: 10.1126/science.aat8464; pmid: 30545857
29. K. P. Lopes *et al*., Genetic analysis of the human microglial transcriptome across brain regions, aging and disease pathologies. *Nat. Genet.* **54**, 4–17 (2022). doi: 10.1038/s41588-021-00976-y; pmid: 34992268
30. S. Kumar, G. Ambrosini, P. Bucher, SNP2TFBS – a database of regulatory SNPs affecting predicted transcription factor binding site affinity. *Nucleic Acids Res.* **45**, D139–D144 (2017). doi: 10.1093/nar/gkw1064; pmid: 27899579
31. S. G. Coetzee, G. A. Coetzee, D. J. Hazelett, *motifbreakR*: An R/Bioconductor package for predicting variant effects at transcription factor binding sites. *Bioinformatics* **31**, 3847–3849 (2015). doi: 10.1093/bioinformatics/btv470; pmid: 26272984
32. R. Tian *et al*., CRISPR interference-based platform for multimodal genetic screens in human iPSC-derived neurons. *Neuron* **104**, 239–255.e12 (2019). doi: 10.1016/j.neuron.2019.07.014; pmid: 31422865
33. N. M. Dräger *et al*., A CRISPRi/a platform in iPSC-derived microglia uncovers regulators of disease states. *Nat. Neurosci.* 10.1038/s41593-022-01131-4 (2022).
34. P. Datlinger *et al*., Pooled CRISPR screening with single-cell transcriptome readout. *Nat. Methods* **14**, 297–301 (2017). doi: 10.1038/nmeth.4177; pmid: 28099430
35. M. Gasperini *et al*., A genome-wide framework for mapping gene regulation via cellular genetic screens. *Cell* **176**, 1516 (2019). doi: 10.1016/j.cell.2019.02.027; pmid: 30849375
36. S. M. Neuner, J. Tcw, A. M. Goate, Genetic architecture of Alzheimer's disease. *Neurobiol. Dis.* **143**, 104976 (2020). doi: 10.1016/j.nbd.2020.104976; pmid: 32565066
37. M. R. Corces *et al*., Single-cell epigenomic analyses implicate candidate causal variants at inherited risk loci for Alzheimer's and Parkinson's diseases. *Nat. Genet.* **52**, 1158–1168 (2020). doi: 10.1038/s41588-020-00721-x; pmid: 33106633
38. G. Novikova *et al*., Integration of Alzheimer's disease genetics and myeloid genomics identifies disease risk regulatory elements and genes. *Nat. Commun.* **12**, 1610 (2021). doi: 10.1038/s41467-021-21823-y; pmid: 33712570
39. M. Druart, C. Le Magueresse, Emerging roles of complement in psychiatric disorders. *Front. Psychiatry* **10**, 573 (2019). doi: 10.3389/fpsyt.2019.00573; pmid: 31496960
40. I. Ferrer *et al*., Glial and neuronal tau pathology in tauopathies: Characterization of disease-specific phenotypes and tau pathology progression. *J. Neuropathol. Exp. Neurol.* **73**, 81–97 (2014). doi: 10.1097/NEN.0000000000000030; pmid: 24335532
41. M. L. Caillet-Boudin, L. Buée, N. Sergeant, B. Lefebvre, Regulation of human MAPT gene expression. *Mol. Neurodegener.* **10**, 28 (2015). doi: 10.1186/s13024-015-0025-8; pmid: 26170022
42. B. Maloney, D. K. Lahiri, Structural and functional characterization of H2 haplotype MAPT promoter: Unique neurospecific domains and a hypoxia-inducible element would enhance rationally targeted tauopathy research for Alzheimer's disease. *Gene* **501**, 63–78 (2012). doi: 10.1016/j.gene.2012.01.049; pmid: 22310385
43. D. G. McEwan *et al*., PLEKHM1 regulates autophagosome-lysosome fusion through HOPS complex and LC3/GABARAP proteins. *Mol. Cell* **57**, 39–54 (2015). doi: 10.1016/j.molcel.2014.11.006; pmid: 25498145
44. P. Bouwman, S. Philipsen, Regulation of the activity of Sp1-related transcription factors. *Mol. Cell. Endocrinol.* **195**, 27–38 (2002). doi: 10.1016/S0303-7207(02)00221-6; pmid: 12354670
45. V. Swarup *et al*., Identification of evolutionarily conserved gene networks mediating neurodegenerative dementia. *Nat. Med.* **25**, 152–164 (2019). doi: 10.1038/s41591-018-0223-3; pmid: 30510257
46. A. C. Yang *et al*., A human brain vascular atlas reveals diverse mediators of Alzheimer's risk. *Nature* **603**, 885–892 (2022). doi: 10.1038/s41586-021-04369-3; pmid: 35165441
47. H. K. Finucane *et al*., Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015). doi: 10.1038/ng.3404; pmid: 26414678
48. T. Town *et al*., Blocking TGF-β–Smad2/3 innate immune signaling mitigates Alzheimer-like pathology. *Nat. Med.* **14**, 681–687 (2008). doi: 10.1038/nm1781; pmid: 18516051
49. N. Y. Tan, L. M. Khachigian, Sp1 phosphorylation and its regulation of gene transcription. *Mol. Cell. Biol.* **29**, 2483–2488 (2009). doi: 10.1128/MCB.01828-08; pmid: 19273606
50. B. A. Citron, J. S. Dennis, R. S. Zeitlin, V. Echeverria, Transcription factor Sp1 dysregulation in Alzheimer's disease. *J. Neurosci. Res.* **86**, 2499–2504 (2008). doi: 10.1002/jnr.21695; pmid: 18449948
51. E. A. Boyle, Y. I. Li, J. K. Pritchard, An expanded view of complex traits: from polygenic to omnigenic. *Cell* **169**, 1177–1186 (2017). doi: 10.1016/j.cell.2017.05.038; pmid: 28622505
52. N. R. Wray, C. Wijmenga, P. F. Sullivan, J. Yang, P. M. Visscher, Common disease is more complex than implied by the core gene omnigenic model. *Cell* **173**, 1573–1580 (2018). doi: 10.1016/j.cell.2018.05.051; pmid: 29906445
53. A. Gordon *et al*., Long-term maturation of human cortical organoids matches key early postnatal transitions. *Nat. Neurosci.* **24**, 331–342 (2021). doi: 10.1038/s41593-021-00802-y; pmid: 33619405
54. S. Heinz *et al*., Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010). doi: 10.1016/j.molcel.2010.05.004; pmid: 20513432
55. D. Szklarczyk *et al*., STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607–D613 (2019). doi: 10.1093/nar/gky1131; pmid: 30476243
56. Y. Cooper, Tauopathy-MPRA/Tauopathy-MPRA: Tauopathy-MPRA, version 1, Zenodo (2022); https://doi.org/10.5281/zenodo.6549628.

ACKNOWLEDGMENTS

We thank Geschwind lab member J. Grundman for assistance with bulk RNA-seq data analysis, as well as the UCLA TCGB and Flow Cytometry cores for their assistance and technical expertise. We also thank H. Goodarzi for valuable discussions. **Funding:** Y.A.C. was supported by National Institute of Aging fellowship 1F30AG064832 and UCLA-Caltech MSTP training grant T32-GM008042. D.H.G. and G.C. were funded by National Institute of Neurological Disorders and Stroke grant 5UG3NS104095-04. D.H.G. was funded by National Institute of Mental Health grant 5U01MH115746-05. D.H.G. and G.C. were additionally funded by award 20180629 from the Rainwater Charitable Foundation. M.K. and D.H.G. were funded by National Institute of Neurological Disorders and Stroke grant U54 NS123746. M.K. was additionally supported by award A130323 from the Rainwater Charitable Foundation. **Author contributions:** Conceptualization: Y.A.C., G.C., and D.H.G. Data curation: Y.A.C. and N.T. Formal analysis: Y.A.C. and N.T. Funding acquisition: Y.A.C., G.C., M.K., and D.H.G.

Investigation: Y.A.C., N.T., N.M.D., Q.G., S.M.S., Z.Y., A.P., and S.W. Methodology: Y.A.C., J.E.D., S.K., and D.H.G. Project administration: Y.A.C., G.C., and D.H.G. Resources: S.K., G.C., M.K., and D.H.G. Software: Y.A.C. Supervision: S.K., G.C., M.K., and D.H.G. Validation: Y.A.C. Visualization: Y.A.C. and N.T. Writing – original draft: Y.A.C. and D.H.G. Writing – review & editing: Y.A.C., J.E.D., S.K., G.C., M.K., and D.H.G. **Competing interests:** S.K. is an employee of and holds equity in Octant, Inc. M.K. holds equity in and serves on the scientific advisory boards of Engine Biosciences, Casma Therapeutics, Cajal Neuroscience, and Alector and advises Modulo Bio and Recursion Therapeutics. M.K. is an inventor on US patent 11,354,933 related to CRISPRi and CRISPRa screening. G.C. is now a full-time employee of Regeneron Pharmaceuticals and has received stock and stock options from Regeneron Pharmaceuticals. The authors have no additional competing interests to declare. **Data and materials availability:** Raw and processed sequencing data for both MPRA experiments are available from the Gene Expression Omnibus (GEO) under accession number GSE163855. Raw CROP-seq data are available from GEO under accession number GSE207099. To facilitate downstream analyses, preprocessed MPRA count data have also been uploaded to Zenodo (*56*). Processed and normalized MPRA data for both experimental stages are provided in data S1. All MPRA frVars with functional annotations as in Fig. 2, H and I, are available for download in data S2. Processed CROP-seq results for all SNP gene comparisons are provided in data S3. Data underlying Figs. 3, F, H, J, and Q, 4A, and 5, C to E, and figs. S3A, S4, S7, S8, and S13A are provided in data S5. As part of a complete description of the methods, additional accession numbers and materials are provided in data S6. Custom Python scripts used for preprocessing MPRA data have been uploaded to Zenodo (*56*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abi8654

Materials and Methods
Supplementary Text
Figs. S1 to S13
Tables S1 and S2
References (*57*–*74*)
MDAR Reproducibility Checklist
Data S1 to S6

View/request a protocol for this paper from *Bio-protocol*.

Submitted 5 April 2021; accepted 19 July 2022
10.1126/science.abi8654

Science

Functional regulatory variants implicate distinct transcriptional networks in dementia

Yonatan A. CooperNoam TeyssierNina M. DrägerQiuyu GuoJessica E. DavisSydney M. SattlerZhongan YangAbdulsamie PatelSarah WuSriram KosuriGiovanni CoppolaMartin KampmannDaniel H. Geschwind

Science, 377 (6608), eabi8654. • DOI: 10.1126/science.abi8654

Zooming in on dementia genes

Neurodegenerative disorders are a major problem worldwide, and their incidence is continuing to increase. These are usually complex polygenic traits, and although many gene loci have been identified in association with these disorders, the functional roles of these loci are unclear and most of them fall within noncoding regions of the genome. Cooper *et al*. used massively parallel reporter assays to screen noncoding variants reported in genome-wide association studies of two neurodegenerative disorders, followed by functional validation in neurons and microglia. With this combination of methods, the authors identified regulatory variants in several different genes that play functional roles in disease pathogenesis as well as interactions between these genes. —YN

Science (ISSN) is published by the American Association for the Advancement of Science. 1200 New York Avenue NW, Washington, DC 20005. The title *Science* is a registered trademark of AAAS.