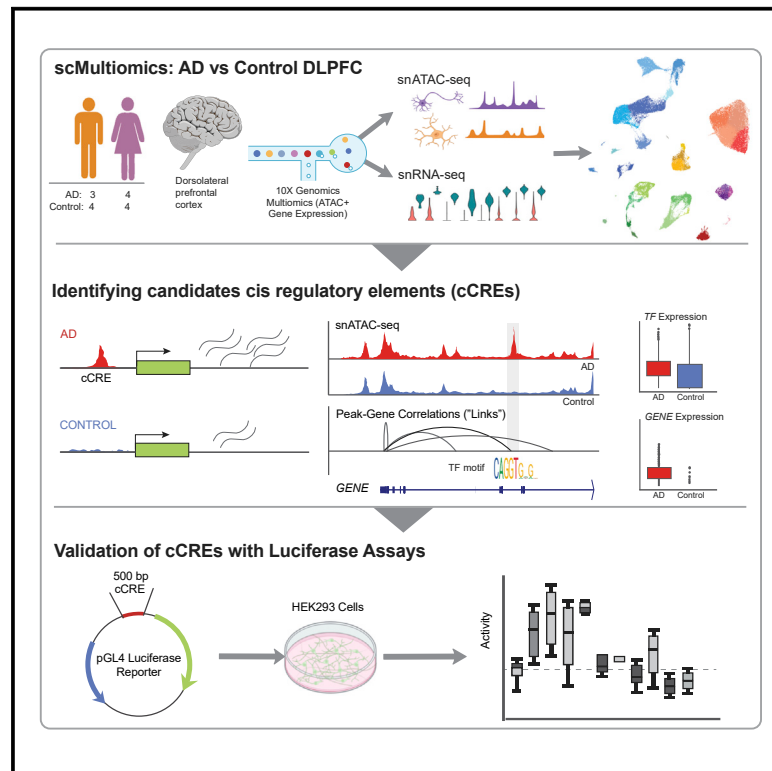


# Single nucleus multiomics identifies ZEB1 and MAFB as candidate regulators of Alzheimer's disease-specific *cis*-regulatory elements

## Graphical abstract



## Authors

Ashlyn G. Anderson, Brianne B. Rogers, Jacob M. Loupe, ..., J. Nicholas Cochran, Richard M. Myers, Lindsay F. Rizzardi

## Correspondence

ncochran@hudsonalpha.org (J.N.C.),  
rmyers@hudsonalpha.org (R.M.M.),  
lrizzardi@hudsonalpha.org (L.F.R.)

## In brief

Linking regulatory elements to target genes in specific cell types can reveal regulatory mechanisms disrupted in disease. Here, Anderson and Rogers et al. identify Alzheimer's-specific elements in cortical tissues and candidate transcription factors regulating their activity. Enhancer-like activity was demonstrated for twelve elements including several regulators of *APP*.

## Highlights

- Single nucleus multiomics to identify cell type-specific regulatory elements in Alzheimer's
- Identified 40,831 candidate regulatory elements specific to Alzheimer's
- Confirmed enhancer-like activity of candidate regulators of *APP* expression
- Nominated ZEB1 and MAFB as candidate Alzheimer's-specific *trans*-regulators



## Article

# Single nucleus multiomics identifies ZEB1 and MAFB as candidate regulators of Alzheimer's disease-specific *cis*-regulatory elements

Ashlyn G. Anderson,<sup>1,5</sup> Brianne B. Rogers,<sup>1,2,5</sup> Jacob M. Loupe,<sup>1</sup> Ivan Rodriguez-Nunez,<sup>1</sup> Sydney C. Roberts,<sup>1</sup> Lauren M. White,<sup>1</sup> J. Nicholas Brazell,<sup>1</sup> William E. Bunney,<sup>3</sup> Blynn G. Bunney,<sup>3</sup> Stanley J. Watson,<sup>4</sup> J. Nicholas Cochran,<sup>1,6,\*</sup> Richard M. Myers,<sup>1,6,\*</sup> and Lindsay F. Rizzardi<sup>1,6,7,\*</sup>

<sup>1</sup>HudsonAlpha Institute for Biotechnology, Huntsville, AL, USA

<sup>2</sup>University of Alabama at Birmingham, Birmingham, AL, USA

<sup>3</sup>Department of Psychiatry and Human Behavior, College of Medicine, University of California, Irvine, Irvine, CA, USA

<sup>4</sup>Mental Health Research Institute, University of Michigan, Ann Arbor, MI, USA

<sup>5</sup>These authors contributed equally

<sup>6</sup>Senior author

<sup>7</sup>Lead contact

\*Correspondence: [ncochran@hudsonalpha.org](mailto:ncochran@hudsonalpha.org) (J.N.C.), [rmyers@hudsonalpha.org](mailto:rmyers@hudsonalpha.org) (R.M.M.), [lrizzardi@hudsonalpha.org](mailto:lrizzardi@hudsonalpha.org) (L.F.R.)  
<https://doi.org/10.1016/j.xgen.2023.100263>

## SUMMARY

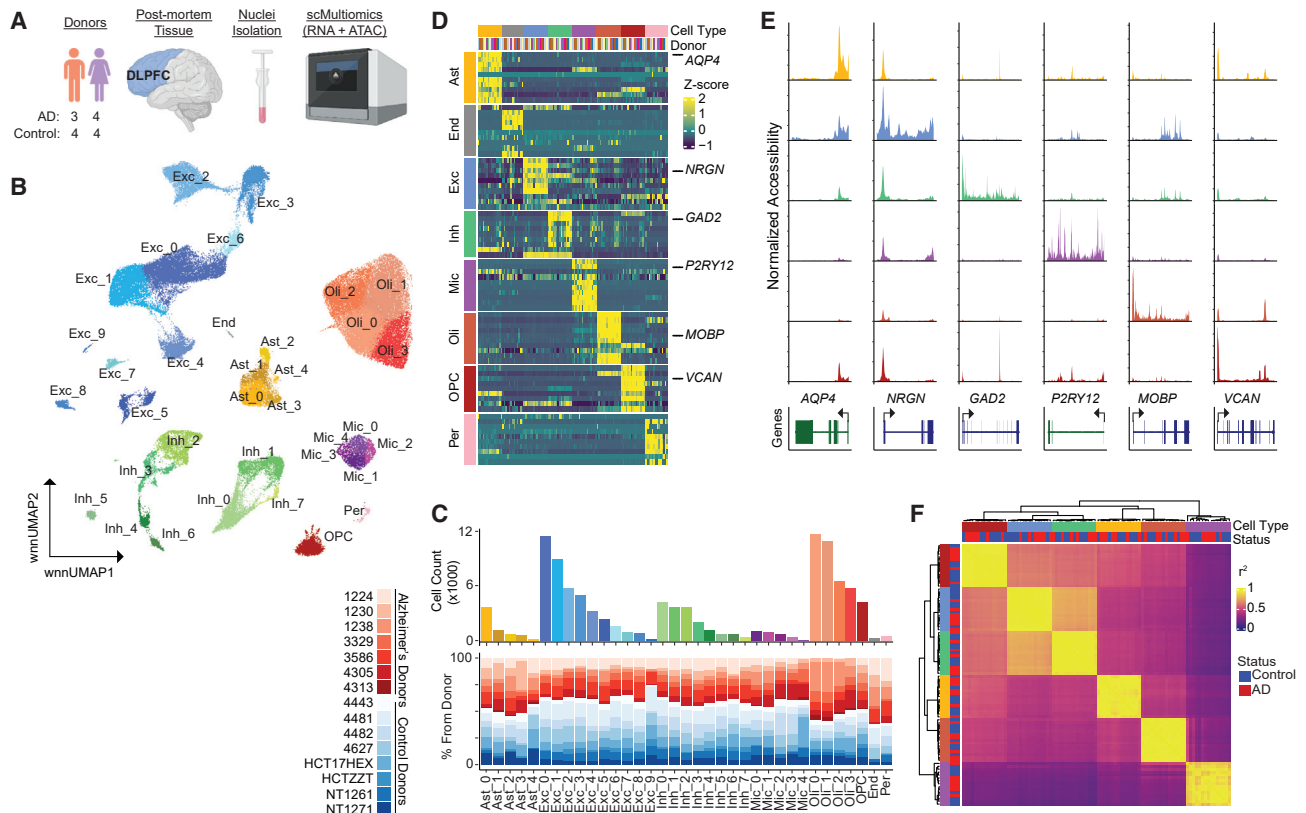
Cell type-specific transcriptional differences between brain tissues from donors with Alzheimer's disease (AD) and unaffected controls have been well documented, but few studies have rigorously interrogated the regulatory mechanisms responsible for these alterations. We performed single nucleus multiomics (snRNA-seq plus snATAC-seq) on 105,332 nuclei isolated from cortical tissues from 7 AD and 8 unaffected donors to identify candidate *cis*-regulatory elements (CREs) involved in AD-associated transcriptional changes. We detected 319,861 significant correlations, or links, between gene expression and cell type-specific transposase accessible regions enriched for active CREs. Among these, 40,831 were unique to AD tissues. Validation experiments confirmed the activity of many regions, including several candidate regulators of *APP* expression. We identified ZEB1 and MAFB as candidate transcription factors playing important roles in AD-specific gene regulation in neurons and microglia, respectively. Microglia links were globally enriched for heritability of AD risk and previously identified active regulatory regions.

## INTRODUCTION

Identification of genetic contributors to Alzheimer's disease (AD) has provided critical insights into potential disease mechanisms. Rare, protein-altering variants in *APP*, *PSEN1*, or *PSEN2* cause early-onset, autosomal-dominant AD,<sup>1</sup> and genome-wide association studies (GWASs) have identified common variants for late-onset AD that increase disease risk to varying degrees.<sup>2–6</sup> However, the majority of GWAS variants are in non-coding regions of the genome and many presumably affect gene regulation. Linkage disequilibrium makes identification of the causal variant difficult, particularly for putative regulatory regions where conservation and deleteriousness estimates may not be as informative. Associating common and rare regulatory variants with affected genes is also challenging.<sup>7–9</sup> In addition, disease-associated variants often function only in specific cell types, further complicating interpretation of their effects.<sup>10,11</sup> Thus, determining which genes are contributing to disease requires assessments in specific cell types.

Recent advances in single cell technologies have allowed profiling of gene expression<sup>12–18</sup> and chromatin accessibility,<sup>10</sup> either separately or in parallel from the same samples.<sup>19,20</sup> Although studies have examined the cell type-specific transcriptional and epigenetic differences between tissues from brain donors with AD and unaffected controls, few have rigorously interrogated the regulatory mechanisms responsible for these alterations.<sup>11,21</sup> Integrating single nucleus RNA-seq (snRNA-seq) and single nucleus assay for transposase-accessible chromatin sequencing (snATAC-seq) data allow the identification of potential *cis*-regulatory elements (CREs) by correlating chromatin accessibility with nearby gene expression. Here, we simultaneously measure both gene expression and chromatin accessibility in the same nuclei to identify cell type-specific regulatory regions and their target genes in dorsolateral prefrontal cortex (DLPFC) tissues from both AD and unaffected donors. In addition, we assessed regulatory mechanisms unique to nuclei from donors with AD.





**Figure 1. Cellular diversity of DLPFC from Alzheimer's disease and unaffected donors revealed by single cell multiomics**

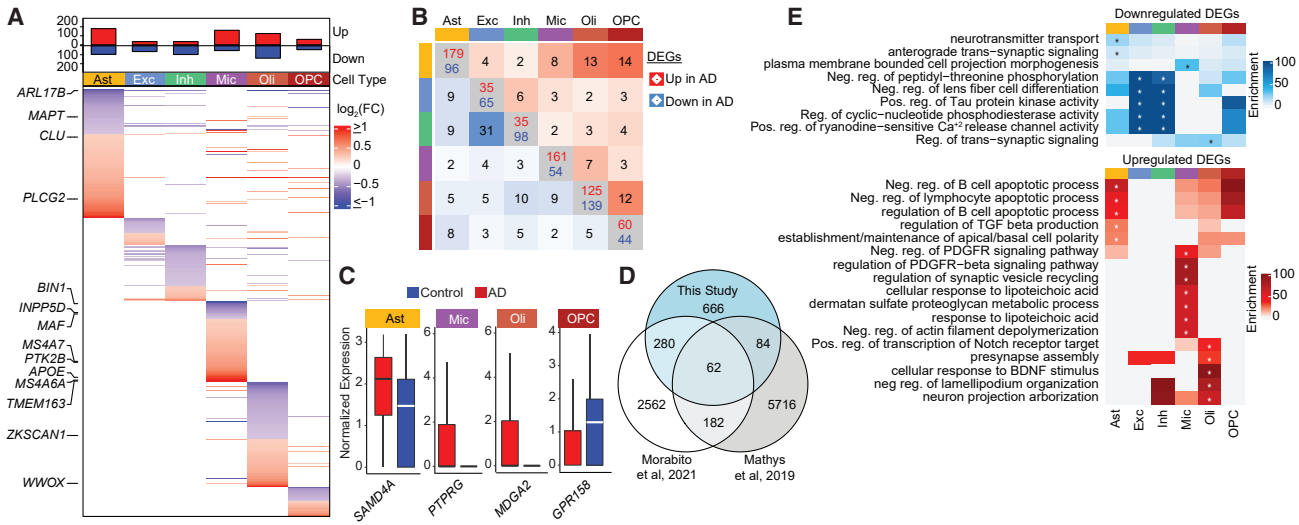
- (A) Experimental design.  
 (B) Uniform manifold approximation and projection (UMAP) visualization of the weighted nearest neighbor (WNN) clustering of single nuclei colored by cell type and cluster assignment.  
 (C) Total number of cells in each subcluster and the proportion of cells from each individual (red, AD donors; blue, unaffected donors) in the subcluster.  
 (D) Row-normalized gene expression of scREAD cell type markers.  
 (E) Chromatin accessibility across cell types for cell type marker genes (indicated below).  
 (F) Correlation of pseudo-bulked cell type-specific expression profiles between individuals. Colors indicating cell type are consistent throughout the figure.

## RESULTS

### Cellular diversity within the human dorsolateral prefrontal cortex

We used the 10x Genomics Multiome technology to perform snATAC-seq and snRNA-seq on nuclei isolated from human postmortem DLPFC tissues from seven individuals diagnosed with AD (mean age 78 years, Braak stages 4–6) and eight sex-matched unaffected control donors (mean age 63 years) (Table S1; Figure 1A). This assay allows direct mapping of both gene expression and chromatin accessibility within the same nuclei without the need to computationally infer cell type identification during cross-modality integration. After removing low quality nuclei and doublets (STAR Methods), we retained a total of 105,332 nuclei with an average of 7,022 nuclei per donor (range 1,410–11,723). We detected a median of 2,659 genes and 11,647 ATAC fragments per cell. We performed normalization and dimensionality reduction for snRNA-seq and snATAC-seq data using Seurat<sup>22</sup> and Signac,<sup>23</sup> respectively. We used weighted nearest neighbor (WNN) analysis to determine a joint

representation of expression and accessibility and identified 36 distinct clusters composed of eight major cell types and their associated subclusters (Figures 1B, S1A, and S1B). Consistent with previous snRNA-seq datasets,<sup>12,13,15,19</sup> we identified all expected cell types in the brain with similar relative abundances across AD and control donors (Figures 1B, 1C, and S1C). Pericytes and endothelial cell clusters contained <500 nuclei and were excluded from further analyses. Cluster annotations were supported by both gene expression and promoter accessibility of well-established cell type marker genes (Figures 1D and 1E). There were strong correlations in global gene expression across donors within each cell type and between excitatory/inhibitory neurons (Figure 1F). The only cell type to display variable correlation values across donors was microglia, a cell type known to be dysregulated in AD. In addition, we identified distinct subpopulations within each major cell type with the exception of oligodendrocyte precursor cells (OPCs), pericytes, and endothelial cells (Figure S2). These subtype annotations were consistent with those from prior studies,<sup>14,22,24</sup> and distributions were similar across AD and control donors, with the exception of



**Figure 2. Cell type-specific transcriptome dysregulation in Alzheimer's DLPPFC**

- (A) MAST  $\log_2(\text{fold change [FC]})$  of all up- and downregulated genes in AD for each cell type.  
 (B) Number of shared DEGs between cell types in both directions (upper triangle, upregulated in AD; lower triangle, downregulated in AD).  
 (C) Normalized expression of the top DEGs in the indicated cell types ( $\log_2[\text{FC}] > 1$ ).  
 (D) Overlap of DEGs with agreement on cell type and direction with Morabito et al.<sup>19</sup> and Mathys et al.<sup>12</sup>  
 (E) Heatmap showing the odds ratio of the top enrichR GO terms for up and downregulated DEGs within each cell type (\*adjusted  $p < 0.01$ ).

microglia subpopulations and two inhibitory neuron subtypes (Inh\_1 and Inh\_2; Figure S2).

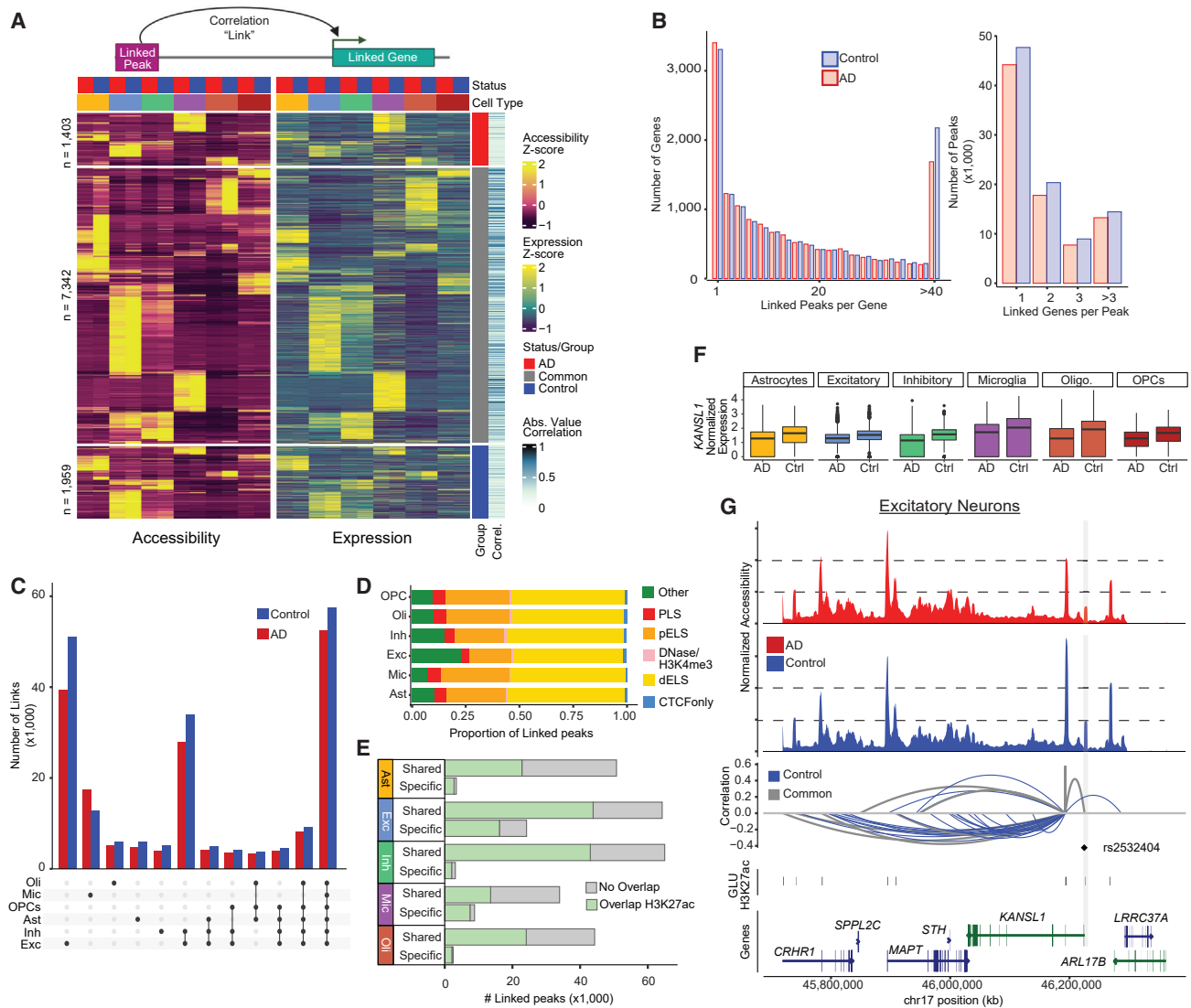
### Cell type-specific transcriptome changes in Alzheimer's DLPPFC

Within each cell type, we identified differentially expressed genes (DEGs) between AD and control tissues. A total of 911 DEGs were identified after considering sex and age as covariates (Figure 2A; Table S3). Although significant sex-specific differences in gene expression between AD and controls have been shown previously,<sup>24</sup> because of our smaller sample size we did not detect such changes. Although the majority of DEGs were cell type specific, 141 were identified across multiple cell types (Figure 2B). Of these DEGs, 62 were also identified as differentially up- or downregulated in the same cell type in both Mathys et al.<sup>12</sup> and Morabito et al.,<sup>19</sup> including *PTPRG* (up in AD Mic) and *GRIA2* (down in AD Ast) (Figures 2C and 2D). *PTPRG* encodes a protein tyrosine phosphatase associated with inflammation<sup>25</sup> and AD disease risk (rs7609954).<sup>26</sup> *GRIA2* encodes the glutamate receptor 2 (GluR2) subunit that reduces calcium channel permeability and may protect against excitotoxicity.<sup>27</sup> A recent meta-analysis<sup>28</sup> of three snRNA studies<sup>12,13,18</sup> found 41 DEGs that were shared across these three studies, only 7 of which were discordant in direction. We also identified 17 of them, including *GRM3* (down in AD Ast) and *SLC38A2* (up in AD Oli) involved in glutamate signaling and *RNF149* (up in AD Mic) encoding an E3 ubiquitin ligase. Most DEGs were upregulated in AD and were enriched for cell type-specific Gene Ontology (GO) terms related to PDGFR beta signaling in microglia, apoptosis in astrocytes, and Notch and BDNF signaling in oligodendrocytes (Figure 2E; Table S4). In contrast, most DEGs downregulated in AD were in neurons and showed enrichment

in GO terms related to regulation of tau activity (*HSP90AB1*, *HSP90AA1*) and calcium channel activity (*CALM2*, *CALM3*) (Figure 2E; Table S4). Example DEGs include *MDGA2* whose overexpression results in reduced excitatory synapse density,<sup>29</sup> *GPR158* which is an osteocalcin receptor associated with cognition<sup>30</sup> and *SAMD4A* whose expression is correlated with AD-associated neuropathology and cognitive status (<https://agora.adknowledgeportal.org>). *SAMD4A*, along with several other DEGs (*NRXN1*, *LUZP2*, *RBMS3*, *ARHGAP15*, *ARHGAP24*, *LRP1B*, and *ATP1B3*), was also newly identified as AD associated in a recent genome-wide association neural network analysis of AD family history in the UK Biobank.<sup>31</sup>

### Identification of candidate CREs

Previous single cell studies have characterized altered gene expression in AD brain tissues and cell types,<sup>12–14,17,19</sup> and we observed signals consistent with those studies. Additionally, we sought to leverage single cell multiomics data to identify cell type- and disease-specific CREs and their target genes by correlating gene expression with chromatin accessibility across all nuclei in the dataset. The Cell Ranger ARC (v2.0) analysis pipeline produces these correlations as “feature linkages” (STAR Methods). A feature linkage, or link, is defined as a significant correlation between accessibility of an ATAC peak and the expression of a gene<sup>32</sup> (Figure 3A). We restricted this correlation analysis to consider only peaks within 500 kb of each transcription start site (TSS), a generous search space, as previous studies have found the majority of enhancers are within 50–100 kb of their target genes,<sup>33</sup> but chosen to allow the identification of longer range interactions that can have high impact.<sup>34</sup> We first took the union of ATAC peaks identified in each cell type and retained only those present in  $\geq 2\%$  of cells in at least one cell type



**Figure 3. Identification of candidate CREs**

(A) Schematic of gene-peak association (top). Heatmap of row-normalized accessibility and expression for the most correlated peak-gene link for each gene (bottom). Columns are pseudo-bulked on cell type and disease status.

(B) Distributions of the number of linked peaks per gene (left) and the number of linked genes per peak (right) for AD (red) and control (blue) samples.

(C) Total number of links per cell type for AD and control. Cell type of the link is assigned by the cell type in which the peak was called.

(D) ENCODE annotation of linked peaks by cell type.

(E) Shared (across cell types) and cell type-specific linked peaks that overlap H3K27ac of the corresponding cell type.

(F) Normalized expression of *KANSL1* from AD and control samples in each cell type. Expression is significantly different in AD versus control for all cell types.

(G) Linkage plot for all links to *KANSL1*. Top: coverage plot of pseudo-bulked accessibility in excitatory neurons separated by status (red, AD; blue, control). Bottom: significant AD and control peak-gene links. Arc height represents strength and direction of correlation. Arc color indicates if the link was identified in both AD and control (common, gray) or control donors only (blue). A linked peak overlapping a single SNP is highlighted in gray.

for a total of 189,925 peaks. Nearly half of all peaks overlapped H3K27ac (46%) from the corresponding cell type, and 43% overlapped ENCODE (Encyclopedia of DNA Elements)<sup>35</sup> distal enhancer-like sequences. Using this peak set, links were then calculated independently using gene expression data from either AD or control nuclei allowing classification of links as AD-specific, control-specific, or common (STAR Methods). We consider the linked peaks to be candidate CREs. Given the gene expres-

sion changes observed in AD, we hypothesized that there would be differential use of CREs between AD and control samples that would be identified in this analysis as AD- or control-specific links. Cell type specificity of each link was determined by the cell type(s) in which the ATAC peak was identified. A total of 319,905 peak-gene links were found involving 15,471 linked genes and 126,213 linked peaks with a minimum absolute correlation value of 0.2 (Figure 3A; Table S5). The median distance

between the linked peak and the TSS of the linked gene was 201,506 bp, and there was an inverse relationship between absolute correlation value and distance to TSS (Figure S3A).

For most genes, we identified a similar number of links in both AD (median = 12) and control samples (median = 13). However, we found 1,294 genes that had only AD links and 1,596 that had only control links (Figure S3B). We observed no significant bias when comparing the number of links identified in either AD or control for a given gene (Figure S3B). Most genes were linked to multiple peaks across all cell types, with a median of 14 linked peaks per gene. However, 16% of genes were linked with 40 or more peaks (Figure 3B), and these genes were significantly longer and more highly expressed than those with fewer links (Figure S3C). This finding is likely due to links being called for peaks within the gene body of longer genes as excluding these peaks abolishes the difference in number of links ( $p = 0.12$ ,  $t$  test). Across the entire dataset, 17.8% of linked peaks are present in the promoter or gene body of the target gene. Although positively correlated links in gene bodies may often be merely a consequence of target gene expression, we retained these peaks in our analyses as enhancers are often located within the introns of their target genes.

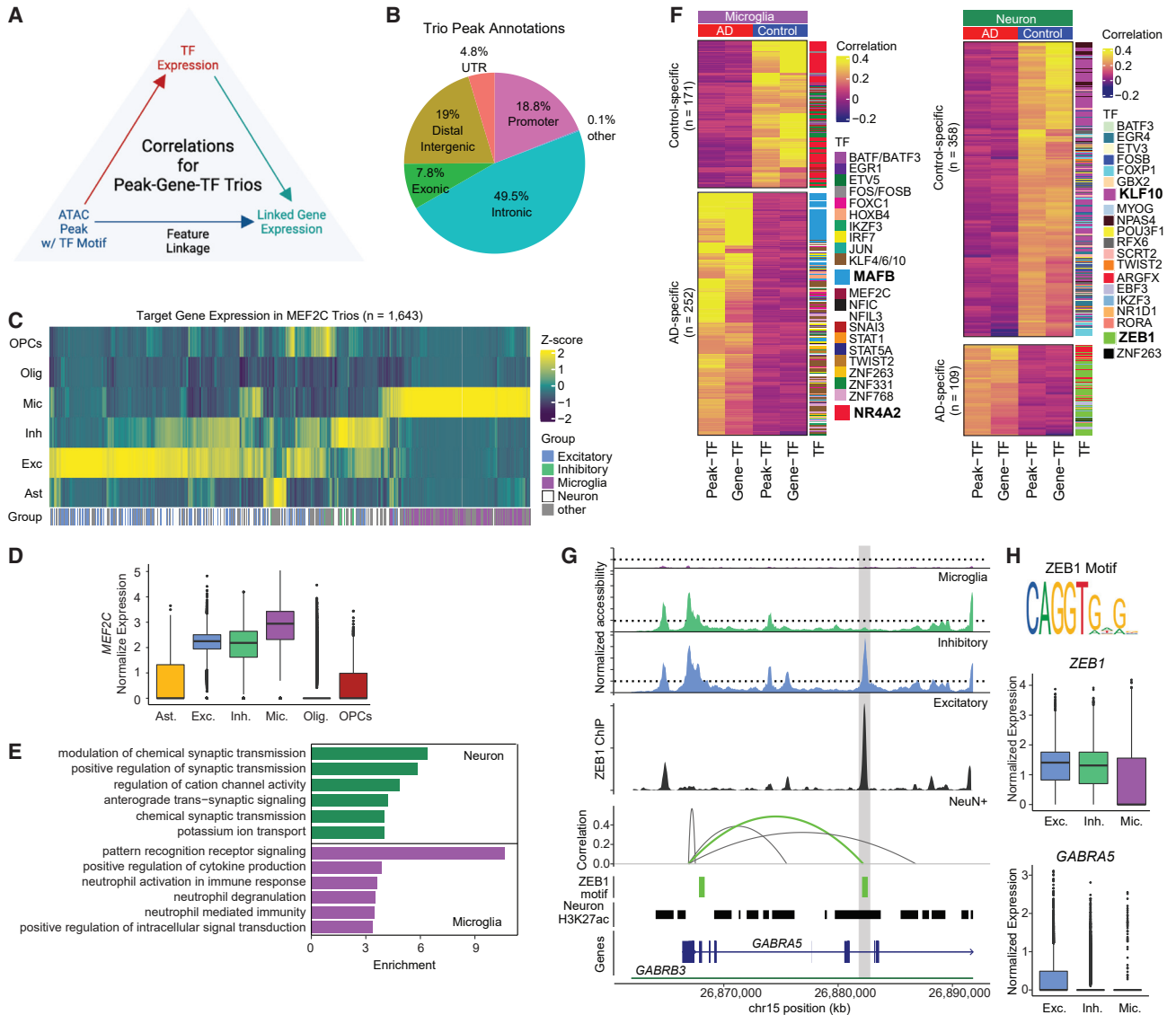
ATAC peaks often interacted with more than one gene. Nearly 70% (126,213) of the ATAC peaks analyzed were linked to a gene with an average of two genes linked to each peak and a range of 1–21 linked genes (Figure 3B). Almost a third (30.24%) of the links were unique to a single cell type, while 21% were common across all cell types (Figure 3C). We identified 40,831 AD-specific links and 74,028 control-specific links with the majority of links identified in both (205,046). We performed permutation analyses and determined that this fraction of AD- and control-specific links (0.36 of total links) was greater than expected by chance ( $p = 0.027$ ,  $Z$  test; Figure S3D). Target genes of cell type-specific links identified in both AD and control samples were enriched in expected pathways (Figure S3E). To evaluate whether linked peaks associate with regulatory regions, we evaluated their overlap with a curated set of candidate CREs identified by ENCODE.<sup>35</sup> We found that linked peaks were significantly enriched for proximal (odds ratio [OR] = 1.24,  $p = 2.4 \times 10^{-15}$ ) and distal (OR = 1.06,  $p = 3.06 \times 10^{-9}$ ) enhancer-like sequences and the proportion of overlap was similar across cell types (Figure 3D). As these annotations were not generated in our particular cell types and tissue, we also intersected these linked peaks with regions of H3K27ac previously identified within cell types isolated from prefrontal cortex tissues.<sup>11,36</sup> We found that on average, 57.5% of linked peaks overlap a H3K27ac peak from the corresponding cell type and this increases to 79% for cell type-specific linked peaks (Figure 3E). The majority (76.11%) of linked peaks were positively correlated with gene expression, as is expected given the association between open chromatin and transcriptional activation,<sup>37–41</sup> though negative correlations may be indicative of repressor binding.<sup>37,40</sup>

We then asked if the differential gene expression we observed could be mediated by candidate CREs. Nearly all (94%) the DEGs identified between AD and control nuclei had a linked peak in the same cell type where the gene was differentially expressed and 85% of these linked peaks overlapped H3K27ac in the same cell type. In addition, we observed that some links to

differentially expressed genes were uniquely identified in either AD or control datasets. For genes upregulated in AD, 72% of their positively correlated links were AD specific, while for downregulated genes 62% were control specific. For example, *BIN1* expression is significantly reduced in AD microglia compared with controls and this reduced expression hampers proinflammatory microglial responses.<sup>42</sup> We identified six control-specific links and no AD-specific links for *BIN1* in microglia. One of these control-specific links was validated as a microglia-specific *BIN1* enhancer in Nott et al.<sup>11</sup> and harbors an AD-associated SNP (rs733839). Together, these findings suggest that this CRE may no longer be used in AD microglia, leading to lower *BIN1* expression, though it remains possible that these observations could also result from less sensitive detection with lower expression. Another example is *KANSL1*, located in the *MAPT* locus, that was downregulated in AD in all cell types (Figure 3F). Twenty-eight of the 37 *KANSL1* linked peaks are unique to control samples and the rest are common to both AD and control (Figure 3G). One of these linked peaks found in the promoter overlaps an expression quantitative trait locus (eQTL)<sup>43</sup> (rs2532404) associated with progressive supranuclear palsy<sup>44</sup> and was recently shown via CRISPRi to regulate *KANSL1* expression in induced pluripotent stem cell (iPSC)-derived neurons.<sup>21</sup> These results are consistent with the hypothesis that other AD- and control-specific regions contribute to the differential expression observed in AD.

### Identification of AD-specific peak-gene-TF trios

We next sought to identify transcription factors (TFs) that could be driving the positive correlations between accessible peaks and target genes. To further investigate the regulatory roles of links, we identified peak-gene-TF “trios” in which (1) there was a correlation between the linked peak and linked gene, (2) the accessibility of a linked peak harboring a specific TF motif was correlated with the expression of that TF, and (3) the expression of the TF was correlated with the expression of the linked gene (Figure 4A; STAR Methods). This approach is conceptually similar to a recently described method called TRIPOD that employs nonparametric models to identify peak-gene-TF associations and from which we adopted the term “trio.”<sup>45</sup> To identify trios, we performed these additional correlation analyses (linked peak:TF expression and TF expression:target gene expression) separately using either AD or control datasets to enable identification of TFs whose activities may be associated with disease. AD- or control-specific trios were those uniquely identified in the AD or control dataset, respectively. Cell type specificity was defined on the basis of the cell type in which the linked peak was identified. We restricted these analyses to links with a correlation value >0.3 that were within 100 kb of the linked gene’s TSS (115,107) and identified 60,120 peak-gene-TF trios involving 17,149 unique peaks and 437 TFs (Table S6). Fewer than 20% of the peaks in these trios are found in promoters, with the majority present in intronic regions (Figure 4B). Trio peaks were enriched for ENCODE distal (OR = 1.26,  $p = 2.2 \times 10^{-16}$ ) and proximal (OR = 1.12,  $p = 5.9 \times 10^{-7}$ ) enhancer-like sequences. There was a median of 37 trios per TF. The TF *MEF2C* was the most common trio participant, appearing in nearly 5% of all trios. Although *MEF2C* was expressed in most



**Figure 4. Identification of AD-specific TF regulatory networks**

(A) Strategy for defining peak-gene-TF trios. A linked peak containing a TF motif must be correlated with that TF and the expression of that TF must be correlated with the linked gene for that peak to be considered a part of trio.

(B) Genome annotations for location of linked peaks within trios.

(C) Heatmap of column-normalized expression of genes within MEF2C trios by cell type.

(D) Normalized expression of *MEF2C* by cell type.

(E) Top enriched GO terms for genes within MEF2C trios from excitatory and inhibitory neurons (green, "neuron") and microglia (purple, "microglia").

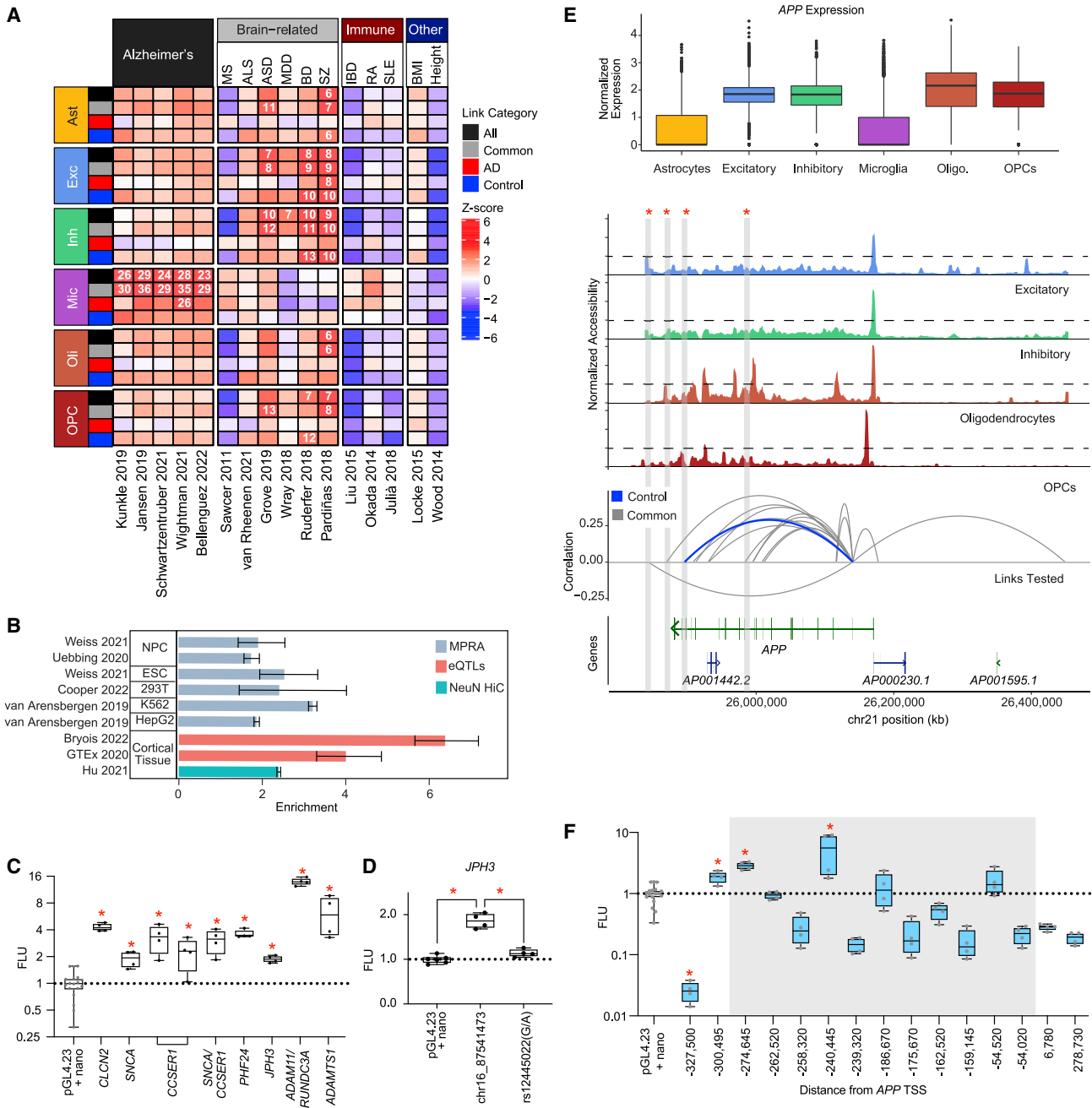
(F) Heatmap of correlation values of AD and control-specific trios identified in microglia (left) and excitatory/inhibitory neurons (right) for TFs involved in at least 3 trios.

(G) Linkage plot for *GABRA5*. Top: coverage plot of pseudo-bulked accessibility in indicated cell types. Middle: coverage plot of ZEB1 ChIP-seq signal from NeuN<sup>+</sup> nuclei isolated from DLPFC tissue from two unaffected donors (1238 and 1242). Bottom: significant peak-gene links; green indicates overlap with ZEB1 motif. Arc height represents strength and direction of correlation. Track of ZEB1 motifs (green) and H3K27ac peaks from neurons (black; Nott et al.<sup>11</sup>). Linked peak of interest is highlighted in gray.

(H) ZEB1 motif from JASPAR 2022 (top). Normalized expression of *ZEB1* and *GABRA5* in excitatory/inhibitory neurons and microglia.

cell types, expression of target genes in MEF2C trios were distinct between cell types (Figures 4C and 4D). In microglia, target genes were enriched in GO terms related to pattern recognition receptor (PRR) signaling, while in neurons they were en-

riched in synaptic transmission in neurons (Figure 4E; Table S7). PRRs consist of several receptor families including Toll-like receptors that are critical for microglial activation.<sup>46</sup> Of the 911 DEGs, we found that 601 participated in a trio in the



**Figure 5. Validation of candidate CREs**

(A) sLDSC results using 16 GWAS traits as indicated with our linked peaks stratified by cell type and group (“All” = all links, “Common” = links identified in both AD and control data, “AD” = links specific to AD, “Control” = links specific to control). Heatmap indicates coefficient Z score from running sLDSC with each set of links combined with the 97 baseline features. Feature-trait combinations with a Z score significantly larger than 0 (one-sided Z test with  $\alpha = 0.05$ , p values corrected within each trait using Benjamini-Hochberg method) are indicated with a numeric value reporting the enrichment score.

(B) Bar plot showing enrichment ( $\pm 95\%$  confidence interval [CI]) of links for previously nominated regulatory regions: active MPRA elements (blue), eQTLs where target gene is same as linked gene (pink), and HiC loops linking region to same target gene (green). MPRA, massively parallel reporter assay; NPC, neural precursor cells; ESC, embryonic stem cells.

(C) Boxplots showing statistically significant ( $p < 0.05$ , ANOVA with Fisher’s LSD) elements representing links tested in luciferase assays. Luciferase elements are denoted by the linked gene for the nominated region.

(D) Boxplots showing comparison of rs12445022 with its corresponding reference element linked to *JPH3* ( $p < 0.05$ , ANOVA with Fisher’s LSD).

(legend continued on next page)



same cell type in which differential expression was observed. Ninety-six DEGs were in AD-specific trios (including *MAPT*, *APOE*, and *BIN1*), while 89 were in control-specific trios (*PADI2/PAD2*, *PDE10A*, and *SNAP25*). The presence of SNAP25 in cerebral spinal fluid is associated with amyloid pathology,<sup>47</sup> and decreased expression in AD brain tissues has been observed,<sup>48</sup> consistent with the decreased expression we find in AD astrocytes and neurons.

Within this set of trios, there was a small subset that were specific to either AD or control groups ( $n = 2,718$ ). Although many of these were specific to a single cell type, 55% were shared across two or more (Figure S4). All cell type-specific trios overlapped H3K27ac peaks from their respective cell types (Table S6). Within microglia trios, NR4A2 (*Nurr1*) was identified most frequently in control-specific trios (Figure 4F). NR4A2 can function as both an activator and repressor and has been shown to repress inflammatory responses in microglia through recruitment of the CoREST complex.<sup>49</sup> Target genes in NR4A2 trios are enriched in neutrophil degranulation ( $OR = 9.01$ ,  $q = 5.3 \times 10^{-6}$ ) and include interleukin genes *IL1A* and *IL1B*, as well as *TGFB1*. Similarly, MAFB was involved in 24% of the AD-specific trios (Figure 4F) where it was linked to the microglial marker gene *CX3CR1* and genes involved in microglial activation (*TLR3*, *CD84*, *HAVCR2*).<sup>50</sup> In healthy microglia, MAFB inhibits inflammatory responses<sup>51</sup> consistent with our finding that target genes in AD-specific trios were enriched for negative regulation of myeloid leukocyte mediated immunity ( $OR = 332$ ,  $q = 0.0004$ ).

Within neuron-specific trios, we identified KLF10 and ZEB1 most frequently in control- and AD-specific trios, respectively (Figure 4F). In neurons, we identified ZEB1 in half of all AD-specific trios with target genes involved in regulating ion channel signaling (*ITPR1*, *CAMK2A*, *CACNB3*, *KCNH3*, *KCNQ5*, *KCNT1*). We found that one ZEB1 target gene encoding a neuronal  $Ca^{2+}$  sensor, *VSNL1*, was downregulated in AD inhibitory neurons consistent with previous studies, and this reduced expression has been correlated with amyloid plaques and neurofibrillary tangles.<sup>52</sup> ZEB1 was never found in control-specific trios. Given the frequency of ZEB1 participation in neuronal AD-specific trios, we performed ZEB1 chromatin immunoprecipitation sequencing (ChIP-seq) in NeuN<sup>+</sup> nuclei isolated from two control donors (1238 and 1242). We found that 33% (55/167) of neuronal ZEB1 trios are bound by ZEB1, and 25 of these are AD-specific trios. The GABA<sub>A</sub> receptor  $\alpha 5$  subunit, encoded by *GABRA5*, is one gene that we find likely to be regulated by ZEB1 in AD (Figure 4G).  $\alpha 5$  GABA<sub>A</sub> receptors are associated with learning and memory, consistent with highest expression of *GABRA5* in hippocampal neurons and association of reduced expression with neurodevelopmental disorders.<sup>53</sup> In our data, *ZEB1* is expressed in both neurons and microglia; however, *GABRA5* is primarily expressed in excitatory neurons (Figure 4G, right). In excitatory neurons, we identified a linked peak correlated with

*GABRA5* expression that was marked with H3K27ac and contained a ZEB1 motif. ChIP-seq data from two of our unaffected donors confirmed ZEB1 binding at this site providing additional evidence to suggest *cis*-regulatory activity of this region for *GABRA5*.

### Genetic variation at candidate CREs

We performed stratified linkage disequilibrium score (sLDSC) regression<sup>54</sup> to determine if our linked peaks were significantly enriched for SNPs associated with complex brain-related traits (Figure 5A; Table S8). Link categories are defined as “AD” or “control” if the links were only identified in the analysis of AD or control samples, respectively. “Common” links were identified in both analyses, and “all” is the union of all linked peaks. Although a peak with multiple links can be duplicated across categories, fewer than a third of peaks with AD-specific links also have control-specific links, emphasizing the specificity of these linked peaks. Within each link category for each cell type, the union of linked peaks was used for this analysis. Cell type was assigned on the basis of the cell type(s) in which the linked peak was identified. Consistent with previous studies,<sup>55,56</sup> linked peaks identified in microglia were significantly enriched for heritability of AD across five different studies<sup>2–6</sup>; however, this was not true for those microglial linked peaks identified in control samples, suggesting that variants in AD-specific linked peaks could have a greater contribution to AD risk. Specificity of microglial linked peaks for AD heritability is also supported by the lack of significant enrichment of these links with risk variants from other brain-related traits,<sup>57–61</sup> or traits where other immune cells play important roles.<sup>62–64</sup> In contrast, linked peaks identified in other cell types were enriched for heritability of brain-related traits including autism spectrum disorder (ASD), bipolar disorder (BD), and schizophrenia (SZ) with AD-specific linked peaks largely excluded from any significant enrichment in these traits. These findings are consistent with previous studies in which candidate CREs identified in excitatory and inhibitory neurons were significantly associated with neuropsychiatric traits.<sup>11</sup> As expected, we did not identify significant enrichments with immune diseases or with other phenotypic traits, such as body mass index (BMI)<sup>65</sup> or height.<sup>66</sup>

### Validation of candidate CREs

We compared the 319,905 links we identified to pre-existing, large-scale functional genomic datasets and identified 67,541 links representing candidate CREs with orthogonal evidence of regulatory activity. This evidence was provided by three data types: (1) massively parallel reporter assay (MPRA),<sup>21,67–69</sup> (2) eQTL studies,<sup>43,70</sup> and (3) HiC<sup>71</sup> datasets. We found significant enrichments of links across each of these datasets despite several MPRA being performed in cancer cell lines (Figure 5B). The MPRA data provided evidence that linked peaks could

(E) Top: normalized expression of *APP* in each cell type. Middle: coverage plot of accessibility in indicated cell types. Bottom: significant control (blue) and common (gray) peak-gene links to *APP* tested in luciferase assays. Arc height represents strength and direction of correlation. Links that contained CREs that increased expression of the luciferase reporter are highlighted in gray.

(F) Boxplots showing all tested luciferase elements representing *APP*-peak links. Elements highlighted in gray are located within the *APP* gene body (\* $p < 0.05$ , ANOVA with Fisher's LSD).

stimulate transcription, but this assay is not capable of identifying the target gene. In contrast, HiC data from NeuN<sup>+</sup> nuclei provided orthogonal validation of a linked peak's target gene, but no evidence of promoting transcriptional activity. We intersected the results from these analyses and found that 1,542 of the 60,473 links that displayed regulatory activity in one or more MPRA also identified the same target gene as the HiC data. In addition, 617 linked peaks overlapped eQTLs and were linked to the same gene providing both evidence of activity and confirming the target gene.

For additional validation, we selected 51 neuronal links for testing in a luciferase reporter assay (Table S9). We performed these assays in the neuroepithelial-derived human embryonic kidney 293 (HEK293 and 293FT) cell lines because of the similar chromatin accessibility landscape to that found in brain tissues.<sup>21</sup> These cell lines are also technically tractable as they are highly transfectable and allow efficient screening of regions of interest. We did not select any AD-specific links for validation, as we are using cell lines from a presumably unaffected individual. Thirteen of these 51 links contained SNPs associated with a brain-related trait (e.g. AD, epilepsy, neurodegeneration), and we tested both alleles of these SNPs (Table S10). Twelve of the elements increased activity of the luciferase reporter including regions linked to *SNCA* ( $\alpha$ -synuclein) and *APP* (amyloid precursor protein) (Figures 5C–5F). Three of these active elements were involved in peak-gene-TF trios (*CCSER1*-MEF2C, *JPH3*-RARB, and *ADAMTS1*-SOX10). ChIP-seq analysis of NeuN<sup>+</sup> nuclei confirmed that MEF2C is bound at the peak linked to *CCSER1*, a gene associated with autism<sup>72</sup> (data not shown). Only one of the 15 variants tested abolished activity, rs12445022, a G/A substitution in a peak linked to *JPH3* ( $p = 0.0003$  by ANOVA with Fisher's least significant difference [LSD]) (Figure 5D). *JPH3* encodes junctophilin-3, important for regulating neuronal excitability.<sup>73</sup> This *JPH3* linked peak was highly correlated ( $r = 0.64$ ,  $q < 2.2 \times 10^{-16}$ ) with *JPH3* expression in both AD and control samples in all cell types except microglia. The linked peak is located 45,503 bp upstream of the *JPH3* TSS and was also linked to *ZCCHC14-DT*, although with a much lower correlation ( $r = 0.36$ ,  $q < 2.2 \times 10^{-16}$ ). Repeat expansions in *JPH3* have been associated with a Huntington disease-like phenotype.<sup>74,75</sup>

Because of its importance in AD pathogenesis, we focused our validation efforts particularly on the *APP* locus (Figure 5E) where we tested 15 elements and identified three that increased expression in the luciferase reporter assay (Figure 5F). *APP* is expressed across all cell types (Figure 5E, top panel), consistent with the high promoter accessibility observed (Figure 5E, middle panels). We also found one element negatively correlated with *APP* expression that significantly reduced reporter activity; however, this assay was not designed to detect repressor activity (as it employs a minimal promoter), thus further experiments would be required to definitively assign a repressive function to this element. Although we have validated several CREs in an unaffected cell line, future work in patient-derived iPSCs would be necessary to evaluate the activity of AD-specific elements identified in this study, with the caveat that this system may still have limitations.

## DISCUSSION

Single cell multiomics has allowed the generation of a rich source of disease- and cell type-specific candidate CREs enriched in variants associated with AD. Our study provides tangible advances by employing snRNA-seq and snATAC-seq in the same cells. Other studies have generated snRNA-seq and snATAC-seq separately and integrated them to identify CREs in AD<sup>19</sup>; however, profiling gene expression and chromatin accessibility simultaneously in the same nuclei allows greater confidence in the correlations linking potential CREs to target genes. As such, we identified five times as many new candidate CREs than previously reported (319,905 links vs 56,552 gene-linked cCREs).<sup>19</sup> To our knowledge to date, only one other study of another human neurodegenerative disease, Parkinson's, used the 10x Genomics Multiomics (ATAC+Gene Expression) technology<sup>76</sup> and identified a similarly large number of peak-gene links (193,732 compared with our 319,905). Our approach is unique in that we identified peak-gene correlations independently in control and AD datasets allowing us to identify 40,831 peak-gene links specific to AD.

We identified many DEGs associated with calcium homeostasis consistent with the calcium hypothesis of AD which postulates that a synergistic relationship between A $\beta$  accumulation and Ca<sup>2+</sup> levels promotes neurodegeneration.<sup>77</sup> In AD neurons, we found decreased expression of ryanodine receptor 3 (*RYR3*) and inositol 1,4,5-trisphosphate receptor type 2 (*ITPR2*) that both release internal stores of Ca<sup>2+</sup> from the endoplasmic reticulum.<sup>78</sup> We also measured decreased expression of the Ca<sup>2+</sup> sensors calmodulin (*CALM1*, *CALM2*, *CALM3*) and VILIP-1 (*VSNL1*), the latter which is associated with neuropathologic lesions.<sup>52,79</sup> In contrast, two genes encoding calcium channel subunits (*CACNA1C* and *CACNA1B*) were upregulated in AD neurons. In addition, astrocytes also demonstrated decreased expression of calneuron 1 (*CALN1*, a Ca<sup>2+</sup> sensor similar to calmodulin), glutamate receptor 2 subunit (*GRIA2*, limits Ca<sup>2+</sup> permeability of AMPA receptors), and glutamate receptor NMDA 2C (*GRIN2C*, a subunit of the NMDA receptors). Both AD neurons and astrocytes showed decreased expression of the glutamate transporter GLT-1 (*SLC1A2*). We identified AD-specific links for all these genes except *GRIN2C*. Altered expression of these calcium-associated proteins is likely to exhibit complex and cell type-specific effects making the resulting network effect on excitability uncertain. However, one possibility is that this altered expression could lead to increased sensitivity of neurons to glutamate and thus neurotoxicity.<sup>78,80</sup> Further study of the candidate regulatory elements we identified for these genes would improve our understanding of how these genes become dysregulated in AD and the emergent resulting effects.

Our study provides two main advances in our understanding of altered gene regulation in AD. First, by leveraging the AD- and control-specific links identified here we constructed peak-gene-TF trios to determine which TFs were particularly involved in regulating AD-specific transcriptional programs. MAFB and ZEB1 were found to be enriched in AD-specific trios in microglia and neurons, respectively. MAFB has previously been implicated in exercise-associated responses in the peripheral immune system in AD<sup>81</sup> and in regulation of the receptor VISTA in microglia,

which is upregulated in AD.<sup>82</sup> In this study, we identify a previously unknown role for ZEB1 in AD-specific transcriptional regulation in neurons. Previously, ZEB1 was shown to play a critical role in epithelial-mesenchymal transition in neural crest migration and glioblastoma<sup>83,84</sup> and further investigation is necessary to fully understand its role in AD. Second, we demonstrated enhancer-like activity for 12 candidate CREs linked to neurodegeneration-associated genes *APP*, *SNCA*, *PHF24*, and *ADAMTS1*. Amyloid precursor protein (*APP*) is the precursor to the AD hallmark pathology A $\beta$ , and, while characteristic of Lewy body diseases,  $\alpha$ -synuclein (*SNCA*) aggregates are highly prevalent in AD postmortem brains as well.<sup>85</sup> *PHF24* is a modulator of GABA<sub>B</sub> receptor activity<sup>86</sup> and was recently identified in a study of AD resilience genes.<sup>87</sup> *ADAMTS1* has been implicated in AD both biochemically<sup>88</sup> and genetically.<sup>6</sup> Our study lays the groundwork for additional functional validation in future studies to confirm these genes as targets of these CREs. Understanding how these genes are regulated and by which TFs could provide new therapeutic targets. In fact, a recent study<sup>89</sup> identified TFs contributing to disruption of gene regulatory networks in AD, demonstrated their ability to predict AD cognitive phenotypes, and used them to prioritize candidate drugs that could be repurposed for AD.

In summary, our study provides important new insights into the contribution of CREs to AD including the roles of TFs ZEB1 and MAFB in neurons and microglia. These findings could provide additional insights for interpreting SNPs associated with AD risk should they disrupt binding motifs for these TFs. Furthermore, these TFs could be therapeutic targets for manipulating aberrant gene regulation in AD. Our study lays the groundwork for future research to expand on the candidate- and literature-based validation approaches taken here. High-throughput CRISPRi screens are well suited to test the necessity and sufficiency of regulatory elements for linked gene expression. Future validation efforts will greatly contribute to advancing our understanding of the effects of non-coding variation on risk for AD.

### Limitations of the study

One limitation of this study is that snATAC-seq data can contain spurious signals, as well as bias from transcribed genes. This limitation underscores the importance of evaluation via orthogonal methods, which we have provided using both published and newly generated data. A second limitation is that our sample size is small. This can be addressed in future studies by increasing sample size; however, the shared signals we observed with larger AD snRNA-seq studies emphasizes the representative nature of our sample set, and that our total number of cells per biological sample is adequate. Finally, as with any study from postmortem tissue, we are measuring by definition the material that remains in a neurodegenerative disease, which can confound interpretation. For this reason, we chose to evaluate DLPCF, which is preserved later into the disease course of AD than tissues affected earlier such as entorhinal cortex and hippocampus. Although this study is focused on identifying CREs, there are other non-coding regulatory mechanisms that could alter gene expression in AD including microRNAs (miRNAs),<sup>90</sup> long non-coding RNAs (lncRNAs),<sup>91</sup> transposable elements,<sup>92</sup> etc., that are not assessed.

### STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **RESOURCE AVAILABILITY**
  - Lead contact
  - Materials availability
  - Data and code availability
- **EXPERIMENTAL MODEL AND SUBJECT DETAILS**
  - Cell lines
  - Human brain tissues
- **METHOD DETAILS**
  - Nuclei isolation from human brain tissues
  - Single nucleus multiomics
  - Sample demultiplexing
  - Joint snRNA-seq and snATAC-seq workflow
  - WNN analysis of snRNA-seq and snATAC-seq
  - Differential expression
  - Annotation of cell subpopulations
  - Gene set enrichment
  - Feature linkage analysis
  - Permutation testing
  - Peak-gene-TF trios
  - Partitioned heritability analysis
  - *APOE* genotyping
  - Comparisons to external data sources
  - Plasmids
  - Transfection
  - Luciferase assays
  - Chromatin preparation for sorted nuclei
  - ChIP-seq protocol
  - ChIP-seq analysis
- **QUANTIFICATION AND STATISTICAL ANALYSIS**

### SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.xgen.2023.100263>.

### ACKNOWLEDGMENTS

We thank Paige I. Hall for assisting L.W. and J.N.B. and Rebecca M. Hauser for *APOE* genotyping. This study was supported in part by NIH grants 5R01MH110472 and 3R01MH110472-03S1 awarded to R.M.M. and NIH grant 5R00AG068271 awarded to J.N.C. Additional funding was generously provided by donors to the HudsonAlpha Foundation Memory and Mobility Program and to the Leo Fund. We especially thank the brain donors and their families without whom this research would not have been possible.

### AUTHOR CONTRIBUTIONS

Conceptualization, L.F.R., J.N.C., and J.M.L.; formal analysis, A.G.A., L.F.R., and I.R.-N.; investigation, J.N.B., L.W., S.R., L.F.R., J.M.L., and B.B.R.; resources, W.E.B., B.G.B., and S.J.W.; data curation, A.G.A.; writing – original draft, L.F.R., A.G.A., and B.B.R.; writing – review & editing, J.N.C., R.M.M., and J.M.L.; supervision, L.F.R., J.N.C., and R.M.M.; funding, J.N.C. and R.M.M.

**DECLARATION OF INTERESTS**

The authors declare no competing interests.

**INCLUSION AND DIVERSITY**

We support inclusive, diverse, and equitable conduct of research.

Received: October 5, 2022

Revised: December 6, 2022

Accepted: January 12, 2023

Published: February 2, 2023

**REFERENCES**

- Neuner, S.M., Tcw, J., and Goate, A.M. (2020). Genetic architecture of Alzheimer's disease. *Neurobiol. Dis.* *143*, 104976.
- 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.
- Jansen, I.E., Savage, J.E., Watanabe, K., Bryois, J., Williams, D.M., Steinberg, S., Sealock, J., Karlsson, I.K., Hägg, S., Athanasiu, L., et al. (2019). Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nat. Genet.* *9*, 63.
- Schwartzentruber, J., Cooper, S., Liu, J.Z., Barrio-Hernandez, I., Bello, E., Kumasaka, N., Young, A.M.H., Franklin, R.J.M., Johnson, T., Estrada, K., et al. (2021). Genome-wide meta-analysis, fine-mapping and integrative prioritization implicate new Alzheimer's disease risk genes. *Nat. Genet.* *53*, 392–402.
- 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.* *53*, 1276–1282.
- Bellenguez, C., Küçükali, F., Jansen, I.E., Kleindam, 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.
- Uffelmann, E., Huang, Q.Q., Munung, N.S., de Vries, J., Okada, Y., Martin, A.R., Martin, H.C., Lappalainen, T., and Posthuma, D. (2021). Genome-wide association studies. *Nat. Rev. Methods Primers* *1*, 59.
- Zhang, F., and Lupski, J.R. (2015). Non-coding genetic variants in human disease. *Hum. Mol. Genet.* *24*, R102–R110.
- Khani, M., Gibbons, E., Bras, J., and Guerreiro, R. (2022). Challenge accepted: uncovering the role of rare genetic variants in Alzheimer's disease. *Mol. Neurodegener.* *17*, 3.
- 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.
- Nott, A., Holtman, I.R., Coufal, N.G., Schlachetzki, J.C.M., Yu, M., Hu, R., Han, C.Z., Pena, M., Xiao, J., Wu, Y., et al. (2019). Brain cell type-specific enhancer-promoter interactome maps and disease risk association. *Science* *366*, 1134–1139.
- Mathys, H., Davila-Velderrain, J., Peng, Z., Gao, F., Mohammadi, S., Young, J.Z., Menon, M., He, L., Abdurrob, F., Jiang, X., et al. (2019). Single-cell transcriptomic analysis of Alzheimer's disease. *Nature* *570*, 332–337.
- Grubman, A., Chew, G., Ouyang, J.F., Sun, G., Choo, X.Y., McLean, C., Simmons, R.K., Buckberry, S., Vargas-Landin, D.B., Poppe, D., et al. (2019). A single-cell atlas of entorhinal cortex from individuals with Alzheimer's disease reveals cell-type-specific gene expression regulation. *Nat. Neurosci.* *22*, 2087–2097.
- Sadick, J.S., O'Dea, M.R., Hasel, P., Dykstra, T., Faustin, A., and Lidde-low, S.A. (2022). Astrocytes and oligodendrocytes undergo subtype-specific transcriptional changes in Alzheimer's disease. *Neuron* *110*, 1788–1805.e10.
- Leng, K., Li, E., Eser, R., Piergies, A., Sit, R., Tan, M., Neff, N., Li, S.H., Rodriguez, R.D., Suemoto, C.K., et al. (2021). Molecular characterization of selectively vulnerable neurons in Alzheimer's disease. *Nat. Neurosci.* *24*, 276–287.
- Zhou, Y., Song, W.M., Andhey, P.S., Swain, A., Levy, T., Miller, K.R., Poliani, P.L., Cominelli, M., Grover, S., Gilfillan, S., et al. (2020). Human and mouse single-nucleus transcriptomics reveal TREM2-dependent and TREM2-independent cellular responses in Alzheimer's disease. *Nat. Med.* *26*, 131–142.
- Alsema, A.M., Jiang, Q., Kracht, L., Gerrits, E., Dubbelaar, M.L., Miedema, A., Brouwer, N., Hol, E.M., Middeldorp, J., van Dijk, R., et al. (2020). Profiling microglia from Alzheimer's disease donors and non-demented elderly in acute human postmortem cortical tissue. *Front. Mol. Neurosci.* *13*, 134.
- Lau, S.-F., Cao, H., Fu, A.K.Y., and Ip, N.Y. (2020). Single-nucleus transcriptome analysis reveals dysregulation of angiogenic endothelial cells and neuroprotective glia in Alzheimer's disease. *Proc. Natl. Acad. Sci. USA* *117*, 25800–25809.
- Morabito, S., Miyoshi, E., Michael, N., Shahin, S., Martini, A.C., Head, E., Silva, J., Leavy, K., Perez-Rosendahl, M., and Swarup, V. (2021). Single-nucleus chromatin accessibility and transcriptomic characterization of Alzheimer's disease. *Nat. Genet.* *53*, 1143–1155.
- Kartha, V.K., Duarte, F.M., Hu, Y., Ma, S., Chew, J.G., Lareau, C.A., Earl, A., Burkett, Z.D., Kohlway, A.S., Lebofsky, R., et al. (2022). Functional inference of gene regulation using single-cell multi-omics. *Cell Genom.* *2*, 100166.
- 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.
- Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. *Cell* *184*, 3573–3587.e29.
- Stuart, T., Srivastava, A., Madad, S., Lareau, C.A., and Satija, R. (2021). Single-cell chromatin state analysis with Signac. *Nat. Methods* *18*, 1333–1341.
- Belonwu, S.A., Li, Y., Bunis, D., Rao, A.A., Solsberg, C.W., Tang, A., Fragiadakis, G.K., Dubal, D.B., Oskotsky, T., and Sirota, M. (2022). Sex-stratified single-cell RNA-seq analysis identifies sex-specific and cell type-specific transcriptional responses in Alzheimer's disease across two brain regions. *Mol. Neurobiol.* *59*, 276–293.
- Boni, C., Laudanna, C., and Sorio, C. (2022). A comprehensive Review of receptor-type tyrosine-protein phosphatase gamma (PTPRG) role in health and non-neoplastic disease. *Biomolecules* *12*, 84.
- Herold, C., Hooli, B.V., Mullin, K., Liu, T., Roehr, J.T., Mattheisen, M., Parrado, A.R., Bertram, L., Lange, C., and Tanzi, R.E. (2016). Family-based association analyses of imputed genotypes reveal genome-wide significant association of Alzheimer's disease with OSBPL6, PTPRG, and PDCL3. *Mol. Psychiatry* *21*, 1608–1612.
- Ceprian, M., and Fulton, D. (2019). Glial cell AMPA receptors in nervous system health, injury and disease. *Int. J. Mol. Sci.* *20*, 2450.
- Wang, X.-L., and Li, L. (2021). Cell type-specific potential pathogenic genes and functional pathways in Alzheimer's Disease. *BMC Neurol.* *21*, 381.
- Connor, S.A., Ammendrup-Johnsen, I., Chan, A.W., Kishimoto, Y., Murayama, C., Kurihara, N., Tada, A., Ge, Y., Lu, H., Yan, R., et al. (2016). Altered cortical dynamics and cognitive function upon haploinsufficiency

- of the autism-linked excitatory synaptic suppressor MDGA2. *Neuron* 97, 1052–1068.
30. Khrimian, L., Obri, A., Ramos-Brossier, M., Rousseaud, A., Moriceau, S., Nicot, A.-S., Mera, P., Kosmidis, S., Karnavas, T., Saudou, F., et al. (2017). Gpr158 mediates osteocalcin's regulation of cognition. *J. Exp. Med.* 214, 2859–2873.
  31. Ghose, U., Sproviero, W., Winchester, L., Fernandes, M., Newby, D., Ulm, B.S., Shi, L., Liu, Q., Adams, C., Albukhari, A., et al. (2022). Genome wide association neural networks (GWANN) identify novel genes linked to family history of Alzheimer's disease in the UK Biobank. Preprint at medRxiv. <https://doi.org/10.1101/2022.06.10.22276251>.
  32. DeTomaso, D., and Yosef, N. (2021). Hotspot identifies informative gene modules across modalities of single-cell genomics. *Cell Syst.* 12, 446–456.e9.
  33. 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.
  34. 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.
  35. ENCODE Project Consortium; Moore, J.E., Purcaro, M.J., Pratt, H.E., Epstein, C.B., Shores, N., Adrian, J., Kawli, T., Davis, C.A., Dobin, A., et al. (2020). Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* 583, 699–710.
  36. Kozlenkov, A., Li, J., Apontes, P., Hurd, Y.L., Byne, W.M., Koonin, E.V., Wegner, M., Mukamel, E.A., and Dracheva, S. (2018). A unique role for DNA (hydroxy)methylation in epigenetic regulation of human inhibitory neurons. *Sci. Adv.* 4, eaau6190.
  37. Starks, R.R., Biswas, A., Jain, A., and Tuteja, G. (2019). Combined analysis of dissimilar promoter accessibility and gene expression profiles identifies tissue-specific genes and actively repressed networks. *Epigenet. Chromatin* 12, 16.
  38. Pastor, W.A., Stroud, H., Nee, K., Liu, W., Pezic, D., Manakov, S., Lee, S.A., Moissiard, G., Zamudio, N., Bourc'his, D., et al. (2014). MORC1 represses transposable elements in the mouse male germline. *Nat. Commun.* 5, 5795.
  39. Ampuja, M., Rantapero, T., Rodriguez-Martinez, A., Palmroth, M., Alarmo, E.L., Nykter, M., and Kallioniemi, A. (2017). Integrated RNA-seq and DNase-seq analyses identify phenotype-specific BMP4 signaling in breast cancer. *BMC Genom.* 18, 68.
  40. Merrill, C.B., Montgomery, A.B., Pabon, M.A., Shabalin, A.A., Rodan, A.R., and Rothenfluh, A. (2022). Harnessing changes in open chromatin determined by ATAC-seq to generate insulin-responsive reporter constructs. *BMC Genom.* 23, 399.
  41. Lara-Astiaso, D., Weiner, A., Lorenzo-Vivas, E., Zaretzky, I., Jaitin, D.A., David, E., Keren-Shaul, H., Mildner, A., Winter, D., Jung, S., et al. (2014). Immunogenetics. Chromatin state dynamics during blood formation. *Science* 345, 943–949.
  42. Sudwants, A., Ramesha, S., Gao, T., Ponnusamy, M., Wang, S., Hansen, M., Kozlova, A., Bitarafan, S., Kumar, P., Beaulieu-Abdelahad, D., et al. (2022). BIN1 is a key regulator of proinflammatory and neurodegeneration-related activation in microglia. *Mol. Neurodegener.* 17, 33.
  43. Bryois, J., Calini, D., Macnair, W., Foo, L., Ulrich, E., Ortmann, W., Iglesias, V.A., Selvaraj, S., Nutma, E., Marzin, M., et al. (2022). Cell-type-specific cis-eQTLs in eight human brain cell types identify novel risk genes for psychiatric and neurological disorders. *Nat. Neurosci.* 25, 1104–1112.
  44. Chen, J.A., Chen, Z., Won, H., Huang, A.Y., Lowe, J.K., Wojta, K., Yokoyama, J.S., Bensimon, G., Leigh, P.N., Payan, C., et al. (2018). Joint genome-wide association study of progressive supranuclear palsy identifies novel susceptibility loci and genetic correlation to neurodegenerative diseases. *Mol. Neurodegener.* 13, 41.
  45. Jiang, Y., Harigaya, Y., Zhang, Z., Zhang, H., Zang, C., and Zhang, N.R. (2022). Nonparametric interrogation of transcriptional regulation in single-cell RNA and chromatin accessibility multiomic data. Preprint at bioRxiv. <https://doi.org/10.1101/2021.09.22.461437>.
  46. Kigerl, K.A., de Rivero Vaccari, J.P., Dietrich, W.D., Popovich, P.G., and Keane, R.W. (2014). Pattern recognition receptors and central nervous system repair. *Exp. Neurol.* 258, 5–16.
  47. Halbgebauer, S., Steinacker, P., Hengge, S., Oeckl, P., Abu Rumeileh, S., Anderl-Straub, S., Lombardi, J., Von Arnim, C.A.F., Giese, A., Ludolph, A.C., et al. (2022). CSF levels of SNAP-25 are increased early in Creutzfeldt-Jakob and Alzheimer's disease. *J. Neurol. Neurosurg. Psychiatry* 93, 1059–1065.
  48. Furuya, T.K., Silva, P.N.O., Payão, S.L.M., Bertolucci, P.H.F., Rasmussen, L.T., De Labio, R.W., Braga, I.L.S., Chen, E.S., Turecki, G., Mechawar, N., et al. (2012). Analysis of SNAP25 mRNA expression and promoter DNA methylation in brain areas of Alzheimer's Disease patients. *Neuroscience* 220, 41–46.
  49. Saijo, K., Winner, B., Carson, C.T., Collier, J.G., Boyer, L., Rosenfeld, M.G., Gage, F.H., and Glass, C.K. (2009). A Nurrl/CoREST transrepression pathway attenuates neurotoxic inflammation in activated microglia and astrocytes. *Cell* 137, 47–59.
  50. Patir, A., Shih, B., McColl, B.W., and Freeman, T.C. (2019). A core transcriptional signature of human microglia: derivation and utility in describing region-dependent alterations associated with Alzheimer's disease. *Glia* 67, 1240–1253.
  51. Holtman, I.R., Skola, D., and Glass, C.K. (2017). Transcriptional control of microglia phenotypes in health and disease. *J. Clin. Invest.* 127, 3220–3229.
  52. Braunewell, K.H. (2012). The visinin-like proteins VILIP-1 and VILIP-3 in Alzheimer's disease—old wine in new bottles. *Front. Mol. Neurosci.* 5, 20.
  53. Jacob, T.C. (2019). Neurobiology and therapeutic potential of  $\alpha$ 5-GABA type A receptors. *Front. Mol. Neurosci.* 12, 179.
  54. Finucane, H.K., Bulik-Sullivan, B., Gusev, A., Trynka, G., Reshef, Y., Loh, P.R., Anttila, V., Xu, H., Zang, C., Farh, K., et al. (2015). Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* 47, 1228–1235.
  55. Tansey, K.E., Cameron, D., and Hill, M.J. (2018). Genetic risk for Alzheimer's disease is concentrated in specific macrophage and microglial transcriptional networks. *Genome Med.* 10, 14.
  56. Novikova, G., Kapoor, M., Tcw, J., Abud, E.M., Efthymiou, A.G., Chen, S.X., Cheng, H., Fullard, J.F., Bendli, 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.
  57. International Multiple Sclerosis Genetics Consortium; Wellcome Trust Case Control Consortium 2; Sawcer, S., Hellenthal, G., Pirinen, M., Spencer, C.C.A., Patsopoulos, N.A., Moutsianas, L., Dilthey, A., Su, Z., et al. (2011). Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature* 476, 214–219.
  58. van Rheenen, W., van der Spek, R.A.A., Bakker, M.K., van Vugt, J.J.F.A., Hop, P.J., Zwamborn, R.A.J., de Klein, N., Westra, H.-J., Bakker, O.B., Deelen, P., et al. (2021). Common and rare variant association analyses in amyotrophic lateral sclerosis identify 15 risk loci with distinct genetic architectures and neuron-specific biology. *Nat. Genet.* 53, 1636–1648.
  59. Grove, J., Ripke, S., Als, T.D., Mattheisen, M., Walters, R.K., Won, H., Pallesen, J., Agerbo, E., Andreassen, O.A., Anney, R., et al. (2019). Identification of common genetic risk variants for autism spectrum disorder. *Nat. Genet.* 51, 431–444.
  60. Wray, N.R., Ripke, S., Mattheisen, M., Trzaskowski, M., Byrne, E.M., Abdellaoui, A., Adams, M.J., Agerbo, E., Air, T.M., Andlauer, T.M.F., et al. (2018). Genome-wide association analyses identify 44 risk variants and

- refine the genetic architecture of major depression. *Nat. Genet.* **50**, 668–681.
61. Pardiñas, A.F., Holmans, P., Pocklington, A.J., Escott-Price, V., Ripke, S., Carrera, N., Legge, S.E., Bishop, S., Cameron, D., Hamshere, M.L., et al. (2018). Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat. Genet.* **50**, 381–389.
  62. Liu, J.Z., van Sommeren, S., Huang, H., Ng, S.C., Alberts, R., Takahashi, A., Ripke, S., Lee, J.C., Jostins, L., Shah, T., et al. (2015). Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat. Genet.* **47**, 979–986.
  63. Okada, Y., Wu, D., Trynka, G., Raj, T., Terao, C., Ikari, K., Kochi, Y., Ohmura, K., Suzuki, A., Yoshida, S., et al. (2014). Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* **506**, 376–381.
  64. Julià, A., López-Longo, F.J., Pérez Venegas, J.J., Bonàs-Guarch, S., Olivé, À., Andreu, J.L., Aguirre-Zamorano, M.Á., Vela, P., Nolla, J.M., de la Fuente, J.L.M., et al. (2018). Genome-wide association study meta-analysis identifies five new loci for systemic lupus erythematosus. *Arthritis Res. Ther.* **20**, 100.
  65. Locke, A.E., Kahali, B., Berndt, S.I., Justice, A.E., Pers, T.H., Day, F.R., Powell, C., Vedantam, S., Buchkovich, M.L., Yang, J., et al. (2015). Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197–206.
  66. Wood, A.R., Esko, T., Yang, J., Vedantam, S., Pers, T.H., Gustafsson, S., Chu, A.Y., Estrada, K., Luan, J., Kutalik, Z., et al. (2014). Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat. Genet.* **46**, 1173–1186.
  67. Weiss, C.V., Harshman, L., Inoue, F., Fraser, H.B., Petrov, D.A., Ahituv, N., and Gokhman, D. (2021). The cis-regulatory effects of modern human-specific variants. *Elife* **10**, e63713.
  68. Uebbing, S., Gockley, J., Reilly, S.K., Kocher, A.A., Geller, E., Gandotra, N., Scharfe, C., Cotney, J., and Noonan, J.P. (2021). Massively parallel discovery of human-specific substitutions that alter enhancer activity. *Proc. Natl. Acad. Sci. USA* **118**, e2007049118.
  69. van Arensbergen, J., Pagie, L., FitzPatrick, V.D., de Haas, M., Baltissen, M.P., Comoglio, F., van der Weide, R.H., Teunissen, H., Vösa, U., Franke, L., et al. (2019). High-throughput identification of human SNPs affecting regulatory element activity. *Nat. Genet.* **51**, 1160–1169.
  70. GTEx Consortium (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330.
  71. Hu, B., Won, H., Mah, W., Park, R.B., Kassim, B., Spiess, K., Kozlenkov, A., Crowley, C.A., Pochareddy, S., Li, Y., et al.; PsychENCODE Consortium (2021). Neuronal and glial 3D chromatin architecture informs the cellular etiology of brain disorders. *Nat. Commun.* **12**, 3968.
  72. Ruzzo, E.K., Pérez-Cano, L., Jung, J.-Y., Wang, L.K., Kashaf-Haghighi, D., Hartl, C., Singh, C., Xu, J., Hoekstra, J.N., Leventhal, O., et al. (2019). Inherited and de novo genetic risk for autism impacts shared networks. *Cell* **178**, 850–866.e26.
  73. Garbino, A., van Oort, R.J., Dixit, S.S., Landstrom, A.P., Ackerman, M.J., and Wehrens, X.H.T. (2009). Molecular evolution of the junctophilin gene family. *Physiol. Genomics* **37**, 175–186.
  74. Bourinaris, T., Athanasiou, A., Efthymiou, S., Wiethoff, S., Salpietro, V., and Houlden, H. (2021). Allelic and phenotypic heterogeneity in Junctophilin-3 related neurodevelopmental and movement disorders. *Eur. J. Hum. Genet.* **29**, 1027–1031.
  75. Schneider, S.A., Marshall, K.E., Xiao, J., and LeDoux, M.S. (2012). JPH3 repeat expansions cause a progressive akinetic-rigid syndrome with severe dementia and putaminal rim in a five-generation african-American family. *Neurogenetics* **13**, 133–140.
  76. Adams, L., Song, M.K., Tanaka, Y., and Kim, Y.-S. (2022). Single-nuclei paired multiomic analysis of young, aged, and Parkinson’s disease human midbrain reveals age- and disease-associated glial changes and their contribution to Parkinson’s disease. Preprint at medRxiv. <https://doi.org/10.1101/2022.01.18.22269350>.
  77. Alzheimer’s Association Calcium Hypothesis Workgroup (2017). Calcium Hypothesis of Alzheimer’s disease and brain aging: a framework for integrating new evidence into a comprehensive theory of pathogenesis. *Alzheimers Dement.* **13**, 178–182.e17.
  78. Cascella, R., and Cecchi, C. (2021). Calcium dyshomeostasis in Alzheimer’s disease pathogenesis. *Int. J. Mol. Sci.* **22**, 4914.
  79. Schnurra, I., Bernstein, H.-G., Riederer, P., and Braunevel, K.-H. (2001). The neuronal calcium sensor protein VILIP-1 is associated with amyloid plaques and extracellular tangles in Alzheimer’s disease and promotes cell death and tau phosphorylation in vitro: a link between calcium sensors and Alzheimer’s disease? *Neurobiol. Dis.* **8**, 900–909.
  80. Burgoyne, R.D., Helassa, N., McCue, H.V., and Haynes, L.P. (2019). Calcium sensors in neuronal function and dysfunction. *Cold Spring Harb. Perspect. Biol.* **11**, a035154.
  81. Chen, Y., Sun, Y., Luo, Z., Chen, X., Wang, Y., Qi, B., Lin, J., Lin, W.-W., Sun, C., Zhou, Y., et al. (2022). Exercise modifies the transcriptional regulatory features of monocytes in Alzheimer’s patients: a multi-omics integration analysis based on single cell technology. *Front. Aging Neurosci.* **14**, 881488.
  82. Borggrewe, M., Grit, C., Den Dunnen, W.F.A., Burm, S.M., Bajramovic, J.J., Noelle, R.J., Eggen, B.J.L., and Laman, J.D. (2018). VISTA expression by microglia decreases during inflammation and is differentially regulated in CNS diseases. *Glia* **66**, 2645–2658.
  83. Powell, D.R., Blasky, A.J., Britt, S.G., and Artinger, K.B. (2013). Riding the crest of the wave: parallels between the neural crest and cancer in epithelial-to-mesenchymal transition and migration. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **5**, 511–522.
  84. Stemmler, M.P., Eccles, R.L., Brabletz, S., and Brabletz, T. (2019). Non-redundant functions of EMT transcription factors. *Nat. Cell Biol.* **21**, 102–112.
  85. Twohig, D., and Nielsen, H.M. (2019).  $\alpha$ -synuclein in the pathophysiology of Alzheimer’s disease. *Mol. Neurodegener.* **14**, 23.
  86. Numakura, Y., Uemura, R., Tanaka, M., Izawa, T., Yamate, J., Kuramoto, T., Kaneko, T., Mashimo, T., Yamamoto, T., Serikawa, T., et al. (2021). PHF24 is expressed in the inhibitory interneurons in rats. *Exp. Anim.* **70**, 137–143.
  87. Telpoukhovskaia, M.A., Hadad, N., Gurdon, B., Dai, M., Ouellette, A.R., Neuner, S.M., Dunn, A.R., Hansen, S., Wu, Y., Dumitrescu, L., et al. (2022). Conserved cell-type specific signature of resilience to Alzheimer’s disease nominates role for excitatory cortical neurons. Preprint at bioRxiv. <https://doi.org/10.1101/2022.04.12.487877>.
  88. Gurses, M.S., Ural, M.N., Gulec, M.A., Akyol, O., and Akyol, S. (2016). Pathophysiological function of ADAMTS enzymes on molecular mechanism of Alzheimer’s disease. *Aging Dis.* **7**, 479–490.
  89. Gupta, C., Xu, J., Jin, T., Khullar, S., Liu, X., Alatar, S., Cheng, F., and Wang, D. (2022). Single-cell network biology characterizes cell type gene regulation for drug repurposing and phenotype prediction in Alzheimer’s disease. *PLoS Comput. Biol.* **18**, e1010287.
  90. O’Brien, J., Hayder, H., Zayed, Y., and Peng, C. (2018). Overview of MicroRNA biogenesis, mechanisms of actions, and circulation. *Front. Endocrinol.* **9**, 402.
  91. Statello, L., Guo, C.-J., Chen, L.-L., and Huarte, M. (2021). Gene regulation by long non-coding RNAs and its biological functions. *Nat. Rev. Mol. Cell Biol.* **22**, 96–118.
  92. Fueyo, R., Judd, J., Feschotte, C., and Wysocka, J. (2022). Roles of transposable elements in the regulation of mammalian transcription. *Nat. Rev. Mol. Cell Biol.* **23**, 481–497.
  93. Huber, W., Carey, V.J., Gentleman, R., Anders, S., Carlson, M., Carvalho, B.S., Bravo, H.C., Davis, S., Gatto, L., Girke, T., et al. (2015). Orchestrating high-throughput genomic analysis with Bioconductor. *Nat. Methods* **12**, 115–121.

94. Germain, P.-L., Lun, A., Meixide, C.G., Macnair, W., and Robinson, M.D. (2022). Doublet identification in single-cell sequencing data using scDbtFinder. *F1000Res.* *10*, 979.
95. Hafemeister, C., and Satija, R. (2019). Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* *20*, 296.
96. Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baiglaenko, Y., Brenner, M., Loh, P.R., and Raychaudhuri, S. (2019). Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* *16*, 1289–1296.
97. Finak, G., McDavid, A., Yajima, M., Deng, J., Gersuk, V., Shalek, A.K., Slichter, C.K., Miller, H.W., McElrath, M.J., Pric, M., et al. (2015). MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol.* *16*, 278.
98. Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G.V., Clark, N.R., and Ma'ayan, A. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinf.* *14*, 128.
99. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based analysis of ChIP-seq (MACS). *Genome Biol.* *9*, R137.
100. Gu, Z., Eils, R., and Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* *32*, 2847–2849.
101. Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M.T., and Carey, V.J. (2013). Software for computing and annotating genomic ranges. *PLoS Comput. Biol.* *9*, e1003118.
102. Lawrence, M., Gentleman, R., and Carey, V. (2009). rtracklayer: an R package for interfacing with genome browsers. *Bioinformatics* *25*, 1841–1842.
103. Yu, G., Wang, L.-G., and He, Q.-Y. (2015). ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* *31*, 2382–2383.
104. Huang, X., and Huang, Y. (2021). Cellsnr-lite: an efficient tool for genotyping single cells. *Bioinformatics* *37*, 4569–4571.
105. Huang, Y., McCarthy, D.J., and Stegle, O. (2019). Vireo: bayesian demultiplexing of pooled single-cell RNA-seq data without genotype reference. *Genome Biol.* *20*, 273.
106. Zhu, Y., Wang, L., Yin, Y., and Yang, E. (2017). Systematic analysis of gene expression patterns associated with postmortem interval in human tissues. *Sci. Rep.* *7*, 5435.
107. Bakken, T.E., Jorstad, N.L., Hu, Q., Lake, B.B., Tian, W., Kalmbach, B.E., Crow, M., Hodge, R.D., Krienen, F.M., Sorensen, S.A., et al. (2021). Comparative cellular analysis of motor cortex in human, marmoset and mouse. *Nature* *598*, 111–119.
108. Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A., et al. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* *44*, W90–W97.
109. Xie, Z., Bailey, A., Kuleshov, M.V., Clarke, D.J.B., Evangelista, J.E., Jenkins, S.L., Lachmann, A., Wojciechowski, M.L., Kropiwnicki, E., Jagodnik, K.M., et al. (2021). Gene set knowledge discovery with enrichr. *Curr. Protoc.* *1*, e90.
110. Castro-Mondragon, J.A., Riudavets-Puig, R., Rauluseviciute, I., Lemma, R.B., Turchi, L., Blanc-Mathieu, R., Lucas, J., Boddie, P., Khan, A., Mansalva Pérez, N., et al. (2022). JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* *50*, D165–D173.
111. Bulik-Sullivan, B.K., Loh, P.R., Finucane, H.K., Ripke, S., Yang, J., Schizophrenia Working Group of the Psychiatric Genomics Consortium; Patterson, N., Daly, M.J., Price, A.L., and Neale, B.M. (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* *47*, 291–295.
112. Bipolar Disorder and Schizophrenia Working Group of the Psychiatric Genomics Consortium Electronic address douglasruderfer@vanderbilt.edu; Bipolar Disorder and Schizophrenia Working Group of the Psychiatric Genomics Consortium (2018). Genomic dissection of bipolar disorder and schizophrenia, including 28 subphenotypes. *Cell* *173*, 1705–1715.e16.
113. Landt, S.G., Marinov, G.K., Kundaje, A., Kheradpour, P., Pauli, F., Batzoglou, S., Bernstein, B.E., Bickel, P., Brown, J.B., Cayting, P., et al. (2012). ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res.* *22*, 1813–1831.
114. Jiang, Y., Matevosian, A., Huang, H.-S., Straubhaar, J., and Akbarian, S. (2008). Isolation of neuronal chromatin from brain tissue. *BMC Neurosci.* *9*, 42.
115. Savic, D., Gertz, J., Jain, P., Cooper, G.M., and Myers, R.M. (2013). Mapping genome-wide transcription factor binding sites in frozen tissues. *Epigenet. Chromatin* *6*, 30.
116. 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.
117. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* *17*, 10–12.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Antibodies</b>		
Rabbit anti-ZEB1 antibody	Bethyl	Cat#: A301-921A; RRID: AB_1524109
Rabbit polyclonal MEF2C antibody	proteintech	Cat#: 18290-1-AP; RRID: AB_2142849
Dynabeads™ M-280 Sheep Anti-Rabbit IgG	ThermoFisher Scientific	Cat#: 11203D; RRID: AB_2783009
Anti-NeuN Antibody, clone A60, Alexa Fluor®488 conjugated	Millipore Sigma	Cat#: MAB377X; RRID: AB_2149209
Recombinant Anti-Olig2 antibody [EPR2673]	Abcam	Cat#: ab109186; RRID: AB_10861310
<b>Biological samples</b>		
Human brain tissue	NIH NeuroBioBank	<a href="https://neurobiobank.nih.gov">https://neurobiobank.nih.gov</a>
Human brain tissue	Pritzker Neuropsychiatric Disorders Research Consortium	<a href="https://pritzkerneuropsych.org/www/about-us/scientific-approach/brain-bank">https://pritzkerneuropsych.org/www/about-us/scientific-approach/brain-bank</a>
<b>Chemicals, peptides, and recombinant proteins</b>		
SPRI Select Reagent	Fisher Scientific	Cat#: NC0406407
Lipofectamine LTX with Plus Reagent	ThermoFisher Scientific	Cat#: 15338100
Protector RNase inhibitor	Millipore Sigma	Cat#: 3335399001
Poly-L-ornithine solution	Millipore Sigma	Cat#: P4957-50ML
UltraPure 1M Tris-HCl, pH 8.0	Fisher Scientific	Cat#: 15568-025
Calcium Chloride Solution	Millipore Sigma	Cat#: 21115-100ML
Magnesium Acetate Solution	Millipore Sigma	Cat#: 63052-100ML
DTT, 1M	ThermoFisher Scientific	Cat#: P2325
EDTA, 0.5M	Millipore Sigma	Cat#: 324506-100ML
Triton X-100	Millipore Sigma	Cat#: T8787-100ML
OmniPur Sucrose	Millipore Sigma	Cat#: 8510-500GM
Bovine Serum Albumin Fraction V, protease-free	Millipore Sigma	Cat#: 3117332001
Tween 20	Millipore Sigma	Cat#: P9416-50ML
NP-40	Fisher Scientific	Cat#: 28324
Digitonin (5%)	ThermoFisher Scientific	Cat#: BN2006
PMSF	Millipore Sigma	Cat#: 10837091001
cOmplete protease inhibitor cocktail	Millipore Sigma	Cat#: 11697498001
Corning® 10X Phosphate-Buffered Saline (PBS), pH 7.4 ± 0.1, Liquid without calcium and magnesium, RNase-/DNase- and protease-free	Corning	Cat#: 46-013-CM
Sodium Deoxycholate 10%	BioWorld	Cat#: 40430018-2
UltraPure SDS Solution, 10%	ThermoFisher Scientific	Cat#: 15553027
Pierce™ 16% Formaldehyde (w/v), Methanol-free	ThermoFisher Scientific	Cat#: 28906
BioUltra Glycine	Millipore Sigma	Cat#: 50046-250G
DAPI Solution (1 mg/mL)	ThermoFisher Scientific	Cat#: 62248
Lithium chloride solution	Millipore Sigma	Cat#: L7026-500ML
Sodium Bicarbonate	Millipore Sigma	Cat#: S5761-500G
Proteinase K	Lucigen	Cat#: MPRK092
RNase A	Qiagen	Cat#: 19101
<b>Critical commercial assays</b>		
Nano-Glo Dual-Luciferase Reporter Assay System	Promega	Cat#: N1630
Global Diversity Array + NeuroBooster	Illumina	Cat#: 20031816

(Continued on next page)



**Continued**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chromium Next GEM Chip J	10X Genomics	Cat#: PN-1000230
Chromium Next GEM Single Cell Multiome ATAC + Gene Expression Reagent Bundle	10X Genomics	Cat#: PN-1000283
Dual Index Kit TT Set A	10X Genomics	Cat#: PN-1000215
DNeasy Blood and Tissue Kit	Qiagen	Cat#: 69506
Qubit dsDNA HS Assay Kit	ThermoFisher Scientific	Cat#: Q32854
Standard Sensitivity NGS Fragment Analysis Kit	Advanced Analytical	Cat#: DNF-473
<b>Deposited data</b>		
Raw and analyzed data	This paper	GEO: GSE214637
Code	<a href="https://doi.org/10.5281/zenodo.7405971">https://doi.org/10.5281/zenodo.7405971</a>	<a href="https://github.com/aanderson54/scMultiomics_AD">https://github.com/aanderson54/scMultiomics_AD</a>
<b>Experimental models: Cell lines</b>		
HEK293	ATCC	Cat# CRL-1573, RRID:CVCL_0045
293FT	ThermoFisher Scientific	Cat#: R70007, RRID:CVCL_6911
<b>Recombinant DNA</b>		
pGL4.23 [ <i>luc2</i> /minP]	Promega	Cat#: E8411
pNL1.1.CMV[ <i>Nluc</i> /CMV]	Promega	Cat#: N1091
<b>Software and algorithms</b>		
Prism v9	Graph Pad	<a href="https://www.graphpad.com/scientific-software/prism/">https://www.graphpad.com/scientific-software/prism/</a>
BioConductor v1.30.16	Huber et al. <sup>93</sup>	<a href="https://www.bioconductor.org">https://www.bioconductor.org</a>
scDbfFinder v1.9.4	Germain et al. <sup>94</sup>	<a href="https://github.com/plger/scDbfFinder">https://github.com/plger/scDbfFinder</a>
Seurat v4.1.1	Hao et al. <sup>22</sup>	<a href="https://github.com/satijalab/seurat">https://github.com/satijalab/seurat</a>
sctransform v0.3.3	Hafemeister and Satija <sup>95</sup>	<a href="https://github.com/satijalab/sctransform">https://github.com/satijalab/sctransform</a>
Harmony v0.1.0	Korsunsky et al. <sup>96</sup>	<a href="https://github.com/immunogenomics/harmony">https://github.com/immunogenomics/harmony</a>
MAST v1.20.0	Finak et al. <sup>97</sup>	<a href="https://github.com/RGLab/MAST">https://github.com/RGLab/MAST</a>
R v4.1.1	CRAN	<a href="https://cran.r-project.org">https://cran.r-project.org</a>
enrichR v3.0.0	Chen et al. <sup>98</sup>	<a href="https://github.com/wjawaid/enrichR">https://github.com/wjawaid/enrichR</a>
MACS2 v2.2.71	Zhang et al. <sup>99</sup>	<a href="https://github.com/macs3-project/MACS">https://github.com/macs3-project/MACS</a>
Signac v1.6.0	Stuart et al. <sup>23</sup>	<a href="https://github.com/stuart-lab/signac">https://github.com/stuart-lab/signac</a>
ComplexHeatmap v2.10.0	Gu et al. <sup>100</sup>	<a href="https://github.com/jokergoo/ComplexHeatmap">https://github.com/jokergoo/ComplexHeatmap</a>
GenomicRanges v1.46.1	Lawrence et al. <sup>101</sup>	<a href="https://bioconductor.org/packages/release/bioc/html/GenomicRanges.html">https://bioconductor.org/packages/release/bioc/html/GenomicRanges.html</a>
rtracklayer v1.54.0	Lawrence et al. <sup>102</sup>	<a href="https://bioconductor.org/packages/release/bioc/html/rtracklayer.html">https://bioconductor.org/packages/release/bioc/html/rtracklayer.html</a>
ChIPseeker v1.30.3	Yu et al. <sup>103</sup>	<a href="https://github.com/YuLab-SMU/ChIPseeker">https://github.com/YuLab-SMU/ChIPseeker</a>
cellSNP 1.2.2	Huang and Huang <sup>104</sup>	<a href="https://github.com/hxj5/cellsnp-lite">https://github.com/hxj5/cellsnp-lite</a>
vireo v0.5.6	Huang et al. <sup>105</sup>	<a href="https://github.com/single-cell-genetics/vireo">https://github.com/single-cell-genetics/vireo</a>
sLDSC v1.0.1	Finucane et al. <sup>54</sup>	<a href="https://github.com/bulik/ldsc">https://github.com/bulik/ldsc</a>
ENCODE ChIP-seq pipeline2	ENCODE DCC	<a href="https://github.com/ENCODE-DCC/chip-seq-pipeline2">https://github.com/ENCODE-DCC/chip-seq-pipeline2</a>
CellRanger arc 2.0.1	10X Genomics	<a href="http://software.10xgenomics.com">http://software.10xgenomics.com</a>

## RESOURCE AVAILABILITY

### Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Lindsay Rizzardi, Ph.D. ([lrizzardi@hudsonalpha.org](mailto:lrizzardi@hudsonalpha.org)).

### Materials availability

This study did not generate new unique reagents.

### Data and code availability

- Single nuclei RNA-seq, single nuclei ATAC-seq, and ChIP-seq data have been deposited at GEO and are publicly available under series accession number GSE214637.
- All original code generated during this study is publicly available at [https://github.com/aanderson54/scMultiomics\\_AD](https://github.com/aanderson54/scMultiomics_AD) and has been deposited at Zenodo. DOIs are listed in the [key resources table](#).
- Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

## EXPERIMENTAL MODEL AND SUBJECT DETAILS

### Cell lines

HEK293 cells were obtained from ATCC (CRL-1573) and grown in DMEM (high glucose, L-glutamine, no sodium pyruvate) (ThermoFisher Scientific Cat#11955-126), supplemented with 10% fetal bovine serum (FBS). 293FT cells were obtained from ThermoFisher Scientific (Cat#R70007) and maintained in DMEM (high glucose, L-Glutamine, 100 mg/L Sodium Pyruvate) (ThermoFisher Scientific Cat#11995081) supplemented with 10% FBS, 1% Glutamax (ThermoFisher Scientific Cat#35050061, 1% non-essential amino acids (NEAA) (ThermoFisher Scientific Cat#11140050), and 500 mg/mL Geneticin (G418 Sulfate) (ThermoFisher Scientific Cat#10131-035). All cells were cultured at 37°C with 5% CO<sub>2</sub>.

### Human brain tissues

Postmortem human brain biospecimens were obtained from the NIH Neurobiobank at the University of Miami and the Human Brain and Spinal Fluid Resource Center (HBSFRC) and from collaborators from the Pritzker Neuropsychiatric Disorders Research Consortium in the Department of Psychiatry and Human Behavior, University of California Irvine (UCI) as noted in [Table S1](#). Flash-frozen tissues were obtained from the dorsolateral prefrontal cortex (BA9/46) of seven donors diagnosed with Alzheimer's (Braak stages 4–6) and eight unaffected controls. AD donors were neuropathologically diagnosed according to CERAD criteria and Braak staging. All AD donors had a clinical diagnosis of AD and evidence of both amyloid beta plaques and neurofibrillary tangles. Demographic information for each donor is presented in [Table S1](#). No statistical methods were used to pre-determine sample sizes, but our sample sizes are similar to those reported in previous publications.<sup>18,19</sup> Data collection and analyses were not performed blind to tissue of origin. We did not pre-select samples based on *APOE* genotype, but genotype information was generated for each sample through TaqMan genotyping assays (see [APOE genotyping](#)).

## METHOD DETAILS

### Nuclei isolation from human brain tissues

Approximately 50–100 mg of frozen tissue per sample was homogenized in 4 mL of nuclei extraction buffer [0.32 M sucrose, 10 mM Tris pH 7.4, 5 mM CaCl<sub>2</sub>, 3 mM Mg acetate, 1 mM DTT, 0.1 mM EDTA, 0.1% Triton X-100, 0.2 U/μL Protector RNase inhibitor (Sigma cat. 3335399001)] by douncing 30 times in a 40 mL dounce homogenizer. Filter through 70 μm filter and spin at 500 x g, 5 min at 4°C in a swinging bucket centrifuge. Resuspend nuclei in 500 μL nuclei extraction buffer and layer over 750 μL sucrose solution (1.8 M sucrose, 10 mM Tris pH 7.4, 3 mM Mg acetate, 1 mM DTT) in a 1.5 mL tube. The samples were then centrifuged at >16,000 x g for 30 min at 4°C. After centrifugation, the supernatant was removed by aspiration and the nuclear pellet was resuspended in 125 μL PBS with 1% BSA and centrifuged 5 min at 500 x g at 4°C in a swinging bucket centrifuge. Permeabilization was performed according to 10X Genomics protocol CG000375 Rev B: nuclei were resuspended in 100 μL lysis buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM MgCl<sub>2</sub>, 1% BSA, 0.01% Tween 20, 0.01% NP-40, 0.001% digitonin, 1 mM DTT, 1 U/μL Protector RNase inhibitor) and incubated 2 min on ice. Nuclei were washed once and resuspended in 30 μL of 1X nuclei buffer with 1 mM DTT and 0.5 U/μL of Protector RNase inhibitor. Nuclei quality and concentrations were determined using the Countess II FL.

### Single nucleus multiomics

Transposition, nuclei isolation, barcoding, and library preparation were performed according to the 10X Genomics Chromium Next GEM Single Cell Multiome protocol CG000338 Rev E with the following alterations. The initial set of eight samples were processed as above (noted as “batch 1” in [Table S1](#)) and each sample was loaded across two lanes of the Chromium Next GEM Chip J. Nuclei were

loaded according to manufacturer's recommendations to target recovery of 10,000 nuclei per lane. The second batch of ten samples were processed as above, but two samples were pooled per lane of the Chromium Next GEM Chip J (each pool is indicated by sub-batch in [Table S1](#)). Each pool consisted of a male and female donor to facilitate assignment of each single cell back to the donor based on genotype and chrY gene expression (see [sample demultiplexing](#)). For these samples, we pooled 20,000 nuclei from each sample and the entire pool was processed according to the multiome protocol. Libraries were sequenced by HudsonAlpha Discovery using Illumina NovaSeq S4 flowcells.

### Sample demultiplexing

For lanes where a male and female sample were pooled together, reads were assigned to samples by genotyping cells. Variants were called from the cellranger output bam file for each cell using cell-snp-lite.<sup>104</sup> High-confidence SNPs from the 1000 Genome Project were used as a reference panel to call variants. Cell genotypes were then split by individual using vireoSNP<sup>105</sup> with the number of donors set to two. Cells were labeled as donor\_0, donor\_1, unassigned, or doublet. Unassigned and doublet cells were removed. Donor ID was assigned to the sample by observing the number of UMIs for genes on chrY. The donor ID with the higher mean counts was assigned to the male sample ([Table S11](#)).

### Joint snRNA-seq and snATAC-seq workflow

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  $> 2$ ). PMI-associated genes<sup>106</sup> were removed from the RNA counts matrix. Peaks that were present in less than 10 cells were removed from the ATAC matrix. Functions within Seurat<sup>22</sup> (v4.1.1, installed via Bioconductor v1.30.16<sup>93</sup>) were used for analysis. RNA counts were normalized with SCTransform<sup>95</sup> 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 second through the 50<sup>th</sup> LSI components. Doublet density was computed using *computeDoubletDensity* from scDbtFinder<sup>94</sup> where doublet score is the ratio of densities of simulated doublets to the density in the data. Cells with a doublet score  $> 3.5$  were removed. Normalization and dimension reduction were performed again on the filtered set with the same parameters. Predicted cell types were determined for each cell using Seurat SCT-normalized reference mapping. Gene expression data was mapped to SCT-normalized DLPFC data<sup>12</sup> and annotated with the cell types of the reference map. Cells with a predicted cell type score less than 0.95 were removed from the data. Batch effects were corrected in RNA ( $\theta = 1$ ) and ATAC ( $\theta = 2$ ) with Harmony (v1.0.0)<sup>96</sup> by removing the effect of sample.

### WNN analysis of snRNA-seq and snATAC-seq

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 harmony-corrected RNA reduction and the second through the 50<sup>th</sup> dimensions from the harmony-corrected ATAC reduction were used to create the graph. The WNN UMAP was created using the wknnc (k = 20) ([Figure S5](#)).

### Differential expression

Differentially expressed genes (DEGs) were determined for AD versus control for each cell type. Within each cell type, the gene expression data was log-normalized with a scale factor of  $1 \times 10^5$ . Pericytes and Endothelial cells were not included in the analysis because of small cell counts. Differential expression was assessed using MAST<sup>97</sup> for genes present in at least 25% of either AD or control cells. Age and sex were included as covariates in the MAST model. Genes with a Bonferroni-adjusted p value  $< 0.01$  and an absolute  $\log_2(\text{fold change}) > 0.25$  were determined to be significant. DEGs between cell types were determined using MAST with age and sex as covariates for genes present in at least 25% of cells. Genes with a Bonferroni adjusted p value  $< 0.01$  and an absolute  $\log_2(\text{fold change}) > 0.5$  were determined to be significant.

### Annotation of cell subpopulations

Cell type subclusters were identified using weighted snRNA and snATAC modalities. Expression data were normalized with *SCTransform*, and chromatin accessibility data were normalized with TFIDF within each cell type. Normalized values were used to construct a multimodal weighted nearest neighbor graph (k = 20). Clusters were identified using wknnc and the SLM algorithm. The resolution (0.3, 0.2, 0.3, 0.3, 0.45) was adjusted for each cell type (Astro, Inh, Exc, Olig, Mic). Any clusters with  $< 100$  cells were excluded from DEG analysis. Within each cell type, cluster DEGs were determined for each subcluster versus all other subclusters. DEGs were defined as those with a Bonferroni adjusted p value  $< 0.01$  using MAST with age and sex as covariates. Only genes that were detected in at least 25% of cells in a subcluster were considered.

Neuronal subclusters were further annotated with Azimuth<sup>22</sup> Human motor cortex<sup>107</sup> clusters to identify known neuronal subpopulations. For each neuronal subcluster, a subtype was assigned by the enrichment for upregulated subcluster DEGs in Azimuth gene sets. Enrichment was performed using *enrichR*<sup>98,108,109</sup> and the Azimuth Cell Types 2021 gene sets. The top subtype annotation was assigned to a subcluster if the adjusted p value was  $< 0.01$ .

AD-specific subclusters and subtypes were determined by observing overrepresentation of cells isolated from AD individuals. Statistically significant overrepresentation was evaluated with a Fisher exact test and adjusted p values.

### Gene set enrichment

The R package *enrichR*<sup>98,108,109</sup> was used for all gene set enrichment analyses. Sets of DEGs and peak-linked genes 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

ATAC peaks were called independently for each cell type using MACS2<sup>99</sup> and Signac<sup>23</sup> *CallPeaks* and the union of these peaks was used in subsequent analyses retaining the cell type annotations. The peaks were then annotated with *ChIPseeker*<sup>103</sup> and *TxDb.Hsapiens.UCSC.hg38.knownGene* where promoters were considered to be 1 kb upstream and 100 bp downstream of the TSS. Only ATAC peaks that were present in at least 2% of cells in at least one cell type were included in the analyses. AD and control links were identified separately via the *cellranger-arc* (v2.0.1) *reanalyze* function using the filtered cell type ATAC peaks and either AD or control expression and accessibility data as input. The maximum interaction distance was restricted to 500 kb. Peak-peak links were not used for downstream analysis. For feature linkage calculation, ATAC and gene expression 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 Gaussian 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.<sup>32</sup> Links with an absolute correlation score <0.2 and linked to a gene with <200 UMIs were removed. Other packages used in these analyses were *GenomicRanges* (v1.46.1)<sup>101</sup> and *ComplexHeatmap* (v2.10.0).<sup>100</sup>

### Permutation testing

We performed 100 sample permutations calling links. Sample sets for the permutations were determined by randomly selecting seven or eight individuals regardless of disease status. Links were called with the same parameters as the true data. Links were independently overlapped with links from all other permutations to get the proportion of links that were group-specific for each pair of permutations. Group-specific links were defined as the links that were not shared between any two permutations. Given the distribution of group-specific links across all permutation pairs, we tested if AD and control-specific links were an outlier for group-specificity using a Z-test.

### Peak-gene-TF trios

Trios were called for a filtered set of links by removing those >100kb from a TSS and those with an absolute score <0.2. Motifs were then called in each linked-peak using Signac *AddMotifs* and the JASPAR 2022<sup>110</sup> CORE PFM. Peaks with >100 motifs were additionally filtered from the link set. TF expression, linked gene expression, and linked peak accessibility matrices for trio correlation were derived from the average counts within metacells. Metacells were determined using WNN clusters for all AD cells and all control cells separately. TF-peak scores are the Pearson correlation between peak accessibility and the expression of the TF whose motif was called in the peak. TF-gene scores are the Pearson correlation between a gene and the TF whose motif was called in the linked peak. Significant associations were defined as those with a p value <0.001. Significant trios were then defined as those with a significant positive TF-peak correlation and a significant TF-gene correlation.

### Partitioned heritability analysis

To evaluate whether linked peaks are enriched for common genetic variants that have been associated with AD or other traits by GWAS, we performed stratified linkage disequilibrium (LD) score regression (sLDSC v1.0.1).<sup>60,111</sup> sLDSC estimates the proportion of genome-wide SNP-based heritability that can be attributed to SNPs within a given genomic feature by a regression model that combines GWAS summary statistics with estimates of linkage disequilibrium from an ancestry-matched reference panel. Summary statistics for AD were downloaded from.<sup>2-6</sup> To estimate SNP heritability from AD GWAS summary statistics, we excluded the *APOE* and *MHC/HLA* genomic regions. Additional GWAS summary statistics were downloaded for brain-related<sup>57-61,112</sup> and other traits.<sup>62-66</sup> Each category (all, common, AD, control) corresponds to the analysis in which the peak-gene link was identified. Cell type is assigned based on the cell type(s) in which the linked peak was identified. Peaks were resized to 1 kb and each set of unique peaks with these categories was tested individually along with the full baseline model (baseline-LD model v2.2.) that included 97 categories capturing a broad set of genomic annotations. Note that a peak can have multiple links that fall in different categories. GWAS summary statistics are available from the websites listed in Table S8. Additional files needed for the sLDSC analysis were downloaded from <https://alkesgroup.broadinstitute.org/LDSCORE/> following instructions at <https://github.com/bulik/ldsc/wiki>.

### APOE genotyping

To determine *APOE* status, TaqMan genotyping assays (cat#: 4371353) were used to genotype SNPs rs429358 and rs7412 (cat#: 4351379, C\_\_3084793\_20 and C\_\_904973\_10, respectively) following the manufacturer's instructions. Genotyping calls were made using QuantStudio software (v1.3) for all individuals in this study. *APOE* status is reported in Table S1.

### Comparisons to external data sources

Cell type-specific H3K27ac peak calls were obtained from<sup>11</sup> and converted to hg38 coordinates using the *liftOver* function from the R package rtracklayer.<sup>102</sup> GABA and GLU neuronal subtype H3K27ac fastqs from Kozlenkov et al.<sup>36</sup> were downloaded from Synapse (syn12033252) and processed as individual replicates using the AQUAS Transcription Factor and Histone ChIP-Seq processing pipeline<sup>113</sup> ([https://github.com/kundajelab/chipseq\\_pipeline](https://github.com/kundajelab/chipseq_pipeline)). Peaks were called using the IDR naive overlapping method with a threshold of 0.05 and the optimal peak sets were used. For each cell type, only peaks identified in at least 3 individuals were retained for downstream analyses. ATAC-seq peaks from non-neuronal cell types were intersected with H3K27ac data from the corresponding cell type obtained from Nott et al.<sup>11</sup> Excitatory and inhibitory neuron ATAC-seq peaks were intersected with H3K27ac peaks identified from GLU (NeuN+/SOX6-) or GABA (NeuN+/SOX6+) neuronal nuclei<sup>36</sup> and from neuronal (NeuN+) nuclei.<sup>11</sup> MPRA data were obtained from,<sup>21,67–69</sup> eQTL data were obtained from,<sup>43,70</sup> and neuronal HiC loop calls were obtained from.<sup>71</sup>

### Plasmids

The pNL1.1.CMV [*Nluc*/CMV] (Cat# N1091) and pGL4.23 [*luc2*/minP] (Cat# E8411) vectors were obtained from Promega. Luciferase elements were generated by selecting 467 bp of the nominated region using hg38 coordinates. Both the forward and reverse complement sequences were ordered as gBlocks from Integrated DNA Technologies (IDT). Gibson assembly was performed by cloning elements into the pGL4.23 [*luc2*/minP] vector digested with *EcoRV*. 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.

### Transfection

HEK293 and 293FT cells were plated at 70,000 cells/cm<sup>2</sup> in a 24-well format. Before plating 293FT cells, culture plates were pre-coated with poly-L-ornithine solution (Millipore Sigma Cat#P4957-50ML). The next day, cells were transfected with 1 μg of plasmid DNA using Lipofectamine LTX with Plus Reagent (ThermoFisher Scientific Cat#15338-100) following the manufacturer's recommendations. Per transfection, 900 ng of luciferase element and 100 ng of pNL1.1.CMV [*Nluc*/CMV] were used. A transfection reaction of 900 ng pGL4.23 [*luc2*/minP] and 100 ng pNL1.1.CMV [*Nluc*/CMV] was used as a baseline control. Both vectors were also transfected as background controls (100 ng) with pmaxGFP (900 ng, Lonza). Cell lysates were harvested by freezing at –80°C 48 h 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.

### Chromatin preparation for sorted nuclei

Buffers required: Nuclei Extraction Buffer (NEB): 0.32 M Sucrose, 5 mM CaCl<sub>2</sub>, 3 mM Mg(Ac)<sub>2</sub>, 0.1 mM EDTA, 10 mM Tris-HCl, 0.1 mM PMSF, 0.1% Triton X-100, 1 mM DTT. Before use, add Roche cOmplete protease inhibitor cocktail according to manufacturer recommendation (Sigma 11697498001). Sucrose Cushion Buffer (SCB): 1.6 M Sucrose, 3 mM Mg(Ac)<sub>2</sub>, 10 mM Tris-HCl, 1 mM DTT. Interphase Buffer: 0.8 M Sucrose, 3 mM Mg(Ac)<sub>2</sub>, 10 mM Tris-HCl. Blocking buffer: 1x PBS, 1% BSA, 1 mM EDTA. Pellet buffer: add up to 200 μL 1 M CaCl<sub>2</sub> to 10 mL SCB. RIPA: 1x PBS, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS.

Methods for extracting and sorting nuclei from postmortem brain are similar to previously published methods.<sup>114</sup> Here, approximately 500 mg of tissue was placed into a chilled 40 mL Dounce homogenizer containing 5 mL of NEB on ice and allowed to partially thaw to ease douncing (2–3 min). Extract nuclei by douncing with “tight” pestle 30–40 times until the tissue is homogenized. Transfer to 15 mL conical tube on ice, wash glassware with 5 mL NEB and add to 15 mL tube. Fix chromatin by adding 625 μL of 16% formaldehyde (methanol free, Thermo 28,906) to a final concentration of 1% and rotate end-over-end at room temperature for 10 min. Halt fixation by adding 500 μL of 2.5 M Glycine and incubate another 5 min rotating at room temperature then place homogenate back on ice. During fixation, prepare sucrose gradient in 2 ultracentrifuge buckets (Beckman Coulter cat: 344058) by layering 5 mL of Interphase buffer atop 10 mL of SCB in each. Carefully layer nuclei homogenate atop sucrose gradient, balance with NEB, then ultracentrifuge at 24,000 rpm for 2 h using SW28 swinging bucket rotor (Beckman Coulter). Upon completion, inspect tubes for a visible pellet of nuclei at the bottom of tube. Remove debris at interphase first by using a 25 mL graduated pipette, then continue removing the remaining sucrose gradient being careful not to disturb the nuclei pellet. Carefully resuspend the pellet in 1 mL cold PBS and transfer to a 15 mL lo-bind tube containing 2 mL PBS on ice. (Optional: if pellet appears to contain large debris then pass through 70 μm filter).

Wash ultracentrifuge tubes with 1 mL cold PBS and combine in 15 mL tube to a final volume of 10 mL, inverting to mix. Centrifuge the nuclei at 1,000 x g for 10 min at 4°C to remove residual sucrose. Label nuclei by resuspending pellet in 5 mL blocking buffer with NeuN-488 antibody (Millipore cat: MAB377X) and OLIG2 antibody (Abcam cat: ab109186) at 1:5,000 each. Incubate nuclei in staining buffer with rotation for at least 1 h at 4°C. Spin nuclei 500 x g for 5 min to pellet, remove supernatant, then resuspend in 5 mL blocking buffer with goat-*anti*-rabbit-647 (ThermoFisher cat: A-21245) at 1:5,000 and DAPI at 1:100,000. Incubate for at least 1 h at 4°C with rotation. Remove stain by centrifuging 500 x g 5 min at 4°C and resuspending in 3 mL cold PBS. Hold on ice and proceed immediately to sorting.

Nuclei were sorted using Sony MA900 with a 70  $\mu$ m nozzle and pressure not exceeding pressure setting of 7. Gates were set to capture those populations that were positive for 488 signal (NeuN<sup>+</sup>), positive for 647 signal (OLIG<sup>+</sup>), or negative for both (NeuN<sup>-</sup>;OLIG<sup>-</sup>). The NeuN<sup>+</sup> population was collected into 5 mL tubes held at 4°C and pooled into 15 mL lo-bind tubes on ice. Purity of selected samples were typically >95% based on reanalysis of sorted samples. To concentrate nuclei for downstream analysis, add approximately 2 mL of pellet buffer per 10 mL of sorted nuclei and rotated at 4°C for 15 min. Centrifuge 500 x g for 10 min at 4°C, after which a pellet should be visible. Remove supernatant and carefully resuspend pelleted nuclei in at least 3 mL cold PBS. Centrifuge 500 x g for 5 min at 4°C.

To generate chromatin for ChIP-seq, resuspend pellet in cold RIPA plus protease inhibitor (Roche, cat: 11836153001) at approximately 3 million nuclei per 250  $\mu$ L. Transfer 250  $\mu$ L of each sample to the Bioruptor tubes (Diagenode, cat: C30010016) and sonicate tissue using a Bioruptor Pico (8 cycles; 30 s on/30 s off). Pool the sonicated chromatin into a 1.5 mL DNA lo-bind conical tube and centrifuge 12,000 x g for 5 min at 4°C to remove any insoluble debris. Collect supernatant into a separate tube, add RIPA to final volume equivalent to 500,000 nuclei per 100  $\mu$ L, then dispense working aliquots into 1.5 mL tubes held on dry ice. Store at -80°C.

### ChIP-seq protocol

ChIP-seq for ZEB1 was performed using chromatin from NeuN<sup>+</sup> nuclei from DLPFC, occipital lobe, and frontal pole from two control donors serving as biological replicates. ChIP-seq for MEF2C was performed on bulk DLPFC tissue from two control donors serving as biological replicates. Protocols for ChIP-seq are similar to those for frozen tissue previously described by our lab<sup>115,116</sup> and consistent with techniques recommended by the ENCODE Consortium ([www.encodeproject.org/documents](http://www.encodeproject.org/documents)). Briefly, 60  $\mu$ L Dynabeads (ThermoFisher, cat: 11203D) were washed with cold 1x PBS +5 mg/mL BSA then combined with 8  $\mu$ L antibody targeting ZEB1 (Bethyl, cat: A301-921A) or MEF2C (proteintech, cat: 18290-1-AP) in a final volume of 200  $\mu$ L 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  $\mu$ L RIPA and brought to 200  $\mu$ L with 100  $\mu$ L chromatin aliquot. Incubate bead/antibody with chromatin using rotation for 1 h 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 Na<sub>2</sub>EDTA). Chromatin was eluted from beads by incubating with intermittent shaking for 1 h at 65°C in IP elution buffer (1% SDS, 0.1 M NaHCO<sub>3</sub>), followed by incubating overnight at 65°C to reverse formaldehyde cross-links. DNA was purified using DNeasy Blood and Tissue kit (Qiagen 69,506) and eluted in a final volume of 50  $\mu$ L 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  $\mu$ L with RIPA and reverse-crosslinked overnight at 65°C. The following morning, samples were incubated an additional 30 min with 20  $\mu$ L Proteinase K and 4  $\mu$ L RNase A (Qiagen 19,101) and subsequently eluted for DNA using DNeasy Blood and Tissue kit. The entirety of the remaining IP DNA (and approximately 250 ng Input control) was used to generate sequencing libraries. Libraries were prepared by blunting and ligating ChIP DNA fragments to sequencing adapters for amplification with barcoded primers (30 s at 98°C; [10 s at 98°C, 30 s at 65°C, 30 s 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 flow cell with 100 bp 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)<sup>117</sup> [-a AGATCGGAAGAGC -m 40]. 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>). All software within the package was run using the default settings or those recommended by the authors for transcription factors. Final peaks were called using the IDR naive overlapping method with a threshold of 0.05. For ZEB1, IDR optimal peaks from the three cortical brain regions were merged for downstream analyses.

### QUANTIFICATION AND STATISTICAL ANALYSIS

The quantitative and statistical analyses are described in the relevant sections of the [method details](#) or in the figure legends.