MAPT expression is regulated by long-range interactions with *cis*-regulatory elements

Brianne B. Rogers[1,2], Ashlyn G. Anderson[1], Shelby N. Lauzon[1], M. Natalie Davis[1], Rebecca M. Hauser[1], Sydney C. Roberts[1], Ivan Rodriguez-Nunez[1], Katie Trausch-Lowther[1], Erin A. Barinaga[1], Mark Mackiewicz[1], Brian S. Roberts[1], Sara J. Cooper[1], Lindsay F. Rizzardi[1], Richard M. Myers[1,4,*], J. Nicholas Cochran[1,3,4,*]

[1] HudsonAlpha Institute for Biotechnology, Huntsville, AL, USA
[2] University of Alabama at Birmingham, Birmingham, AL, USA
[3] Lead contact
[4] Senior Author

* Co-corresponding authors: ncochran[at]hudsonalpha.org, rmyers[at]hudsonalpha.org

Summary

Tau is encoded by *MAPT* and abnormal aggregates of tau are a hallmark of a group of neurodegenerative diseases called tauopathies. *MAPT* is lowly expressed in neural progenitor cells (NPCs), but it is more highly expressed in oligodendrocytes, astrocytes, and neurons that derive from NPCs. This expression switch at differentiation suggests that *MAPT* expression is controlled by transcription factors and *cis*-regulatory elements specific to these differentiated cell types, including neurons. Using three-dimensional chromatin conformation capture (Capture-C), single-nucleus multiomics (RNA-seq and ATAC-seq), and chromatin marks (H3K27ac, H3K4me1, and chromatin accessibility (DNase hypersensitivity and ATAC-seq)), we nominated 94 unique regions as candidate *cis*-regulatory elements for *MAPT* in human NPCs, NPCs differentiated to neurons, and pure inhibitory or excitatory neuron cultures. We then functionally assessed nominated regions using luciferase assays to test sufficiency for regulatory activity as well as with CRISPR inhibition experiments to assess the effect of nominated regions on *MAPT* expression. We identified both proximal and distal regulatory elements for *MAPT* and confirmed the funciton for several regions, including three regions centromeric to *MAPT* beyond the well-described H1/H2 haplotype inversion breakpoint.

Introduction

The microtubule associated protein tau (encoded by *MAPT*) is most known for its role in neurodegenerative tauopathies like Alzheimer's disease (AD) and tau-positive frontotemporal dementia (FTD) including progressive supranuclear palsy (PSP). Tau is highly abundant in the brain and functions to stabilize microtubules, which are critical for axonal growth and guidance.[1–3] There are six isoforms of tau expressed in the brain and different isoforms have either three (3R) or four (4R) microtubule-binding domains that are regulated by phosphorylation at sites within or adjacent to the binding domain. Under pathological conditions, tau undergoes hyperphosphorylation at these sites, causing it to become unbound from microtubules and oligomerize.[1,4] Hyperphosphorylation of tau is correlated with a mislocalization of tau from its primary location in axons into dendrites.[5,6] Somatodendritic mislocalization of tau occurs early in disease pathogenesis before neurodegeneration is observable and has been shown to mediate synaptic dysfunction leading to neuronal loss.[5–8] These mislocalized tau oligomers form large insoluble fibrils that are the major

components of neurofibrillary tangles (NFTs). These intracellular NFTs occur in the hippocampus during normal aging, but abnormal loads elsewhere in the cortex are pathological hallmarks of tauopathies.[1,9,10] In animal models, reducing endogenous tau has been successful in preventing early mortality, cognitive deficits, and excitotoxicity even in the presence of amyloid-beta.[11–16] While studies [17–20] examining altered expression levels of *MAPT* in AD have produced mixed results, rare copy number events provide an example of the deleterious effect of increased *MAPT* expression. For example, *MAPT* duplication resulting in 1.6 – 1.9 fold higher levels of *MAPT* mRNA leads to a tauopathy with AD-like presentation.[21] Thus, therapies aimed at reducing tau could be beneficial for the full spectrum of tauopathies.

Many studies have been successful in reducing tau levels by targeting either tau modulators or knocking out tau itself.[11–13,15,22] In mice, total tau ablation causes subtle motor deficits later in life; however, these deficits were not observed with partial tau reduction, which still provides therapeutic benefit.[11,23] Targeting tau through regulatory mechanisms controlling its expression may be a tractable method for reducing, instead of ablating, tau expression with limited off-target effects. Large-scale genome-wide association studies (GWASs) have identified 83 common variants that are associated with late-onset Alzheimer's disease (LOAD) and increase disease risk to varying degrees.[24–27] The majority of all GWAS nominated variants are located in non-coding regions of the genome, pointing to the importance of understanding the roles these regions have in contributing to disease and of the potential of these regions as therapeutic targets.[28,29] The leading hypothesis is that these regions are *cis* regulatory elements (CREs), like enhancers, contributing to gene expression.[30,31] Large efforts, like ENCODE, have been made to identify CREs; however, these types of experiments in brain cell types are sparse, resulting in false negatives due to the condition and cell type-specificity of enhancers. Additionally, GWAS SNPs are typically assigned to the nearest gene in the locus, but because of chromatin looping enhancers may actually be in closer three-dimensional proximity with promoters of genes farther away in linear distance.[32–35] A further complication is linkage disequilibrium (LD) which makes causal variant identification difficult, especially in non-coding regions where additional information on conservation and predicted deleteriousness is not available or as informative. Importantly, *MAPT* falls within the largest LD block known in the human genome, spanning over 1 Mb. The locus can be divided into two haplotypes, with a 900 kb inversion of the H2 haplotype with respect to the H1 haplotype encompassing several genes, including *MAPT*. The H1 haplotype is the most common and has been associated with progressive supranuclear palsy (PSP), corticobasal degeneration (CBD), and Parkinson's disease(PD).[36–40] Additionally, the haplotype-tagging single nucleotide polymorphisms (SNPs) have been associated with AD.[24,41] In the H2 haplotype, *MAPT* expression has been shown to be decreased corresponding to a lower risk for AD.[42] Given the complexity of this locus, it is not only important to know the functionality of candidate CREs (cCREs) within this region, but also to identify their target gene (or genes).

Here, we nominate putative *cis*-regulatory elements of *MAPT* using orthogonal genomics approaches. We integrated single nucleus RNA-seq (snRNA-seq) and ATAC-seq (snATAC-seq), from both cultured neurons and previously published DLPFC tissue[20], to correlate chromatin accessibility with *MAPT* expression. We performed chromatin conformation assays (HiC and Capture-C) to determine the 3D interactions with the *MAPT* promoter in NPC and matched differentiated neurons, as well as pure populations of human glutamatergic and GABAergic neurons. Finally, we used the UCSC genome browser to manually inspect regions previously nominated by other studies,[43,44] as well nominate more proximal regions with characteristic histone modifications (H3K27ac and H3K4me1), high conservation, transcription factor (TF) binding motifs, and ENCODE cCRE annotation. To determine the efficacy of these nomination methods, we functionally assessed these nominated regions for sufficiency to induce transcriptional activity through reporter assays and for necessity for the target gene's expression using CRISPR inhibition (CRISPRi) experiments.

Results

Identification of candidate CREs by single nucleus multiomics

To identify regions that may regulate tau expression, we used orthogonal approaches to nominate candidate interactors with the *MAPT* promoter (**Figure 1A**). First, we performed single-nucleus multiomics (snRNA-seq and snATAC-seq) using the 10x Genomics Multiome technology on nuclei isolated from neural progenitor cells (NPCs) and neurons differentiated from these NPCs for 14 and 21 days.[45] This differentiation protocol produces a mixed culture of inhibitory and excitatory neurons as well as immature glial cells. Using this multiomics assay allows for direct mapping of gene expression and chromatin accessibility within the same nuclei without the need to infer cell type identities computationally during cross-modality integration. We removed low quality nuclei and doublets (Methods), and we retained a total of 3,881 nuclei with an average of 1,293 per time point (range of 856 – 1601). We detected a median of 3,559 genes and 55,996 ATAC fragments per cell. We performed normalization and dimensionality reduction for snRNA-seq and snATAC-seq data using Seurat (v4) and Signac (v5), respectively. We used weighted-nearest neighbor (WNN) analysis to determine a joint representation of expression and accessibility and identified 3 distinct clusters (**Figures 1B and S1A-D**). Cluster 1 represents a mostly NPC cluster and is defined by expression of *SOX5*, which regulates embryonic development and cell fate determination.[46] Cluster 2 is a mixture of Day 14 and Day 21 neurons and is defined by expression of *NEAT1*, which in iPSC-derived neurons has been shown to directly regulate neuronal excitability.[47] Cluster 3 is also a mixture of time points of differentiated neurons and is defined by expression of *DCX*, which encodes doublecortin that is expressed in migrating neurons through the central and peripheral nervous system during embryonic and postnatal development, representing an immature neuron population.[48] Cluster 3 also is the only cluster that expresses *MAPT*. This clustering was expected given the resulting mixed culture produced from this differentiation protocol. Differentially expressed genes (DEGs) were identified for each cluster by comparing to the other two remaining clusters. A total of 792 DEGs were identified, including *MAPT* ($\log_2$ Fold Change=0.75; adjusted p =5.39x10^{-24}).

Using the cellranger-arc (v2.0) analysis pipeline, we were able to calculate correlations, "feature linkages," of gene expression and chromatin accessibility to nominate CREs. A feature linkage, or "link," is an ATAC-seq peak ("linked peak") that has a significant correlation between its accessibility across all nuclei and the expression of a "linked gene" (**Figure 1A, top panel**). We restricted this correlation analysis to consider only peaks within 1 megabase (Mb) of each transcription start site (TSS). Previous studies have shown that enhancers are typically within 100 kb of their target genes but this distance can range to beyond a megabase.[49–52] Additionally, it has previously been reported that the genome is partitioned into megabase-sized topological domains (TADs),[53,54] and enhancer-promoter interactions occur more frequently within TAD boundaries.[49,55–57] In choosing to restrict to 1 Mb from the gene's TSS, we are restricting to the most likely regulators of *MAPT* without missing long-range interactions. A description of link calls can be found in our previous publication [20] and in Methods. A total of 54,879 peak-gene links were found, including 9 *MAPT*-specific links (**Figure 1C, Tables 1 and S1**). To evaluate if the linked peaks were associated with regulatory regions, we overlapped the linked peaks with ENCODE curated CREs.[58] The linked peaks were enriched for promoter-like sequences (OR = 6.11; $p < 2.2\text{x}10^{-16}$) and proximal enhancer-like sequences (OR = 1.59; $p < 2.2\text{x}10^{-16}$), with the majority (88.9%) of *MAPT* linked peaks overlapping distal enhancer-like sequences (**Figure 1D**). ATAC peaks were not always linked to just one gene; we identified a median of 1 link per peak globally (range 1 – 31) and 2 links per peak for *MAPT* linked peaks (range 1 – 3; **Figure S1E**). Additionally, the average peak distance from gene TSS was 440 kb globally, but for *MAPT* linked peaks, the peaks tended to be farther from the TSS, with an average distance of 670 kb (**Figure S1F**).

In addition to the single nucleus multiomics data generated from cultured neurons, we also identified *MAPT* links from our previously published data.[20] Using nuclei from DLPFC of 7 AD and

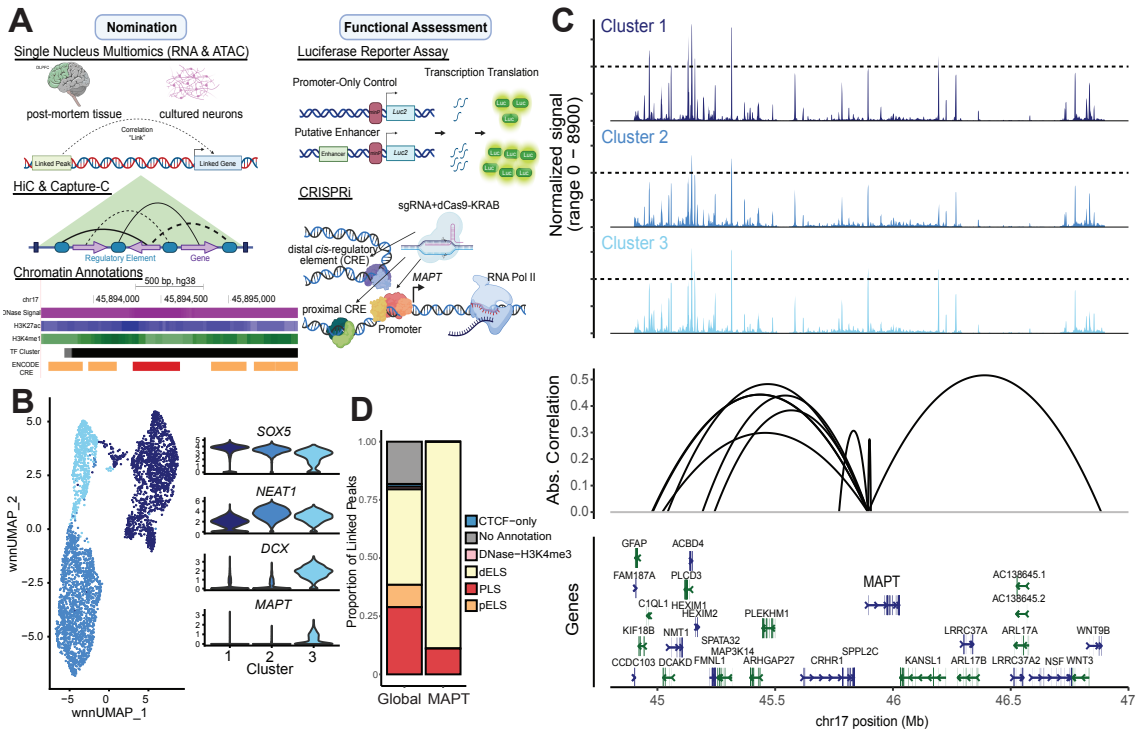

Figure 1. Nomination of candidate CREs by snRNA- I& snATAC-seq in iPSC-derived neuronal cultures.. **A)** Experimental design. **B)** UMAP visualization of the weighted nearest neighbor (WNN) clustering of single nuclei colored by marker genes and cluster assignment. **C)** Linkage plot for all links to *MAPT*. Top panels: coverage plot of normalized accessibility separated by cluster assignment (light blue = NPC; blue = intermediate differentiation; dark blue = immature neurons). Bottom panel: significant peak-gene links. Arc height represents the strength of correlation shown as absolute correlation values **D)** Comparison of proportion of linked peaks by ENCODE regulatory element annotation of global and *MAPT* links.

8 control donors, we identified cell type- and disease-specific CREs and their target genes. Focusing on the *MAPT* locus, we expanded our original restriction of 500 kb from each gene's TSS to 1 Mb to match our current analysis. We identified 58 *MAPT* links; 28 of these links were called in both AD and controls (common), while there were 26 AD-specific and 4 control-specific links (**Figure S2A and TableS2**). We found that many of the linked peaks were negatively correlated with *MAPT* expression. These are likely technical artifacts due to the links being called across all cell types and the combination of high expression of *MAPT* in oligodendrocytes and low variability of *MAPT* expression in neurons (data not shown); however, we included these negatively correlated regions in the validation set to comprehensively evaluate all links. We overlapped the *MAPT* linked peaks with ENCODE CREs and found that they were significantly enriched for both proximal (OR = 4.50; p = $1.58\text{x}10^{-13}$) and distal enhancer-like sequences (OR = 1.98; p = 0.011) and promoter-like sequences (OR = 4.51; p = $1.58\text{x}10^{-13}$; **Figure S2B**).[58] *MAPT* linked peaks had a median of 7.5 (range 1 – 15) links, while globally the median links per ATAC peak was only 3 (range 1 – 31; **Figure S2C**). *MAPT* links also tended to be farther away from the TSS in this data set with an average peak distance of 500 kb compared to 440 kb globally (**Figure S2D**).

Candidate CREs identification by structural analysis and chromatin marks

A prominent theory in the field is that enhancers exert their function by being in physical contact with the target gene's promoter.[33,35,49,59] Identifying contacts with promoters provides a means for determining the regulatory regions contributing to that gene's expression. We sought to identify candidate CREs of *MAPT* using HiC and Capture C assays. Using the Arima HiC kit, we performed HiC in iCell GlutaNeurons from FujiFilm Cellular Dynamics, which produce a > 90 % pure population of human glutamatergic (excitatory) neurons. Using the Juicer Tools pipeline (v1.21) and HICCUPs, we identified 15,918 loops with 2 loops contacting the *MAPT* promoter region (chr17: 45830001 – 45835000 (hg38); **Figures 2A and S3A; Tables 1 and S3**). When overlapping loops with ENCODE annotations, about half are interactions with enhancer-like sequences (enhancer:enhancer 25.2% and enhancer:promoter 23.8%) and 17.7% are promoter:promoter interactions (**Figures 2B and S3B**).[58] For the loops interacting with the *MAPT* promoter, both are regions annotated as enhancer-like sequences (ELS).

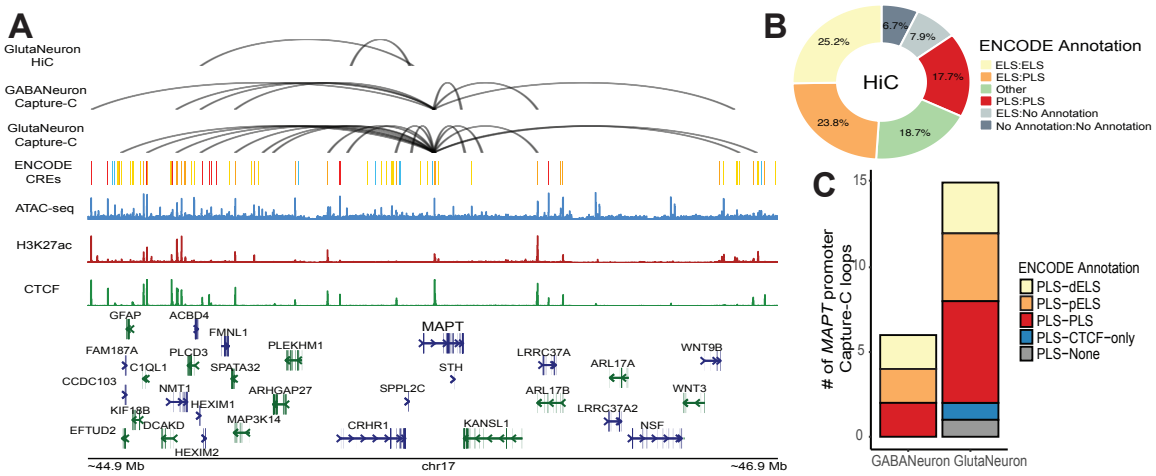

Figure 2. Nomination of candidate CREs using HiC and Capture C in pure excitatory and inhibitory neuron cultures. . **A)** Loop plot of HiCCUPs nominated loops to the *MAPT* promoter. Top panel: Nominated loops by dataset. Top: HiC loops to the *MAPT* promoter from iCell GlutaNeuron cultures. Middle: Capture-C loops to the *MAPT* promoter from iCell GlutaNeuron cultures. Bottom: Capture-C loops to the *MAPT* promoter from iCell GABANeuron cultures. Bottom panel: Top: Track of ENCODE annotated candidate CREs. Middle: ATAC-seq peaks from KOLF2.1J-hNGN2 differentiated neurons. Bottom: H3K27ac and CTCF ChIP-seq peaks from Day 14 BrainPhys differentiated neurons. **B)** Proportion of HiC global loop calls by ENCODE annotation. **C)** Proportion of Capture-C *MAPT* promoter loops by ENCODE annotation for both iCell GlutaNeurons and GABANeurons.

To enrich for regions specifically interacting with the *MAPT* promoter, we used AgilentTechnologies SureSelect DNA probes spanning a 2kb region (chr17: 45892780 – 45893184 and chr17: 45893527 – 45895196; hg38) to perform Capture-C in iCell GlutaNeurons and iCell GABANeurons (> 95% pure population of GABAergic (inhibitory) neurons) cultures from FujiFilm Cellular Dynamics. Loops were called on the Knight-Ruiz normalized combined hic file using HiCCUPs at a resolution of 5kb with the *MAPT* promoter region defined as chr17: 45890001 – 4589500. Window width and peak width were set to 20 and 10, respectively. Loops with a FDR $<$ 0.2 were determined to be significant. Resulting loops were then filtered to a maximum interaction distance of 1Mb, matching our restriction space for single nucleus multiomics nominations. We identified 15 and 6 *MAPT* promoter contact loops in GlutaNeuron and GABANeurons, respectively

(**Figures 2A and S3A;Tables 1, S4, and S5**). About half of these loops are regions annotated as either proximal or distal enhancer-like sequences with the majority of the remaining loops being promoter-like sequences (**Figure 2C**).

In addition to the previous methods described, we also nominated candidate CREs by manual inspection using the UCSC genome browser (http://genome.ucsc.edu) incorporating ENCODE chromatin annotations (DNase Hypersensitivity, H3K27ac, and H3K4me1)[60], transcription factor clusters (ENCODE)[60,61], conservation (Vertebrate Multiz Alignment and Conservation (100 Species)), and previously published data.[43,44] Limiting the search area to the 1 Mb upstream or downstream of the *MAPT* promoter, we nominated an additional 25 regions not overlapping any of the experimental nomination methods ("Other" Method **Table 1, Table S6**). Of these nominated regions, more than half (56%) overlap distal enhancer-like sequences while a large proportion are not annotated by ENCODE; however, ENCODE annotations are defined by non-neuronal cell types (**Figure S3C**).

Number of *MAPT* nominated cis-regulatory elements		
Nomination Method	**Total #**	**# Overlapping**
snMultiomics (DLPFC tissue)	58	9
snMultiomics (cultured neurons)	9	2
GABANeuron Capture-C	6	4
GlutaNeuron Capture-C	15	6
GlutaNeuron HiC	2	1
Other	25	0
Total # of unique tested regions	**94**	

Table 1: Total number of MAPT nominated cis-regulatory elements by nomination method. . For each nomination method, the total number of nominated regions and number of regions that were also nominated by another method are shown. The total number of unique regions tested across methods is shown at the bottom.

Differential analysis of CREs nominated in NPCs and neurons

In our culture system, *MAPT* is very lowly expressed at the NPC stage, but is more highly expressed in the differentiated neuron state (**Figure 3A**). This expression switch makes this model very useful in examining *MAPT* regulation. To interrogate putative CREs controlling this biological switch in *MAPT* expression, we performed differential analysis of Capture-C data generated in NPCs and differentiated neurons. For differential Capture-C, we chose bins of 500 bp, and we identified 49 regions differentially interacting with the *MAPT* promoter between NPC and neurons. Of these interacting regions, 27 were specific to neurons (**Figure 3B**). One of these neuron-specific regions falls within a single nucleus multiomics linkage located 933,730 bps upstream of *MAPT* in the *C1QL1* 3' UTR. Additionally, we wanted to determine if any of the interacting regions were specific to excitatory or inhibitory neurons (**Figure 3B**). We identified 12 regions differentially interacting with the *MAPT* promoter. Of these differential regions, 9 overlapped our nominated *MAPT* regulatory regions with 5 being nominated from single nucleus linkages. For the remaining overlapping regions, 3 were previously identified in our excitatory neuron Capture-C and 1 was identified in inhibitory neuron Capture-C. Lastly, Corces, et al. previously identified a region > 650

kb upstream of the *MAPT* promoter that harbored putative neuron-specific regulatory elements.[44] We also identified a neuron-specific region in our differential Capture-C analysis within this region (**Figure 3C**). To further explore whether this putative regulatory region was specific to neurons, we evaluated chromatin accessibility within this locus using snATAC-seq data generated in our previous publication.[20] We found that this region was specifically accessible in both excitatory and inhibitory neurons, providing further evidence that this region functions as a neuron-specific regulatory element.

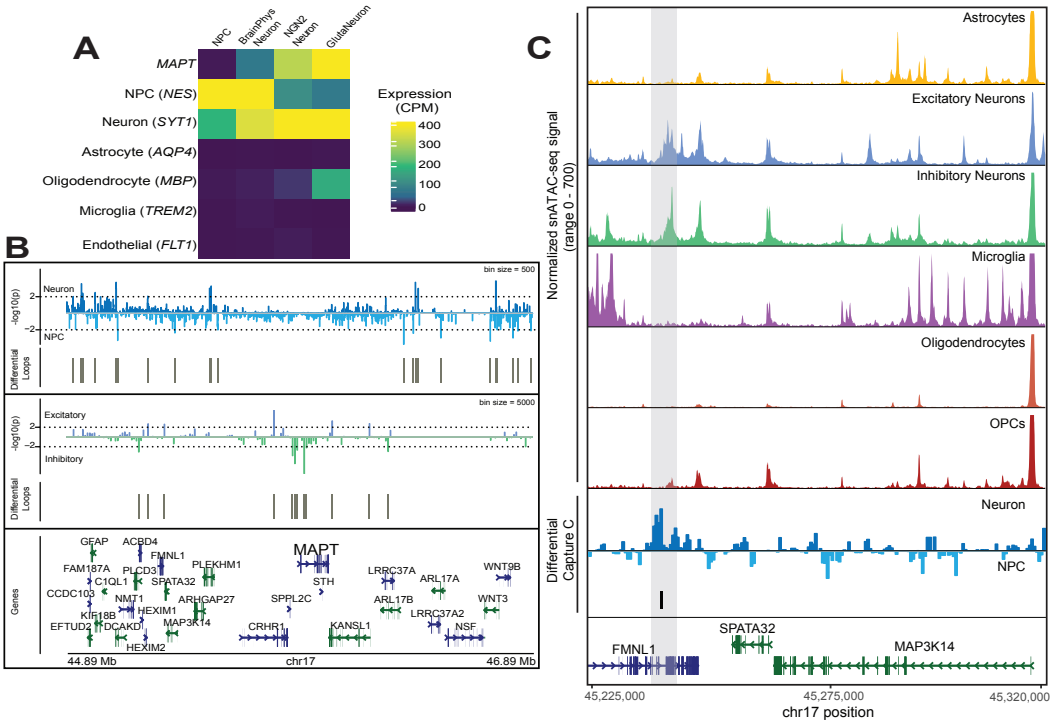

Figure 3. Identification of neuron-specific candidate regulatory regions. A) Heatmap showing expression of *MAPT* and cell type marker genes in NPCs, Day 14 BrainPhys differentiated neurons (BrainPhys Neurons), KOLF2.1J-hNGN2 Neurons (NGN2 Neuron), and iCell GlutaNeurons. **B)** Differential Capture-C directional p-value track. Top panel: Comparison of NPCs and Day 14 BrainPhys Neurons. Middle panel: Comparison of iCell GlutaNeurons (Excitatory) and iCell GABANeurons (Inhibitory). **C)** Zoom in of chr17: 45,225,000-45,320,000. Top panels: Chromatin accessibility across cell types generated by snATAC-seq in DLPFC nuclei. Bottom panels: Differential capture-C comparing BrainPhys differentiated neurons and NPCs with the neuron-specific region highlighted in gray.

Functional assessment of candidate CREs

As previous literature has evaluated, it is important to not only determine regulatory function of a region but also to determine the target gene, or genes.[49,62] We sought to functionally assess these 94 nominated regions using human neurons differentiated from NPCs (BrainPhys), glutamatergic neurons generated from KOLF2.1J-hNGN2 iPSCs, and iCell GlutaNeurons. All of these models highly express the neuronal marker gene *SYT1* and *MAPT* (**Figure 3A**). All cell lines were determined to be H1/H1. We first tested the regions for enhancer-like activity by testing for sufficiency to induce transcriptional activity using a reporter assay. We selected 39 of the nominated

regions for testing in a luciferase reporter assay (**Table S6**). We performed these assays in pure cultures of human iPSC-derived glutamatergic (excitatory) neurons (iCell GlutaNeurons and KOLF2.1J hNGN2).[63] We selected regions with high levels of conservation, high DNase hypersensitivity signal, histone modifications characteristic of active elements (H3K27ac and H3K4me1), and overlapping transcription factor motif clusters60 and ENCODE candidate CREs.[58] We excluded regions annotated as promoter-like sequences because of the prior hypothesis that these regions would be active but not informative for understanding *MAPT* regulation. 21 of the nominated regions were tested in iCell GlutaNeurons, while the remaining 18 were tested in KOLF2.1J-hNGN2 derived neurons (**Figures S4A and S4B**). Eleven of the regions increased activity of the luciferase reporter as measured by fold luminescence units (**Figures 4A and 4B**).

We then sought to answer this question using publicly available datasets. Seventeen of our nominated regions overlap significant elements from either the Cooper, et al.[43] MPRA or the van Arensbergen, et al.[64] SuRE-seq datasets. Additionally, we overlapped our nominations with HiChIP for H3K27ac data generated from multiple brain regions by Corces, et al.[44] Eighteen of our nominated regions overlapped a H3K27ac HiChIP peak ($q < 0.01$), including 5 regions overlapping either MPRA (**Table S7**). Importantly, the region corresponding to the element 77,758 bp away from the *MAPT* promoter that had significant luciferase activity in KOLF2.1J-hNGN2 neurons was also significant in both MPRAs analyzed and also overlapped a H3K27ac HiChIP peak (**Figure 4B**).

We used CRISPRi to determine the target gene of the nominated regulatory regions, as well as their necessity for that gene's expression. CRISPRi uses a dCas9 fused to a transcriptional repressor KRAB, and it has been previously shown to be effective at reducing gene expression in neurons with minimal off-target effects.[65] At least one guide RNA (gRNA) was tested per candidate region, and positive controls were designed to target a region encompassing the *MAPT* promoter and TSS. We also designed gRNAs to both GFP and the AAVS1 safe harbor locus as negative controls where gRNAs should have no effect on gene expression. gRNAs were introduced into NPCs that were subsequently differentiated into neurons and harvested at the two week time point. We isolated RNA and performed both quantitative reverse-transcription PCR(qRT-PCR) and 3' mRNA-sequencing to assess gene expression changes for 62 of the nominated regions (**Figure S4C, Table S8**). The remaining regions either resulted in cell death or did not meet quality control metrics. We observed robust knockdown of *MAPT* expression when targeting the promoter and 300 bp downstream. We observed significant knockdown of 11 other regions, 5 of which were confirmed in both 3' mRNA-sequencing and RT-qPCR (**Figures 4C-E**).

Four of the validated regions are all upstream of the *MAPT* promoter and three are long-range interactions beyond the centromeric inversion breakpoint. The gRNA target region -652,338 resulted in 35% knockdown of *MAPT* expression (**Figure 4C and 5A**) and that was nominated in both DLPFC and cultured neuron single nucleus multiomics datasets. Interestingly, this link was specific to AD donors. This region corresponds to a bidirectional promoter regulating both *FMNL1* and *FMNL1-AS1* expression. However, *FMNL1* is not expressed in our culture model (average cpm < 1), so we could not determine the effect of targeting this region on its expression. Two regions within the *ARHGAP27* locus resulted in more than 50% knockdown of *MAPT* expression (**Figure 4C**). Targeting both of these regions also significantly reduced expression of *EFTUD2* and *KIF18B*, while *MAP3K14* expression was significantly increased (**Figure 4D, Table S8**). These regions were both nominated by single nucleus multiomics, and the linked peaks identified were also correlated with *MAP3K14* expression. The region -464,677 was additionally nominated by Capture-C and specifically looped into the *MAPT* promoter in the iCell GlutaNeurons. This region is just upstream of the *ARHGAP27* promoter, and the gRNA target region encompasses the luciferase region corresponding to -464,861, which showed enhancer-like activity in the KOLF2.1J-hNGN2 neuron dataset (**Figure 4B and 5B**). Therefore, we provide strong evidence that this region is an enhancer regulating *MAPT*.

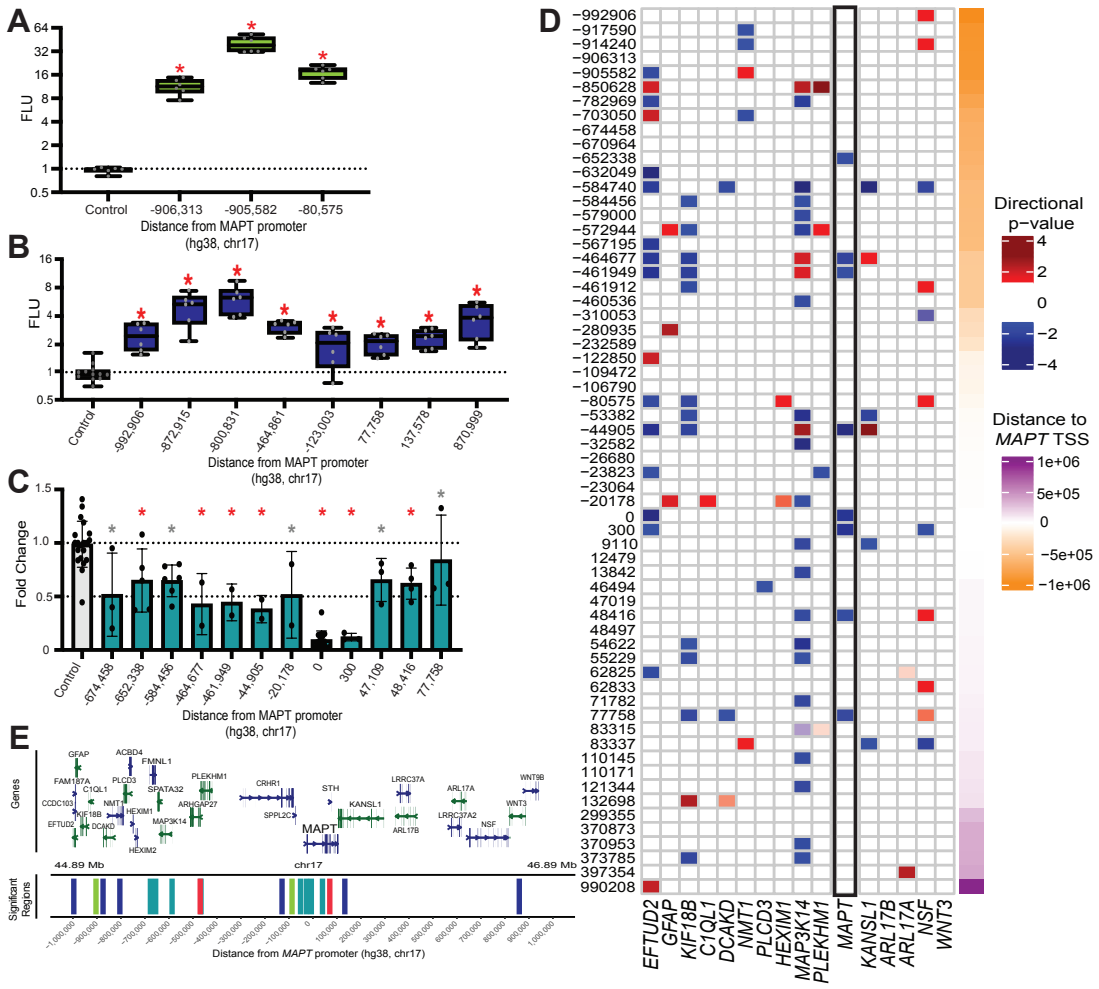

Figure 4. Functional validation of nominated CREs.. **A)** Boxplots showing statistically significant (*p<0.05, ANOVA with Fisher's LSD) elements tested in luciferase assays in iCell GlutaNeurons. **B)** Boxplots showing statistically significant (*p<0.05, ANOVA with Fisher's LSD) elements tested in luciferase assays in KOLF2.1J-hNGN2 neurons. **C)** RT-qPCR of *MAPT* expression. Barplots show statistically significant regions tested in Day 14 BrainPhys differentiated neurons by CRISPRi experiments. Red asterisk indicates regions significant by 3' mRNA-seq and RT-qPCR. Gray asterisk indicates regions significant in either RT-qPCR or 3' mRNA-seq. (*p<0.05, ANOVA with Fisher's LSD). **D)** Heatmap showing differential expression of *MAPT* and all genes expressed (cpm > 30) within 1 Mb of *MAPT* TSS. Left y-axis indicates the distance of the midpoint of the target region tested in CRISPRi experiments from the *MAPT* promoter (chr17:45894000). textbfE) Browser view of significant nominated regions. Green: iCell GlutaNeuron luciferase assay. Blue: KOLF2.1J-hNGN2 neuron luciferase assay. Teal: CRISPRi. Red: significant in luciferase and CRISPRi.

The region -461,949 also resulted in more than 50% knockdown of *MAPT* expression and spans *ARHGAP27* exons 3 and 4 (**Figure 4C**). *ARHGAP27* is very lowly expressed in our culture system (average cpm 2), so we could not evaluate any effect on its expression. The closest validated region (-44,905) lies within the first intron of *MAPT-AS1* and was nominated by single cell multiomics with linkages to 11 genes including *MAPT*, *MAPT-AS1*, and *KANSL1* (**Figure 5C**). Targeting this region resulted in a more than 60% reduction of *MAPT* expression, significant reduction in

EFTUD2 and *KIF18B*, and significant increase in *MAP3K14* and *KANSL1* expression (**Figures 4C and 4D**). The only downstream region validated in both 3' mRNA-seq and RT-qPCR was 48,416 which is in the first intron of *MAPT* (**Figure 5D**). This region was nominated by single nucleus multiomics and is linked to 10 genes, including *MAPT* and *NSF*. Targeting this region resulted in about a 40% reduction of *MAPT* expression, significant reduction in *MAP3K14*, and significant increase in *NSF* expression (**Figure 4C and 4D**). This region also overlaps a H3K27Ac HiChIP peak.[44] Six regions were significant in either RNA-seq or RT-qPCR and cannot be validated with high confidence (**Figure 4C**, gray asterisks, and **Figure 4D**). These regions are interesting but require further validation.

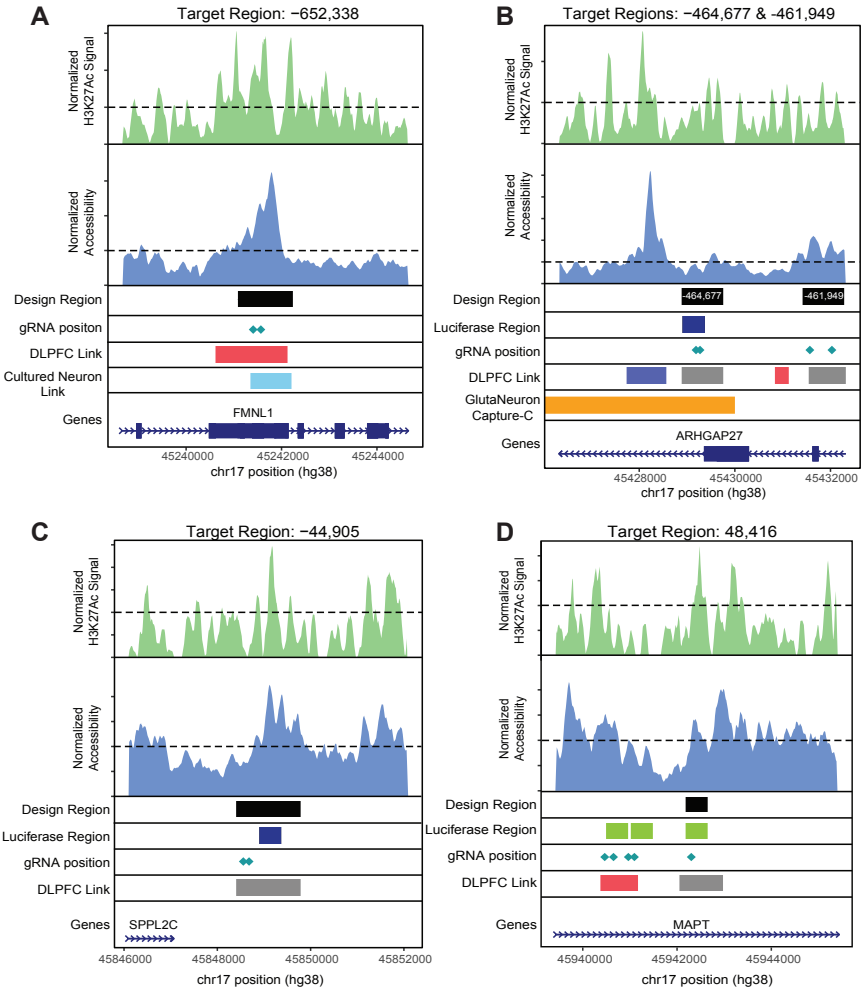

Figure 5. Key CREs regulating *MAPT* expression.. Zoom in on regions validated by qPCR and 3' mRNA-seq (**A)** -652,338, **B)** -464,677 I& -461,949, **C)** -44,905, **D)** 48,416). Top panel: H3K27ac signal from Day14 BrainPhys Neurons. Middle panel: Normalized Accessibility from KOLF2.1J-hNGN2 differentiated Neurons. Bottom panel: Tracks of design region, luciferase region (Green: tested in iCell Gluta Neurons; Blue: tested in KOLF2.1J-hNGN2 neurons), gRNA position, and nominated regions (DLPFC Links: Blue: control-specific link; gray: common link; Red: AD-specific link).

Discussion

Understanding how non-coding regions of the genome contribute to disease has been a major area of focus in modern genetics. Many methods have been developed to better identify candidate regulatory regions by assessing chromatin accessibility, modifications, and structure to determine interactions contributing to gene expression. However, each of these methods have limitations allowing for both false positives and negatives. By combining orthogonal genomics methods, we identified 94 *MAPT* candidate *cis*-regulatory elements and then sought to functionally interrogate these nominations to determine the most likely regulators of *MAPT*. We tested all regions using CRISPRi to determine their necessity for *MAPT* expression, and using 3' RNA-seq, we were able to determine if these regions were necessary for the expression of any other genes within the region. To determine if these regions had enhancer-like activity, we selected a subset of our nominated regions for testing in a luciferase reporter assay.

We identified two regions in the *FMNL1* locus as putative regulators of *MAPT* expression. We provide evidence in three independent chromatin conformation datasets that the region corresponding to -674,458 interacts with the *MAPT* promoter in excitatory and inhibitory neurons. While this region was not confirmed by 3' mRNA-seq, we observed significant reduction of *MAPT* expression by RT-qPCR. Importantly, we were able to validate a region 20 kb downstream (-652,338) in both RT-qPCR and 3' mRNA-seq, and this region was previously reported by Corces, et al. to have elevated interaction with the *MAPT* promoter by H3K27Ac HiChIP specifically in the H1 haplotype.[44] Of genes within 1 Mb of *MAPT*, we only observed differential expression of *MAPT* when targeting this region for repression. Corces, et al. also provided evidence by single cell ATAC-seq that this region could harbor neuron-specific regulatory elements. We also found that this region is 1.5 kb downstream of a neuron-specific Capture-C interaction (chr17: 45,239,001-45,239.500) with the *MAPT* promoter, and we confirmed using snATAC-seq that this region was specifically accessible in excitatory and inhibitory neurons. While we only examined the function of this region in H1 haplotype neurons, we provide further evidence that this region is important for *MAPT* regulation, but specific testing in H2 haplotype cell lines would be required to validate this cell type- and haplotype-specific regulation.

Within the *ARHGAP27* locus, we identified two regions as candidate CREs of *MAPT*, corresponding to regions -464,677 and -461,949. Both of these regions were nominated by single nucleus multiomics, and these ATAC-seq peaks were also linked to *ARHGAP27*, *FMNL1*, and *MAP3K14*. Both of these regions showed significant reduction in *MAPT* expression when targeted with CRISPRi. The region -464,677 also encompasses the region -464,861 that had enhancer-like activity in the luciferase reporter assay, providing two lines of evidence that this region functions as an enhancer and that *MAPT* is the target of this enhancer region. Interestingly, both regions also showed a significant increase in *MAP3K14* expression. We also observe this inverse differential expression change when targeting a region within the first intron of *MAPT-AS1* (-44,905). This region was also nominated by single nucleus multiomics and linked to several other genes, but not to *MAP3K14*. Additionally, in our RNA-seq analysis we observe that 19 regions tested significantly reduce *MAP3K14* expression. *MAP3K14* is located over 500 kb upstream of *MAPT* near *FMNL1*. Several of these regions also show a trend of decreased *MAPT* expression that would suggest that these may be involved in *MAPT* regulation but not required for its expression. These observations suggest that there may be significant interaction between the *MAP3K14/FMNL1* region and *MAPT*, but this relationship requires further evaluation.

In summary, our study provides strong evidence of several CREs contributing to *MAPT* expression. Identification of these CREs is critical for understanding how SNPs in these regions contribute to disease risk. Future studies aimed at identifying TFs bound to these regulatory regions could also point to new therapeutic targets. Because we have identified neuron-specific regulatory

regions, drug screens targeting TFs bound to these regions could provide therapeutic targets with both target gene and tissue specificity. Our study also lays the groundwork for high-throughput approaches for combinatorial knockdown of CREs.[66] We identify several potential CREs with trending effects of decreasing *MAPT* expression. These regions cannot be ruled out as enhancers because redundancy of the *MAPT* enhancer repertoire may result in some CREs contributing to *MAPT* expression that are not required. By targeting different combinations of putative CREs, a more comprehensive understanding of *MAPT* regulation can be obtained.

One limitation of this study is that all experiments were performed in H1 haplotype cell lines. Therefore, regulatory regions identified here may not translate to the H2 haplotype as the inversion of the locus may result in the loss of interaction of these regions with the *MAPT* promoter. [44] We also performed CRISPRi experiments in a mixed culture of inhibitory and excitatory neurons and astrocytes. Given the variability in purity of this culture system, we may miss some neuron-specific regulatory regions. Future experiments examining these putative CREs in high-throughput CRISPR screens, like Perturb-seq,[67] in pure culture of neurons might reveal more neuron-specific regulatory regions. Additionally, due to the low-throughput nature of luciferase assays, we were not able to assess all regions for enhancer-like activity. Therefore, further investigation of these regions through methods like massively parallel reporter assays (MPRAs) are warranted to fully understand their function. While we focused on identifying CREs, specifically enhancers, there are several other non-coding regulatory mechanisms that are not assessed here

Acknowledgements

We thank Jacob M. Loupe for assisting A.G.A. with ChIP-seq analysis. This study was supported by BrightFocus fellowship A2019129F and NIH grants K99AG068271/5R00AG068271 awarded to J.N.C. as well as donors to the HudsonAlpha Foundation Memory and Mobility Program and Leo Fund.

Author contributions

Conceptualization, J.N.C; Formal Analysis, A.G.A. and B.B.R.; Investigation, B.B.R., S.N.L., S.R., B.S.R., M.N.D., I.R., R.M.H., J.N.C., and E.B.; Data Curation, A.G.A. and B.B.R.; Writing– Original Draft, B.B.R., A.G.A., and J.N.C.; Writing– Review and Editing, L.F.R., J.N.C., S.J.C. and R.M.M.; Supervision, S.J.C., R.M.M., and J.N.C.; Funding, R.M.M. and J.N.C.

Declaration of interests

The authors declare no competing interests.

Table Descriptions

Table S1: *MAPT* single nucleus multi-omics links in cultured neurons.

Table S2: *MAPT* single nucleus multi-omics links in postmortem human brain.

Table S3: HiC *MAPT* HICCUPs called loops in GlutaNeurons.

Table S4: GlutaNeuron *MAPT* promoter Capture-C loops.

Table S5: GABANeuron *MAPT* promoter Capture-C loops.

Table S6: Regions selected for testing in luciferase assay.

Table S7: Regions overlapping publicly available functional genomics sets.

Table S8: RNA-seq differentially expressed genes.

Table S9: Cluster DEGs from cultured neuron single nucleus multi-omics before filtering low quality clusters.

Table S10: Significantly enriched GO terms from cultured neuron single nucleus multi-omics before filtering low quality clusters.

Table S11: Designed gRNAs.

Methods

Experimental model and subject details

Cell Lines

XCL4 Neural Progenitor Cells (NPCs) were obtained from StemCell Technologies. BC1 NPCs were obtained from MTI GlobalStem. All NPCs were maintained in NPC Media: 2:1 DMEM (high glucose, L-Glutamine, 100 mg/L Sodium Pyruvate): Ham's F-12 Nutrient Mix, and supplemented with 50X serum-free B-27 Supplement (ThermoFisher Scientific). hFGF (40 μg/mL), hEGF (20 μg/mL), and heparin (5 ug/mL) were added daily to NPC media. iCell GlutaNeurons and iCell GABA Neurons were obtained from FujiFilm and maintained according to the manufacturer protocols for 14 days. 293FT cells were obtained from ThermoFisher Scientific (R70007) and maintained in DMEM (high glucose, L-Glutamine, 100 mg/L Sodium Pyruvate) supplemented with 10% FBS, 1% Glutamax, 1% non-essential amino acids (NEAA), and 500 mg/mL Geneticin (G418 Sulfate) (ThermoFisher Scientific). KOLF2.1J and KOLF2.1J-hNGN2 iPSCs were obtained from the Jackson Laboratory and maintained in mTeSR Plus medium (StemCell Technologies). All cells were cultured at 37°C with 5% CO2. KOLF2.1J NPCs were generated following the StemCell Technologies neural induction protocols.

Method details

Neuron Differentiation

NPCs were differentiated according to the Bardy, et al. protocol.[45] Cells were maintained in the neuronal differentiation medium for 14 or 21 days. iCell GlutaNeurons and GABAneurons were differentiated according to the manufacturer recommendations (FujiFilm Cellular Dynamics). Cells were maintained in their respective BrainPhys complete media for 14 days. KOLF2.1J-hNGN2 iPSCs were dissociated using Accutase and plated as single cells at 50,000 cells/cm^2 in Induction Medium: DMEM/F12 medium with HEPES (ThermoFisher), 100x N2 supplement (StemCell Technologies), 100X Non-essential amino acids (NEAA, ThermoFisher), and 100X Glutamax (ThermoFisher). For plating cells, Induction Medium was supplemented with 5mM Y-27632 and 2 mg/mL Doxycycline. Induction medium supplemented with Doxycycline was renewed daily for two days. On day 3 neurites were present, and cells were renewed with Cortical Neuron Culture Media: BrainPhys neuronal medium (StemCell Technologies), 50X SM1 supplement (StemCell Technologies), 10 μg/mL BDNF (StemCell Technologies), 10 μg/mL NT-3 (StemCell Technologies), and 1 mg/mL Mouse Laminin (Gibco). On day 3, Cortical Neuron Culture Media was supplemented with 2 mg/mL Doxycycline. One-half media changes were performed every other day for a total of 14 days.

Single nucleus multiomics

Nuclei from cultured neurons were isolated as detailed in 10x Genomics demonstrated protocol CG000375 Rev A with modifications. Cells were washed with cold 1X DPBS then 500 uL of NP-40 Lysis Buffer was added and cells were gently scraped off the dish while on ice, transferred to a 1.5 mL tube, and incubated 5 min on ice. Cells were centrifuged at 500 x g for 5 min at 4C and 500 uL 1X PBS + 1% BSA + DAPI + 0.5U/uL Protector RNAse Inhibitor (Roche) was added. Cells were

spun again, gently resuspended 5X in 0.1X Lysis Buffer, and incubated 2 min on ice. After incubation, 1 mL of wash buffer was added and cells were immediately spun at 500 x g for 5 min at 4C. Nuclei were resuspended in <100 uL of Diluted Nuclei Buffer and counted using Countess FL II aiming for 3,000 - 5,000 nuclei/uL. Transposition, barcoding, and library preparation were performed according to the 10x Genomics Chromium Next GEM Single Cell Multiome protocol CG000338 Rev E. Protocols for single nucleus multiomics in DLPFC tissue were described in our previous publication.[20]

Joint snRNA-seq and snATAC-seq workflow

Count matrices were obtained for each time point using cellranger-arc count (v2.0.1) then aggregated using cellranger-arc aggr. Low-quality cells were filtered on gene expression data (nFeatures > 200, nFeatures $< 10{,}000$, and mitochondrial percent < 5) and chromatin accessibility data (nucleosome signal < 2 and TSS enrichment > 2). Peaks that were present in less than 10 cells were removed from the ATAC matrix. RNA counts were normalized with SCTransform with mitochondrial percent per cell regressed out. Principal component analysis (PCA) was performed on RNA, and UMAP was run on the first 30 principal components (PCs). The optimum number of PCs was determined to be 30 PCs using an elbow plot. The ATAC counts were normalized with term-frequency inverse-document-frequency (TFIDF). Dimension reduction was performed with singular value decomposition (SVD) of the normalized ATAC matrix. The ATAC UMAP was created using the 2nd through the 50th LSI components. The weighted nearest neighbor (WNN) graph was determined with Seurat's FindMultiModalNeighbors to represent a weighted combination of both modalities. The first 30 dimensions of the RNA PCA and the 2nd through the 50th dimensions from the ATAC LSI were used to create the graph. The WNN UMAP was created using the wknn (k=20). Clusters were identified from the wknn graph, and markers for each cluster were determined using a Wilcoxon rank-sum test. Clusters with marker genes that were enriched for ribosomal genes were filtered from the data (**Table S9 and S10**). After filtering low quality clusters, all normalization and dimension reduction were repeated as previously described.

Differential Expression from snRNA-seq

Differentially expressed genes were determined for one cluster versus all other clusters using a Wilcoxon Rank Sum test. Only genes that were expressed in at least 10% of cells in a cluster were tested. Genes with a Bonferroni adjusted p-value < 0.01 were determined to be significant. Cluster markers were found for initial clusters to define and filter low quality clusters. Cluster DEGs were recalled with the same method after filtering.

Gene Set Enrichment

The R package enrichR[72–74] was used for gene set enrichment analyses. Sets of positive DEGs for each cluster were used as input to look for enrichment in GO Biological Process 2021, GO Molecular Function 2021, GO Cellular Component 2021, and KEGG 2021 databases. Terms with an adjusted p-value less than 0.05 were considered to be enriched.

Feature Linkage Analysis

Feature linkages were called using cellranger-arc. The maximum interaction distance was restricted to 1Mb, and feature linkages with a correlation score < 0.2 were removed and not used for downstream analysis. For feature linkage calculation, ATAC and GEX counts were normalized independently using depth-adaptive negative binomial normalization. To account for sparsity in the data, the normalized counts were smoothed by taking the weighted sum of the 30 closest neighbors from the KNN graph. The cell weights are determined by using a Guassian kernel transformation of the euclidean distance. Feature linkage scores were calculated by taking the Pearson correlation

between the smoothed counts, while the significance of the correlation was determined using the Hotspot algorithm. For cultured multiomics, the ATAC peaks were called jointly across all time points. Cells from all time points were used to call links. For links called from AD and control tissue, ATAC peaks were called for each cell type and the union of these peaks was used to call links. Cells from all cell types were used in the feature linkage calculation.

Chromatin Conformation Assays

HiC was performed in iCell GlutaNeurons, and Capture-C was performed in NPCs, Day 14 BrainPhys differentiated neurons, iCell GlutaNeurons and GABANeurons. All experiments were performed in triplicate with independent cell grow-ups of approximately 15 million cells for biological replication. For HiC, we followed manufacturer recommendations from the Arima HiC kit- User Guide for Mammalian cell lines (A510008, V.Oct.2019) using the Arima HiC Library Prep using Swift Biosciences Accel-NGS 2S plus DNA library kit (A510008, V. Nov2018). iCell GlutaNeuron and GABANeuron Capture-C was performed following the HiC protocols, with the addition of the Agilent Technology SureSelect XT HS/SureSelect XT Low input target enrichment with pre-capture pooling protocol. 102 SureSelect DNA probes spanning a 2kb region (chr17:45892780-45893184 and chr17:45893527-45895196; hg38) were ordered from AgilentTechnologies. The region between the two probes could not be synthesized because of low sequence complexity. NPC and Day 14 BrainPhys differentiated neuron Capture-C was performed following the NG Capture-C Protocol (v2.4).[68] For this experiment, enrichment for *MAPT* promoter contacts was performed by double capture using two biotinylated oligonucleotides targeting chr17:45892836–45892923 and chr17:45895092–45895179 (both hg38). Libraries were sequenced using an Illumina NovaSeq S4 with XP kit for 2.5 billion reads total, or 833 million reads per replicate for HiC as well as both Capture-C experiments.

HiC and Capture-C Analysis

HiC data was analyzed using the Juicer pipeline (v1.6) with Juicer Tools (v1.22.01). Capture-C data was analyzed using the Juicer Tools pipeline (v1.21). Libraries from each replicate were first run individually. Restriction site positions were generated with generate_site_positions.py with Arima specified as the enzyme. Reads were aligned to hg38 for HiC, but only to chr17 of hg38 for capture-C. For both HiC and capture-C, all 3 replicates were merged to create a combined hic file with mega.sh. Loops were called on the Knight-Ruiz normalized combined hic files using HiCCUPs at a resolution of 5 kb. Window width and peak width were set to 20 and 10 respectively. Loops with a FDR < 0.2 were determined to be significant. Resulting loops were then filtered to a maximum interaction distance of 1Mb.

Differential Capture-C Analysis

Knight-Ruitz normalized counts were pulled from the hic file for each replicate for Gluta and GABA capture-C at a resolution of 5kb and merged by contact positions. Counts that were not connected to the *MAPT* promoter were removed from the count matrix. Differential regions were tested using DESeq2 for 5kb bins. Neuron and NPC capture-C were processed using capC-MAP[69] with the target set at the *MAPT* Promoter (chr17:45892837-45895177) and the restriction enzyme set as DpnII. Contact counts were taken from capC-MAP output at a step size of 500bp and window size of 1kb. Differential regions were tested using DESeq2 with cell line as a covariate. For both analyses, significant regions were defined as those with an adjusted p-value < 0.01. DESeq2 results were formatted as a bigwig for plotting using the directional p-value (-log10(adjusted p-value) * sign($\log_2$FC)).

ATAC-seq Protocol

ATAC-seq was performed in the KOLF2.1J-hNGN2 cell line after 14 days of neuronal differentiation with two biological replicates per cell line. Our protocol is similar to those previously published by Buenrostro, et al.[70,71] Briefly, we harvested 100,000 cells by mechanical dissociation and washed with 50 μL cold 1X PBS and centrifuged at 300xg for 5 minutes. The cell pellet was then resuspended in 50 μL cold Lysis Buffer (10 mM Tris-HCl at pH 7.5, 10 mM NaCl, 3 mM MgCl2, 0.1% NP-40, 0.1% Tween-20, 0.01% Digitonin, and nuclease-free H2O) and incubated on ice for 3 minutes. After incubation, 1 mL of Wash Buffer (10 mM Tris-HCl at pH 7.5, 10 mM NaCl, 3 mM MgCl2, 0.1% Tween-20, and nuclease-free H2O) was added, samples were inverted gently to mix, and centrifuged at 500xg for 10 minutes at 4°C. The supernatant was discarded and the remaining nuclei were resuspended in 50 μL Transposition Mix (2X TD Buffer, TDE1, and nuclease-free H2O) and incubated at 37°C on shaker at 1000 rpm for 30 minutes. Samples were immediately purified using the Qiagen MinElute Reaction Cleanup kit and eluted in a final volume of 11 μL. Recovered DNA was then used to generate sequencing libraries using primers from the Nextera XT Index Kit (15055293) and Q5 Hot Start Master Mix and amplified (30 sec at 98°C; [10 sec at 98°C, 30 sec at 63°C, and 72°C for 1 min] x10 cycles). Libraries were quantified with Qubit dsDNA HS Assay kit and visualized with BioAnalyzer High Sensitivity DNA Analysis kit (Agilent 5067-4626) and 2100 BioAnalyzer Instrument (Agilent). Libraries were sequenced using Illumina NovaSeq flowcell with 50-bp paired-end runs.

ChIP-seq Protocol

ChIP-seq for H3K27Ac and CTCF were performed using chromatin from neurons differentiated from KOF2.1J NPCs (day 14) with biological replicates. Protocols for ChIP-seq are consistent with techniques previously described by our lab and available from the ENCODE Consortium (www.encodeproject.org/documents).[72–74] Briefly, 60 uL Dynabeads (ThermoFisher, cat: 11203D) were washed with cold 1x PBS + 5 mg/mL BSA then combined with 8 uL antibody targeting H3K27Ac (ActiveMotif, Cat: 39133) or CTCF (ActiveMotif, Cat:61311) in a final volume of 200 uL and held at 4°C overnight with rotation. Tubes of aliquoted chromatin are thawed on ice and bead/antibody complex is washed with PBS + 5 mg/mL BSA solution. Beads are ultimately resuspended in 100 uL RIPA and brought to 200 uL with 100 uL chromatin aliquot. Incubate bead/antibody with chromatin using rotation for one hour at room temperature then move to 4°C for another hour. After incubation, bead complexes were washed five times with a LiCl wash buffer (100 mM Tris at pH 7.5, 500 mM LiCl, 1% NP-40, 1% sodium deoxycholate) and wash with 1 mL of cold TE (10 mM Tris-HCl at pH 7.5, 0.1 mM Na2EDTA). Chromatin was eluted from beads by incubating with intermittent shaking for 1 hour at 65°C in IP elution buffer (1% SDS, 0.1 M NaHCO3), followed by incubating overnight at 65°C to reverse formaldehyde cross-links. DNA was purified using DNeasy Blood and Tissue kit (Qiagen 69506) and eluted in a final volume of 50 uL EB. Recovered DNA was quantified using Qubit dsDNA HS Assay kit (Thermo Q32854). For input controls, one aliquot of each tissue was brought to 200 uL with RIPA and reverse-crosslinked overnight at 65°C. The following morning, samples were incubated an additional 30 minutes with 20 uL Proteinase K and 4 uL RNase A (Qiagen 19101) and subsequently eluted for DNA using DNeasy Blood and Tissue kit. The entirety of the remaining IP DNA (and approximately 250 ng Input control) were used to generate sequencing libraries. Libraries were prepared by blunting and ligating ChIP DNA fragments to sequencing adapters for amplification with barcoded primers (30 sec at 98°C; [10 sec at 98°C, 30 sec at 65°C, 30 sec at 72°C] x 15 cycles; 5 min at 72°C). Libraries were quantified with Qubit dsDNA HS Assay kit and visualized with Standard Sensitivity NGS Fragment Analysis Kit (Advanced Analytical DNF-473) and Fragment Analyzer 5200 (Agilent). Libraries were sequenced using Illumina NovaSeq flowcell with 100bp single-end runs.

ChIP-seq analysis

Prior to analysis, reads were processed to remove optical duplicates with clumpify (BBMap v38.20; https://sourceforge.net/projects/bbmap/) [dedupe=t optical=t dupedist=2500] and remove adapter reads with Cutadapt (v1.16)[-a AGATCGGAAGAGC -m 40][75]. Input reads were capped at 40 million using Seqtk (v1.2; https://github.com/lh3/seqtk). Individual experiments were constructed following ENCODE guidelines (https://www.encodeproject.org/about/experiment-guidelines/) and analyzed with the chip-seq-pipeline2 processing pipeline (https://github.com/ENCODE-DCC/chip-seq-pipeline2). FASTQs were cropped to 50bp and all other parameters were run using the default setting. Final peaks were called using pseudoreps and the IDR naïve overlapping method with a threshold of 0.05.

Plasmids

The pNL1.1.CMV [Nluc/CMV] and pGL4.23 [luc2/minP] vectors were obtained from Promega. Luciferase elements were generated by selecting 467 bp of the nominated region and both the forward and reverse complement sequences were ordered as gBlocks from Integrated DNA Technologies (IDT). Elements were cloned into the pGL4.23 [luc2/minP] vector digested with EcoRV by Gibson Assembly. Element insertion was confirmed by Sanger sequencing (MCLAB). Each element was individually prepped 3 times for a total of 6 individual plasmid preparations per nominated region. The pMD2.G, psPAX2, FUGW-H1-GFP-neomycin, and pLV hU6-sgRNA hUbC-dCas9-KRAB-T2A-Puro plasmids were obtained from Addgene (#12259, #12260, #37632, and #71236 respectively). sgRNA sequences were designed using https://benchling.com and ordered as premixed primer pools from IDT (**Table S11**). The sgRNA sequences were then inserted into the pLV hU6-sgRNA hUbC-dCas9-KRAB-T2A-Puro plasmid by digesting with Esp3I and subsequent ligation.[76,77]

Lentivirus Production

293FT cells were plated at 70,000 cells/cm^2 in poly-L-ornithine coated 6-well culture plates. The next day, the media was renewed with OptiMEM Reduced Serum Media supplemented with 300 mg D-glucose. Cells were transfected with 1 μg pLV hU6-sgRNA hUbC-dCas9-KRAB-T2A-Puro plasmid with inserted sgRNA sequence using Lipofectamine LTX with Plus Reagent, following manufacturer recommendations. 48 hours post transfection, supernatant was harvested and filtered through a 0.45 μm syringe filter into a 15 mL conical tube on ice.

CRISPRi Experiments

NPCs were plated at 57,000 cells/cm^2 in reduced growth factor Matrigel (Corning #354230) coated 12 well culture plates. The next day, culture media was renewed with NPC media supplemented with protamine sulfate (10 mg/mL). To transduce NPCs, 500 μL of cold, filtered lentivirus was added to the NPC media and plates were centrifuged at 2000 rpm for 1 hour. 24 hours post transduction, cells were renewed with fresh NPC media with growth factors and 0.5 μg/mL puromycin for selection of successfully transduced cells. Cells were selected for 24-48 hours, until the control well (transduced with FUGW-H1-GFP-neomycin) had minimal remaining living cells. NPCs were then swapped to neuronal differentiation media following the Bardy, et al. protocol[45]. Neurons were harvested for RNA isolation at day 14 of differentiation. sgRNAs were designed upstream of the *MAPT* TSS in the promoter region (chr17:45884000; hg38) as a positive control. sgRNAs targeting the AAVS1 safe harbor locus were used as non-targeting controls.

3'mRNA-seq Library Generation

RNA isolation was performed using the Norgen Total RNA isolation and RNase-Free DNase I kits (Norgen 17250 and 25720, respectively) and quantified using the Qubit RNA HS Assay Kit (Thermo

Q32855). Libraries were prepared using the QuantSeq 3 mRNA-Seq Library Prep Kit FWD for Illumina and UMI Second Strand Synthesis Module for QuantSeq FWD (Illumina, Read 1) from Lexogen (015.96 and 081.96, respectively). Libraries were quantified using the Qubit DNA HS Assay Kit (Thermo Q32854) and visualized with the BioAnalyzer High Sensitivity DNA Analysis kit (Agilent 5067-4626) and 2100 BioAnalyzer Instrument (Agilent). Libraries were sequenced by HudsonAlpha Discovery using Illumina NovaSeq S1 100 cycle or S4 200 cycle flowcells.

RT-qPCR

MAPT expression was also determined by reverse transcription-quantitative polymerase chain reaction (RT-qPCR). cDNA was generated from 100 ng RNA using SuperScript IV Master Mix (Thermo 11756050). Two Taqman probes were used to measure *MAPT* relative abundance (Thermo Hs00213491_m1 and Hs00902194_m1). Taqman probes to AP1G1 (Thermo Hs00964419_m1) and GAPDH (Thermo Hs99999905_m1) were used as housekeeping controls. For each sample, Δ ΔCt values were calculated using the average of the median housekeeper Ct values. Samples with average housekeeper Ct values $\pm$ 1 standard deviation from the average of the other samples in each run were removed.

RNAseq Analysis

UMIs were first extracted from the reads with UMI-tools extract with the extraction method set as regex. Reads were trimmed with bbduk then aligned to hg38 with STAR using the Lexogen recommended parameters for QuantSeq. Bams were deduplicated by UMI and mapping coordinate using UMI-tools dedup. Counts were generated with htseq-count using the intersection-nonempty method. The count matrix was normalized to counts per million (CPM). After normalization, each sample was scored based on the level of differentiation and astrocyte presence using Seurat's AddModuleScore. Module scores were calculated using a list of markers for differentiation (*CACNA1C, ENO2, MAP2*) and astrocytes (*AQP4, SLC1A3*) (Figure S5). Differentially expressed genes (DEGs) were determined for each target region versus non-targeting controls for all genes in the *MAPT* locus (+1Mb). Differential expression was assessed with a linear model with differentiation score, astrocyte score, and batch as covariates. Genes with a p-value$<$ 0.05 were determined to be significant.

Nucleofection

iCell GlutaNeurons and KOLF2.1J-hNGN2 differentiated neurons were plated at 200,000 cells/cm^2 in 24-well matrigel pre-coated plates. Cells were differentiated according to their respective protocols for 14 days. One hour before nucleofection, media was renewed to fresh media plus 500X Rock Inhibitor (Y27632). Neurons were nucleofected with 10 μ total DNA plasmid DNA using the AD1 4D-Nucleofector Y kit (Lonza V4YP-1A24) with the ED158 (iCell GlutaNeurons) and EH166 (KOLF2.1J-hNGN2 neurons) pulse settings. Per nucleofection, 9 μg element cloned into pGl4.23 and 1 μg pNL1.1.CMV [*Nluc*/CMV] were used. A transfection reaction of 9 μg pGL4.23 [*luc2*/minP] and 1 μg pNL1.1.CMV [*Nluc*/CMV] was used as a baseline control. Both vectors were also transfected as background controls (1 μg) with pmaxGFP (9 μg, Lonza). 24 hours post nucleofection, cells were renewed with fresh media without Rock inhibitor. Cell lysates were harvested by freezing at -80 °C 48 hours post-transfection.

Luciferase Assays

Luciferase assays were performed using the Nano-Glo Dual-Luciferase Reporter Assay System (Promega Cat#N1630) following the manufacturer's protocol. Cell lysis was performed on the 24 well plate and aliquoted across 4 wells of a white 96-well plate for 4 technical replicates per biological replicate. Assays were completed in quadruplicate. Firefly luminescence was first normalized across

the average plate luminescence and then normalized to the average control luminescence. For each biological replicate, the median fold luminescence value was determined for the four technical replicates. Four biological replicates were compared to the pGL4.23 [luc2/minP]/ pNL1.1.CMV [Nluc/CMV] control using an ordinary one-way ANOVA with Fisher's LSD.

Data and code availability

All supplementary tables are available upon request. All the code generated during this study is available at github.com/aanderson54/MAPT_cre

References

1. Chong, F.P., Ng, K.Y., Koh, R.Y., and Chye, S.M. (2018). Tau Proteins and Tauopathies in Alzheimer's Disease. Cell. Mol. Neurobiol. 38, 965–980. 10.1007/s10571-017-0574-1.

2. Coupland, K.G., Kim, W.S., Halliday, G.M., Hallupp, M., Dobson-Stone, C., and Kwok, J.B.J. (2016). Role of the long non-coding RNA MAPT-AS1 in regulation of microtubule associated protein tau (MAPT) expression in Parkinson's disease. PLoS ONE 11. 10.1371/JOURNAL.PONE.0157924.

3. Hefti, M.M., Farrell, K., Kim, S., Bowles, K.R., Fowkes, M.E., Raj, T., and Crary, J.F. (2018). High-resolution temporal and regional mapping of MAPT expression and splicing in human brain development. PLOS ONE 13, e0195771. 10.1371/journal.pone.0195771.

4. Cherry, J.D., Esnault, C.D., Baucom, Z.H., Tripodis, Y., Huber, B.R., Alvarez, V.E., Stein, T.D., Dickson, D.W., and McKee, A.C. (2021). Tau isoforms are differentially expressed across the hippocampus in chronic traumatic encephalopathy and Alzheimer's disease. Acta Neuropathol. Commun. 9, 86. 10.1186/s40478-021-01189-4.

5. Hoover, B.R., Reed, M.N., Su, J., Penrod, R.D., Kotilinek, L.A., Grant, M.K., Pitstick, R., Carlson, G.A., Lanier, L.M., Yuan, L.-L., et al. (2010). Tau mislocalization to dendritic spines mediates synaptic dysfunction independently of neurodegeneration. Neuron 68, 1067–1081. 10.1016/j.neuron.2010.11.030.

6. Cochran, J.N., Hall, A.M., and Roberson, E.D. (2014). The dendritic hypothesis for Alzheimer's disease pathophysiology | Elsevier Enhanced Reader. Brain Res. Bull. 103, 18–28. 10.1016/j.brainresbull.2013.12.004.

7. Miller, E.C., Teravskis, P.J., Dummer, B.W., Zhao, X., Huganir, R.L., and Liao, D. (2014). Tau phosphorylation and tau mislocalization mediate soluble A oligomer-induced AMPA glutamate receptor signaling deficits. Eur. J. Neurosci. 39, 1214–1224. 10.1111/ejn.12507.

8. Ittner, L.M., Ke, Y.D., Delerue, F., Bi, M., Gladbach, A., van Eersel, J., Wölfing, H., Chieng, B.C., Christie, M.J., Napier, I.A., et al. (2010). Dendritic Function of Tau Mediates Amyloid-Toxicity in Alzheimer's Disease Mouse Models. Cell 142, 387–397. 10.1016/j.cell.2010.06.036.

9. Braak, H., Alafuzoff, I., Arzberger, T., Kretzschmar, H., and Del Tredici, K. (2006). Staging of Alzheimer disease-associated neurofibrillary pathology using paraffin sections and immunocytochemistry. Acta Neuropathol. (Berl.) 112, 389–404. 10.1007/s00401-006-0127-z.

10. Braak, H., Thal, D.R., Ghebremedhin, E., and Tredici, K.D. (2011). Stages of the Pathologic Process in Alzheimer Disease: Age Categories From 1 to 100 Years.

11. Roberson, E.D., Scearce-Levie, K., Palop, J.J., Yan, Y., Cheng, I.H., Wu, T., Gerstein, H., Yu, G.-Q., and Mucke, L. (2007). Reducing endogenous tau ameliorates amyloid beta-induced deficits in an Alzheimer's disease mouse model. Science 316, 750–754. 10.1126/SCIENCE.1141736.

12. Roberson, E.D., Halabisky, B., Yoo, J.W., Yao, J., Chin, J., Yan, F., Wu, T., Hamto, P., Devidze, N., Yu, G.-Q., et al. (2011). Amyloid-/Fyn-induced synaptic, network, and cognitive

impairments depend on tau levels in multiple mouse models of Alzheimer's disease. J. Neurosci. Off. J. Soc. Neurosci. 31, 700–711. 10.1523/JNEUROSCI.4152-10.2011.

13. DeVos, S.L., Miller, R.L., Schoch, K.M., Holmes, B.B., Kebodeaux, C.S., Wegener, A.J., Chen, G., Shen, T., Tran, H., Nichols, B., et al. (2017). Tau reduction prevents neuronal loss and reverses pathological tau deposition and seeding in mice with tauopathy. Sci. Transl. Med. 9, eaag0481. 10.1126/scitranslmed.aag0481.

14. Wegmann, S., Maury, E.A., Kirk, M.J., Saqran, L., Roe, A., Devos, S.L., Nicholls, S., Fan, Z., Takeda, S., Cagsal-Getkin, O., et al. (2015). Removing endogenous tau does not prevent tau propagation yet reduces its neurotoxicity. EMBO J. 34, 3028–3041. 10.15252/embj.201592748.

15. DeVos, S.L., Corjuc, B.T., Commins, C., Dujardin, S., Bannon, R.N., Corjuc, D., Moore, B.D., Bennett, R.E., Jorfi, M., Gonzales, J.A., et al. (2018). Tau reduction in the presence of amyloid-/*Beta* prevents tau pathology and neuronal death in vivo. Brain J. Neurol. 141, 2194–2212. 10.1093/BRAIN/AWY117.

16. Chang, C.-W., Evans, M.D., Yu, X., Yu, G.-Q., and Mucke, L. (2021). Tau reduction affects excitatory and inhibitory neurons differently, reduces excitation/inhibition ratios, and counteracts network hypersynchrony. Cell Rep. 37, 109855. 10.1016/j.celrep.2021.109855.

17. Huin, V., Deramecourt, V., Caparros-Lefebvre, D., Maurage, C.-A., Duyckaerts, C., Kovari, E., Pasquier, F., Buée-Scherrer, V., Labreuche, J., Behal, H., et al. (2016). The MAPT gene is differentially methylated in the progressive supranuclear palsy brain. Mov. Disord. 31, 1883. 10.1002/mds.26820.

18. Fukasawa, J.T., Labio, R.W. de, Rasmussen, L.T., Oliveira, L.C. de, Chen, E., Villares, J., Tureck, G., Smith, M. de A.C., and Payao, S.L.M. CDK5 and MAPT Gene Expression in Alzheimer's Disease Brain Samples. Curr. Alzheimer Res. 15, 182–186.

19. Jiang, J., Wang, C., Qi, R., Fu, H., and Ma, Q. (2020). scREAD: A Single-Cell RNA-Seq Database for Alzheimer's Disease. iScience 23. 10.1016/j.isci.2020.101769.

20. Anderson, A.G., Rogers, B.B., Loupe, J.M., Rodriguez-Nunez, I., Roberts, S.C., White, L.M., Brazell, J.N., Bunney, W.E., Bunney, B.G., Watson, S.J., et al. (2023). Single nucleus multiomics identifies ZEB1 and MAFB as candidate regulators of Alzheimer's disease-specific cis-regulatory elements. Cell Genomics 0. 10.1016/j.xgen.2023.100263.

21. Le Guennec, K., Quenez, O., Nicolas, G., Wallon, D., Rousseau, S., Richard, A.-C., Alexander, J., Paschou, P., Charbonnier, C., Bellenguez, C., et al. (2017). 17q21.31 duplication causes prominent tau-related dementia with increased MAPT expression. Mol. Psychiatry 22, 1119–1125. 10.1038/mp.2016.226.

22. Kim, J., de Haro, M., Al-Ramahi, I., Garaicoechea, L.L., Jeong, H.-H., Sonn, J.Y., Tadros, B., Liu, Z., Botas, J., and Zoghbi, H.Y. (2023). Evolutionarily conserved regulators of tau identify targets for new therapies. Neuron, S0896627322010856. 10.1016/j.neuron.2022.12.012.

23. Morris, M., Hamto, P., Adame, A., Devidze, N., Masliah, E., and Mucke, L. (2013). Age-appropriate cognition and subtle dopamine-independent motor deficits in aged Tau knockout mice. Neurobiol. Aging 34, 1523–1529. 10.1016/j.neurobiolaging.2012.12.003.

24. Bellenguez, C., Küçükali, F., Jansen, I.E., Kleineidam, L., Moreno-Grau, S., Amin, N., Naj, A.C., Campos-Martin, R., Grenier-Boley, B., Andrade, V., et al. (2022). New insights into the genetic etiology of Alzheimer's disease and related dementias. Nat. Genet. 54, 412–436. 10.1038/s41588-022-01024-z.

25. Wightman, D.P., Jansen, I.E., Savage, J.E., Shadrin, A.A., Bahrami, S., Holland, D., Rongve, A., Børte, S., Winsvold, B.S., Drange, O.K., et al. (2021). A genome-wide association study with 1,126,563 individuals identifies new risk loci for Alzheimer's disease. Nat. Genet. 2021 539 53, 1276–1282. 10.1038/s41588-021-00921-z.

26. Kunkle, B.W., Grenier-Boley, B., Sims, R., Bis, J.C., Damotte, V., Naj, A.C., Boland, A., Vronskaya, M., van der Lee, S.J., Amlie-Wolf, A., et al. (2019). Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates A/*Beta*, tau, immunity and lipid processing. Nat. Genet. 51, 414–430. 10.1038/s41588-019-0358-2.

27. Novikova, G., Kapoor, M., Tcw, J., Abud, E.M., Efthymiou, A.G., Chen, S.X., Cheng, H., Fullard, J.F., Bendl, J., Liu, Y., et al. (2021). Integration of Alzheimer's disease genetics and myeloid genomics identifies disease risk regulatory elements and genes. Nat. Commun. 12, 1610. 10.1038/s41467-021-21823-y.

28. Watanabe, K., Stringer, S., Frei, O., Umićević Mirkov, M., de Leeuw, C., Polderman, T.J.C., van der Sluis, S., Andreassen, O.A., Neale, B.M., and Posthuma, D. (2019). A global overview of pleiotropy and genetic architecture in complex traits. Nat. Genet. 51, 1339–1348. 10.1038/s41588-019-0481-0.

29. Welter, D., MacArthur, J., Morales, J., Burdett, T., Hall, P., Junkins, H., Klemm, A., Flicek, P., Manolio, T., Hindorff, L., et al. (2014). The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Res. 42, D1001–D1006. 10.1093/nar/gkt1229.

30. Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds, A.P., Sandstrom, R., Qu, H., Brody, J., et al. (2012). Systematic Localization of Common Disease-Associated Variation in Regulatory DNA. Science 337, 1190–1195. 10.1126/science.1222794.

31. Trynka, G., Sandor, C., Han, B., Xu, H., Stranger, B.E., Liu, X.S., and Raychaudhuri, S. (2013). Chromatin marks identify critical cell types for fine mapping complex trait variants. Nat. Genet. 45, 124–130. 10.1038/ng.2504.

32. Reed, K.S.M., Davis, E.S., Bond, M.L., Cabrera, A., Thulson, E., Quiroga, I.Y., Cassel, S., Woolery, K.T., Hilton, I., Won, H., et al. (2022). Temporal analysis suggests a reciprocal relationship between 3D chromatin structure and transcription. Cell Rep. 41. 10.1016/j.celrep.2022.111567.

33. Carter, D., Chakalova, L., Osborne, C.S., Dai, Y., and Fraser, P. (2002). Long-range chromatin regulatory interactions in vivo. Nat. Genet. 32, 623–626. 10.1038/ng1051.

34. Chakraborty, S., Kopitchinski, N., Zuo, Z., Eraso, A., Awasthi, P., Chari, R., Mitra, A., Tobias, I.C., Moorthy, S.D., Dale, R.K., et al. (2023). Enhancer–promoter interactions can bypass CTCF-mediated boundaries and contribute to phenotypic robustness. Nat. Genet., 1–11. 10.1038/s41588-022-01295-6.

35. Whalen, S., Truty, R.M., and Pollard, K.S. (2016). Enhancer–promoter interactions are encoded by complex genomic signatures on looping chromatin. Nat. Genet. 48, 488–496. 10.1038/ng.3539.

36. Dickson, D.W., Rademakers, R., and Hutton, M.L. (2007). Progressive Supranuclear Palsy: Pathology and Genetics. Brain Pathol. 17, 74–82. 10.1111/j.1750-3639.2007.00054.x.

37. Kouri, N., Ross, O.A., Dombroski, B., Younkin, C.S., Serie, D.J., Soto-Ortolaza, A., Baker, M., Finch, N.C.A., Yoon, H., Kim, J., et al. (2015). Genome-wide association study of corticobasal degeneration identifies risk variants shared with progressive supranuclear palsy. Nat. Commun. 6, 7247. 10.1038/ncomms8247.

38. Vandrovcova, J., Anaya, F., Kay, V., Lees, A., Hardy, J., and de Silva, R. (2010). Disentangling the role of the tau gene locus in sporadic tauopathies. Curr. Alzheimer Res. 7, 726–734. 10.2174/156720510793611619.

39. Strickland, S.L., Reddy, J.S., Allen, M., N'songo, A., Burgess, J.D., Corda, M.M., Ballard, T., Wang, X., Carrasquillo, M.M., Biernacka, J.M., et al. (2020). MAPT haplotype–stratified GWAS reveals differential association for AD risk variants. Alzheimers Dement. 16, 983–1002. 10.1002/alz.12099.

40. Nalls, M.A., Pankratz, N., Lill, C.M., Do, C.B., Hernandez, D.G., Saad, M., DeStefano, A.L., Kara, E., Bras, J., Sharma, M., et al. (2014). Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. Nat. Genet. 46, 989–993. 10.1038/ng.3043.

41. Jun, G., Ibrahim-Verbaas, CA, Vronskaya, M, Lambert, J.-C, Chung, J, et al. (2016). A novel Alzheimer disease locus located near the gene encoding tau protein. Mol. Psychiatry 18, 108–117. 10.1038/mp.2015.23.

42. Allen, M., Kachadoorian, M., Quicksall, Z., Zou, F., Chai, H.S., Younkin, C., Crook, J.E., Pankratz, V.S., Carrasquillo, M.M., Krishnan, S., et al. (2014). Association of MAPT haplotypes with Alzheimer's disease risk and MAPT brain gene expression levels. Alzheimers Res. Ther. 6, 39. 10.1186/alzrt268.

43. Cooper, Y.A., Teyssier, N., Dräger, N.M., Guo, Q., Davis, J.E., Sattler, S.M., Yang, Z., Patel, A., Wu, S., Kosuri, S., et al. (2022). Functional regulatory variants implicate distinct transcriptional networks in dementia. Science 377, eabi8654. 10.1126/science.abi8654.

44. Corces, M.R., Shcherbina, A., Kundu, S., Gloudemans, M.J., Frésard, L., Granja, J.M., Louie, B.H., Eulalio, T., Shams, S., Bagdatli, S.T., 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. 10.1038/s41588-020-00721-x.

45. Bardy, C., Van Den Hurk, M., Eames, T., Marchand, C., Hernandez, R.V., Kellogg, M., Gorris, M., Galet, B., Palomares, V., Brown, J., et al. (2015). Neuronal medium that supports basic synaptic functions and activity of human neurons in vitro. Proc. Natl. Acad. Sci. U. S. A. 112, E2725–E2734. 10.1073/pnas.1504393112.

46. She, Z.-Y., and Yang, W.-X. (2015). SOX family transcription factors involved in diverse cellular events during development. Eur. J. Cell Biol. 94, 547–563. 10.1016/j.ejcb.2015.08.002.

47. Barry, G., Briggs, J.A., Hwang, D.W., Nayler, S.P., Fortuna, P.R.J., Jonkhout, N., Dachet, F., Maag, J.L.V., Mestdagh, P., Singh, E.M., et al. (2017). The long non-coding RNA NEAT1 is responsive to neuronal activity and is associated with hyperexcitability states. Sci. Rep. 7, 40127. 10.1038/srep40127.

48. Gleeson, J.G., Lin, P.T., Flanagan, L.A., and Walsh, C.A. (1999). Doublecortin Is a Microtubule-Associated Protein and Is Expressed Widely by Migrating Neurons. Neuron 23, 257–271. 10.1016/S0896-6273(00)80778-3.

49. Snetkova, V., and Skok, J.A. (2018). Enhancer talk. Epigenomics 10, 483–498. 10.2217/epi-2017-0157.

50. Fulco, C.P., Nasser, J., Jones, T.R., Munson, G., Bergman, D.T., Subramanian, V., Grossman, S.R., Anyoha, R., Doughty, B.R., Patwardhan, T.A., et al. (2019). Activity-by-contact model of enhancer–promoter regulation from thousands of CRISPR perturbations. Nat. Genet. 51, 1664–1669. 10.1038/s41588-019-0538-0.

51. Lettice, L.A., Heaney, S.J.H., Purdie, L.A., Li, L., de Beer, P., Oostra, B.A., Goode, D., Elgar, G., Hill, R.E., and de Graaff, E. (2003). A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. Hum. Mol. Genet. 12, 1725–1735. 10.1093/hmg/ddg180.

52. Bahr, C., von Paleske, L., Uslu, V.V., Remeseiro, S., Takayama, N., Ng, S.W., Murison, A., Langenfeld, K., Petretich, M., Scognamiglio, R., et al. (2018). A Myc enhancer cluster regulates normal and leukaemic haematopoietic stem cell hierarchies. Nature 553, 515–520. 10.1038/nature25193.

53. Jin, F., Li, Y., Dixon, J.R., Selvaraj, S., Ye, Z., Lee, A.Y., Yen, C.-A., Schmitt, A.D., Espinoza, C.A., and Ren, B. (2013). A high-resolution map of the three-dimensional chromatin interactome in human cells. Nature 503, 290–294. 10.1038/nature12644.

54. Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 485, 376–380. 10.1038/nature11082.

55. Ron, G., Globerson, Y., Moran, D., and Kaplan, T. (2017). Promoter-enhancer interactions identified from Hi-C data using probabilistic models and hierarchical topological domains. Nat. Commun. 8, 2237. 10.1038/s41467-017-02386-3.

56. Doyle, B., Fudenberg, G., Imakaev, M., and Mirny, L.A. (2014). Chromatin Loops as Allosteric Modulators of Enhancer-Promoter Interactions. PLOS Comput. Biol. 10, e1003867. 10.1371/journal.pcbi.1003867.

57. Symmons, O., Uslu, V.V., Tsujimura, T., Ruf, S., Nassari, S., Schwarzer, W., Ettwiller, L., and Spitz, F. (2014). Functional and topological characteristics of mammalian regulatory domains. Genome Res. 24, 390–400. 10.1101/gr.163519.113.

58. Moore, J.E., Purcaro, M.J., Pratt, H.E., Epstein, C.B., Shoresh, N., Adrian, J., Kawli, T., Davis, C.A., Dobin, A., Kaul, R., et al. (2020). Expanded encyclopaedias of DNA elements in the human and mouse genomes. Nature 583, 699–710. 10.1038/s41586-020-2493-4.

59. Pennacchio, L.A., Bickmore, W., Dean, A., Nobrega, M.A., and Bejerano, G. (2013). Enhancers: five essential questions. Nat. Rev. Genet. 14, 288–295. 10.1038/nrg3458.

60. ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature 489, 57–74. 10.1038/nature11247.

61. Luo, Y., Hitz, B.C., Gabdank, I., Hilton, J.A., Kagda, M.S., Lam, B., Myers, Z., Sud, P., Jou, J., Lin, K., et al. (2020). New developments on the Encyclopedia of DNA Elements (ENCODE) data portal. Nucleic Acids Res. 48, D882–D889. 10.1093/nar/gkz1062.

62. Gasperini, M., Tome, J.M., and Shendure, J. (2020). Towards a comprehensive catalogue of validated and target-linked human enhancers. Nat. Rev. Genet. 21, 292–310. 10.1038/s41576-019-0209-0.

63. Pantazis, C.B., Yang, A., Lara, E., McDonough, J.A., Blauwendraat, C., Peng, L., Oguro, H., Kanaujiya, J., Zou, J., Sebesta, D., et al. (2022). A reference human induced pluripotent stem cell line for large-scale collaborative studies. Cell Stem Cell 29, 1685-1702.e22. 10.1016/j.stem.2022.11.004.

64. van Arensbergen, J., Pagie, L., FitzPatrick, V.D., de Haas, M., Baltissen, M.P., Comoglio, F., van der Weide, R.H., Teunissen, H., Võsa, U., Franke, L., et al. (2019). High-throughput identification of human SNPs affecting regulatory element activity. Nat. Genet. 51, 1160–1169. 10.1038/s41588-019-0455-2.

65. Zheng, Y., Shen, W., Zhang, J., Yang, B., Liu, Y.-N., Qi, H., Yu, X., Lu, S.-Y., Chen, Y., Xu, Y.-Z., et al. (2018). CRISPR interference-based specific and efficient gene inactivation in the brain. Nat. Neurosci. 21, 447–454. 10.1038/s41593-018-0077-5.

66. Yeung, T.K., Lau, H.W., Ma, H.T., and Poon, R.Y.C. (2021). One-step multiplex toolkit for efficient generation of conditional gene silencing human cell lines. Mol. Biol. Cell 32, 1320–1330. 10.1091/mbc.E21-02-0051.

67. Dixit, A., Parnas, O., Li, B., Chen, J., Fulco, C.P., Jerby-Arnon, L., Marjanovic, N.D., Dionne, D., Burks, T., Raychowdhury, R., et al. (2016). Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. Cell 167, 1853-1866.e17. 10.1016/j.cell.2016.11.038.

68. Downes, D.J., Smith, A.L., Karpinska, M.A., Velychko, T., Rue-Albrecht, K., Sims, D., Milne, T.A., Davies, J.O.J., Oudelaar, A.M., and Hughes, J.R. (2022). Capture-C: a modular and flexible

approach for high-resolution chromosome conformation capture. Nat. Protoc. 17, 445–475. 10.1038/s41596-021-00651-w.

69. Buckle, A., Gilbert, N., Marenduzzo, D., and Brackley, C.A. (2019). capC-MAP: software for analysis of Capture-C data. Bioinformatics 35, 4773–4775. 10.1093/bioinformatics/btz480.

70. Buenrostro, J.D., Wu, B., Chang, H.Y., and Greenleaf, W.J. (2015). ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. Curr. Protoc. Mol. Biol. 109. 10.1002/0471142727.mb2129s109.

71. Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods 10, 1213–1218. 10.1038/nmeth.2688.

72. Partridge, E.C., Chhetri, S.B., Prokop, J.W., Ramaker, R.C., Jansen, C.S., Goh, S.-T., Mackiewicz, M., Newberry, K.M., Brandsmeier, L.A., Meadows, S.K., et al. (2020). Occupancy maps of 208 chromatin-associated proteins in one human cell type. Nature 583, 720–728. 10.1038/s41586-020-2023-4.

73. Reddy, T.E., Pauli, F., Sprouse, R.O., Neff, N.F., Newberry, K.M., Garabedian, M.J., and Myers, R.M. (2009). Genomic determination of the glucocorticoid response reveals unexpected mechanisms of gene regulation. Genome Res. 19, 2163–2171. 10.1101/gr.097022.109.

74. Johnson, D.S., Mortazavi, A., Myers, R.M., and Wold, B. (2007). Genome-wide mapping of in vivo protein-DNA interactions. Science 316, 1497–1502. 10.1126/science.1141319.

75. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17, 10–12. 10.14806/ej.17.1.200.

76. Sanjana, N.E., Shalem, O., and Zhang, F. (2014). Improved vectors and genome-wide libraries for CRISPR screening. Nat. Methods 11, 783–784. 10.1038/nmeth.3047.

77. Shalem, O., Sanjana, N.E., Hartenian, E., Shi, X., Scott, D.A., Mikkelson, T., Heckl, D., Ebert, B.L., Root, D.E., Doench, J.G., et al. (2014). Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. Science 343, 84–87. 10.1126/science.1247005.

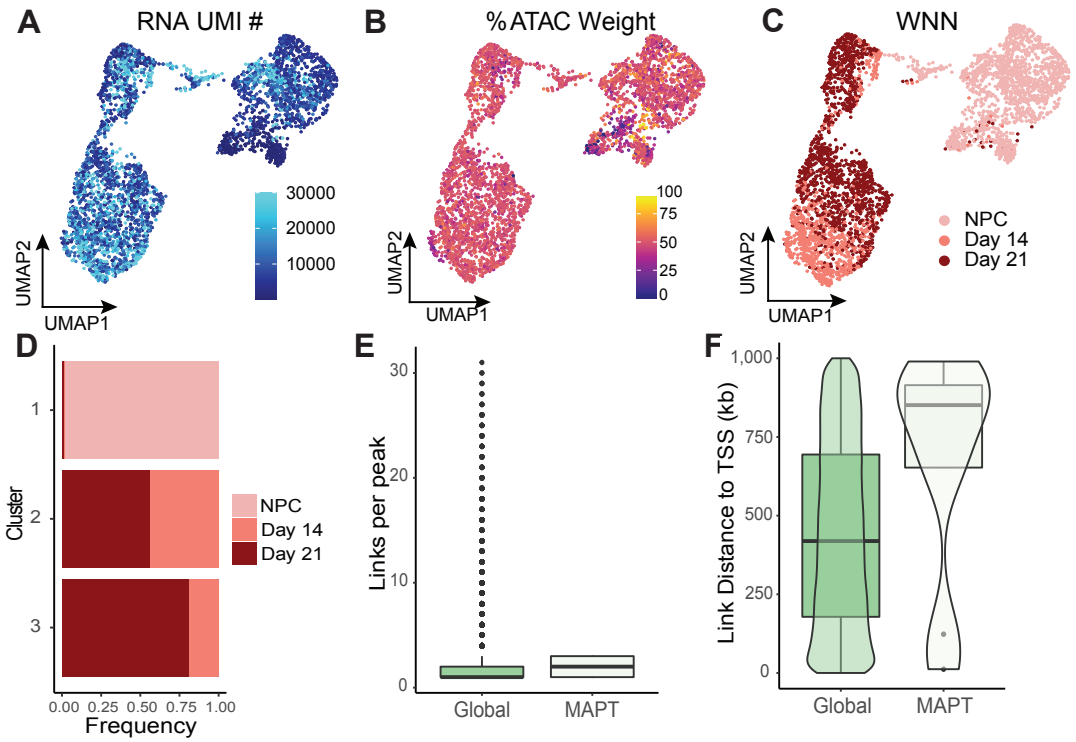

Figure S1. ntegrating snRNA-seq and snATAC-seq and feature linkage description for iPSC-derived neuronal cultures, Related to Figure 1. A) WNN UMAP colored by RNA UMI counts. **B)** WNN UMAP colored by the proportion weight given to the snATAC-seq for creating the WNN graph. **C)** WNN UMAP colored by time point. **D)** The proportion of cells assigned to a cluster from each time point. **E)** Comparison of the distribution of links per ATAC peak of global link calls to MAPT links. **F)** Comparison of the distribution of the distance from linked ATAC peak to linked gene's TSS for global links and MAPT links.

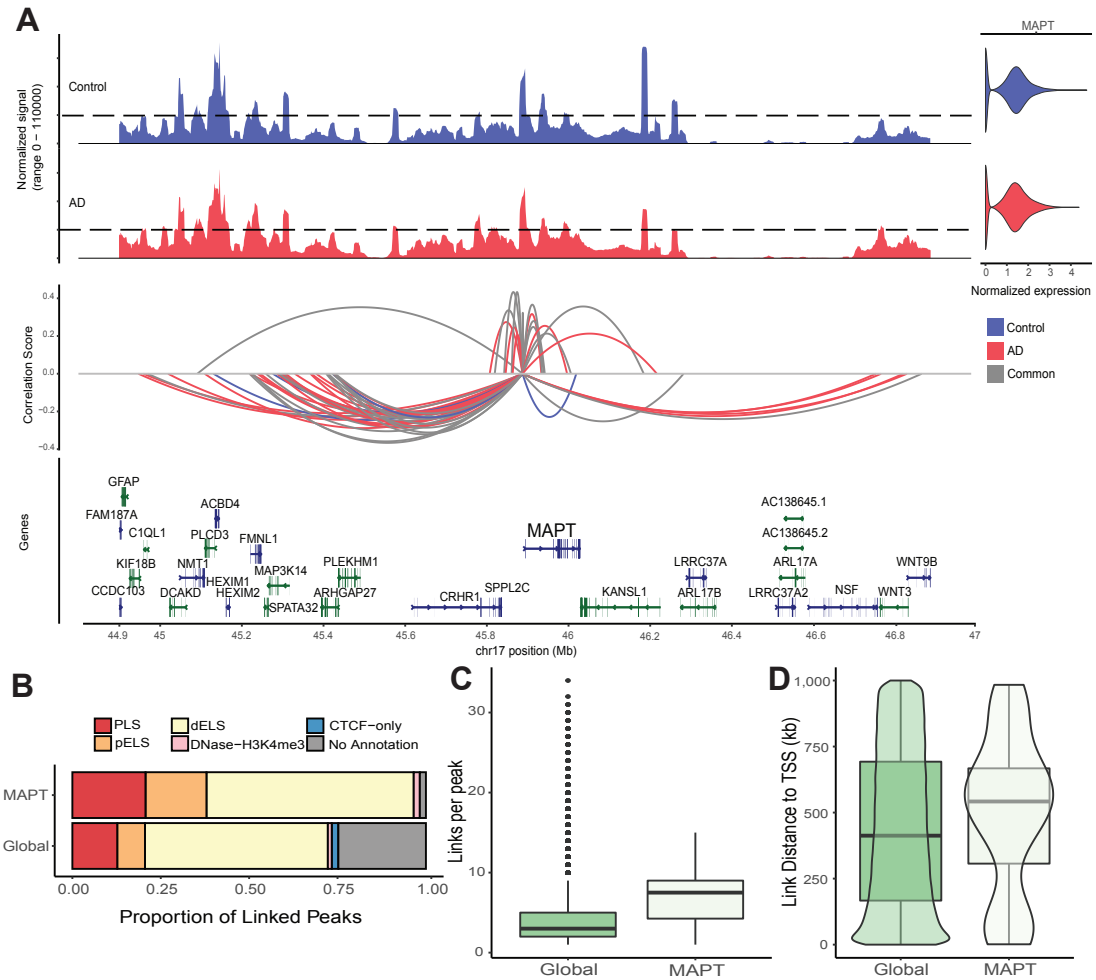

Figure S2. Nomination of candidate CREs using publicly available AD vs control single nucleus multiomics data. **A)** Linkage plot for all links to *MAPT*. Top panels: coverage plot of normalized accessibility across all cell types separated by donor status (red = AD; blue = control). Bottom panel: significant AD and control peak-gene links. Arc height represents strength and direction of the correlation. Arc color indicates if the link was AD-specific (red), control-specific (blue), or shared in both AD and control ("common", gray). **B)** Comparison of proportion of linked peaks by ENCODE regulatory element annotation of global and *MAPT* links. **C)** Comparison of the distribution of links per ATAC peak of global links to *MAPT* links. D) Comparison of the distribution of the distance from linked ATAC peak to linked gene's TSS of global links to *MAPT* links.

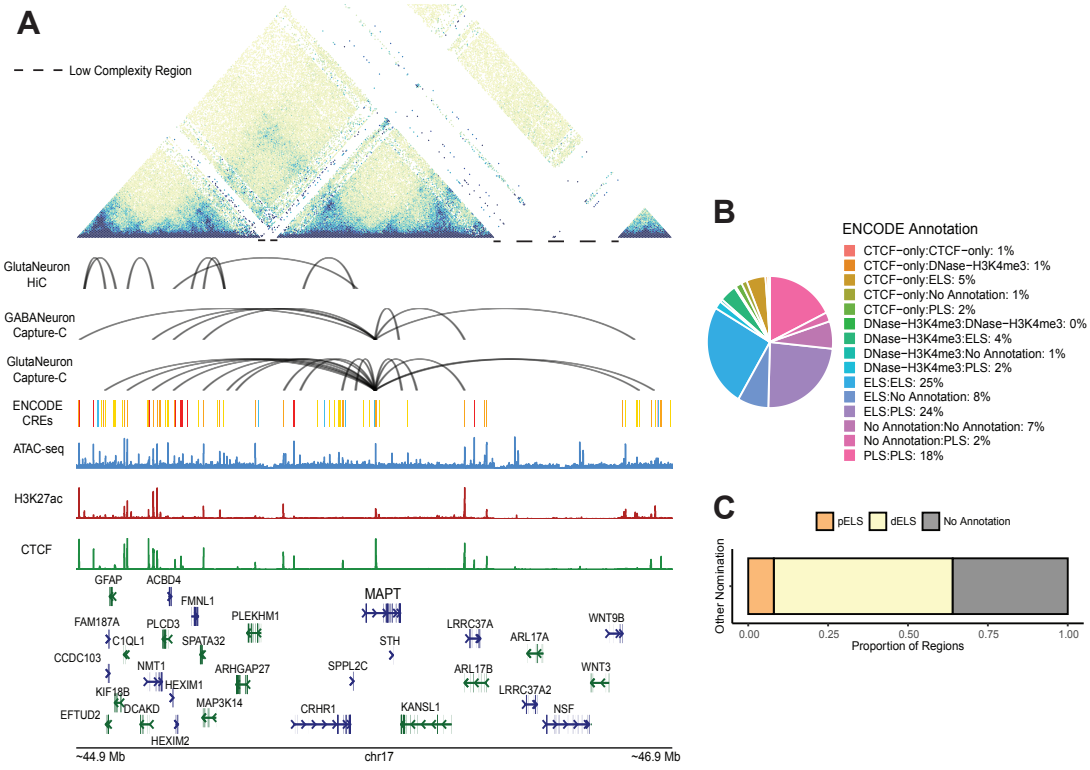

Figure S3. Description of chromatin annotation and structural analysis nominated CREs, Related to Figure 2. A) Loop plot of HiCCUPs nominated loops in the *MAPT* locus. Top panel: Triangular plot showing TAD hierarchy. White spaces underlined with dashed lines represent regions of low complexity where sequences could not be adequately mapped. Middle panel: HiCCUPs loops in *MAPT* locus for HiC in iCell GlutaNeurons (top), Capture-C in iCell GlutaNeurons (middle), and Capture-C in iCell GABANeurons (bottom). Bottom panel: Track displaying ENCODE cCREs in the locus. ATAC-seq peaks and H3K27ac and CTCF ChIP-seq peaks generated from iPSC-derived neuronal cultures differentiated for 14 days. **B)** Distribution of HiC loops by ENCODE annotation. **C)** Distribution of regions nominated from chromatin annotations and previously published data by ENCODE annotation.

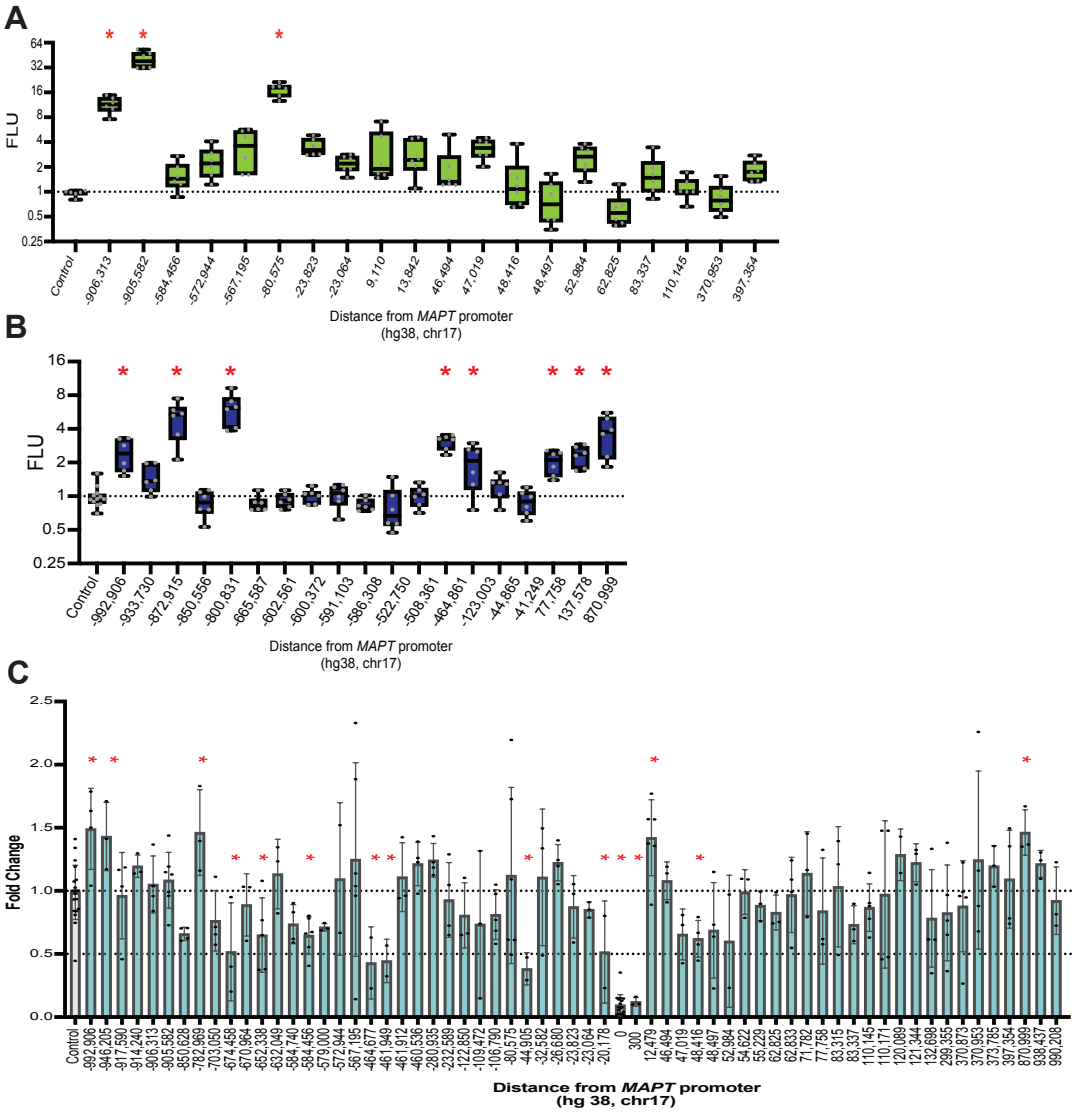

Figure S4. All functional validation, Related to Figure 4. **A)** Boxplots showing elements tested in luciferase assays in iCell GlutaNeurons (*p<0.05, ANOVA with Fisher's LSD). **B)** Boxplots showing elements tested in luciferase assays in KOLF2.1J-hNGN2 neurons (*p<0.05, ANOVA with Fisher's LSD). **C)** RT-qPCR of *MAPT* expression. Barplots show regions tested in Day 14 BrainPhys differentiated neurons by CRISPRi experiments (*p<0,05, ANOVA with Fisher's LSD).

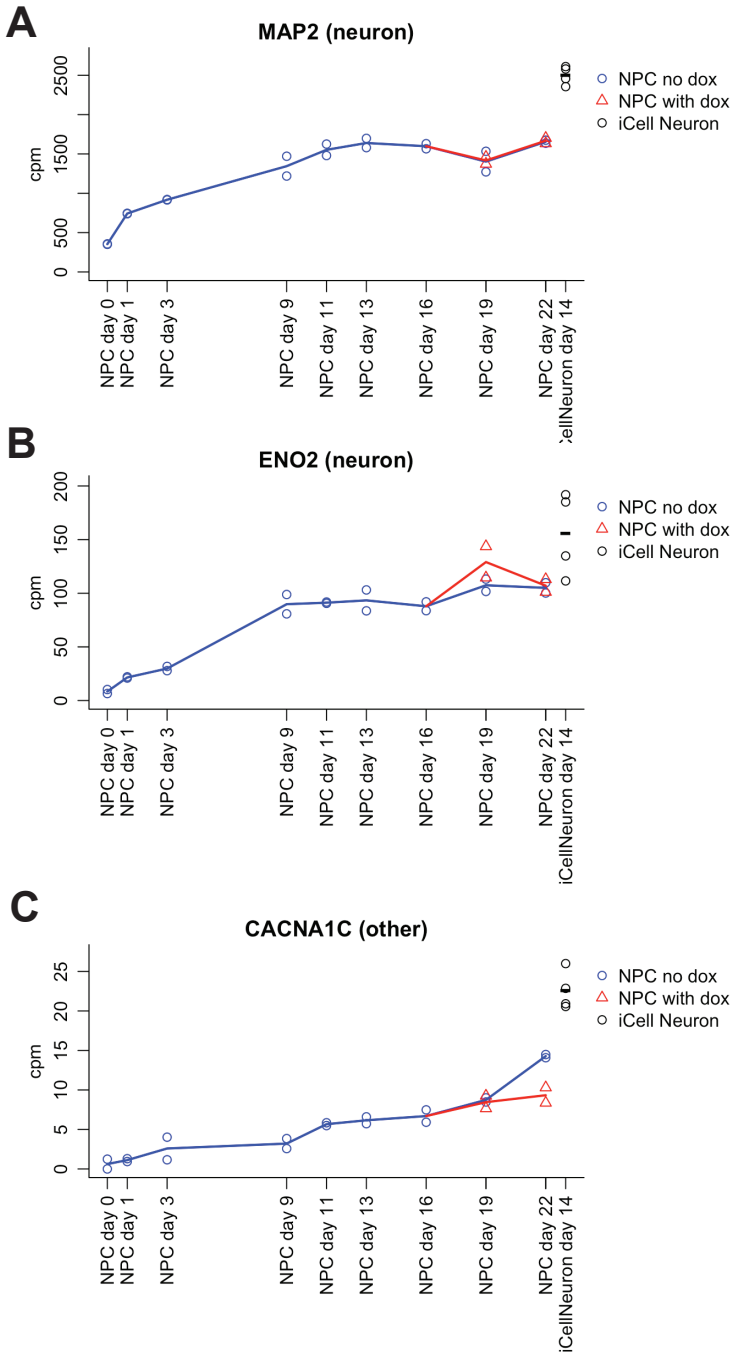

Figure S5. BrainPhys neuron differentiation timecourse for calculating differentiation score, Related to Methods. Expression of key genes (**A)** *MAP2*, **B)** *ENO2*, and **C)** *CACNA1C*) used to determine level of differentiation of each replicate of differentiated neurons for CRISPRi experiments across 22 days compared to pure culture of excitatory neurons (iCell GlutaNeuron day 14).