


RESEARCH

Open Access



Integrative single-nucleus multi-omics analysis prioritizes candidate *cis* and *trans* regulatory networks and their target genes in Alzheimer's disease brains

Julia Gamache^{1,2†}, Daniel Gingerich^{1,2†}, E. Keats Shwab^{1,2†}, Julio Barrera^{1,2}, Melanie E. Garrett³, Cordelia Hume^{1,2}, Gregory E. Crawford^{2,5,6*}, Allison E. Ashley-Koch^{3,4*} and Ornit Chiba-Falek^{1,2*} 

Abstract

Background The genetic underpinnings of late-onset Alzheimer's disease (LOAD) are yet to be fully elucidated. Although numerous LOAD-associated loci have been discovered, the causal variants and their target genes remain largely unknown. Since the brain is composed of heterogeneous cell subtypes, it is imperative to study the brain on a cell subtype specific level to explore the biological processes underlying LOAD.

Methods Here, we present the largest *parallel* single-nucleus (sn) multi-omics study to simultaneously profile gene expression (snRNA-seq) and chromatin accessibility (snATAC-seq) to date, using nuclei from 12 normal and 12 LOAD brains. We identified cell subtype clusters based on gene expression and chromatin accessibility profiles and characterized cell subtype-specific LOAD-associated differentially expressed genes (DEGs), differentially accessible peaks (DAPs) and *cis* co-accessibility networks (CCANs).

Results Integrative analysis defined disease-relevant CCANs in multiple cell subtypes and discovered LOAD-associated cell subtype-specific candidate *cis* regulatory elements (cCREs), their candidate target genes, and *trans*-interacting transcription factors (TFs), some of which, including *ELK1*, *JUN*, and *SMAD4* in excitatory neurons, were also LOAD-DEGs. Finally, we focused on a subset of cell subtype-specific CCANs that overlap known LOAD-GWAS regions and catalogued putative functional SNPs changing the affinities of TF motifs within LOAD-cCREs linked to LOAD-DEGs, including *APOE* and *MYO1E* in a specific subtype of microglia and *BIN1* in a subpopulation of oligodendrocytes.

Conclusions To our knowledge, this study represents the most comprehensive systematic interrogation to date of regulatory networks and the impact of genetic variants on gene dysregulation in LOAD at a cell subtype resolution. Our findings reveal crosstalk between epigenetic, genomic, and transcriptomic determinants of LOAD pathogenesis and define catalogues of candidate genes, cCREs, and variants involved in LOAD genetic etiology and the cell

[†]Julia Gamache, Daniel Gingerich and E. Keats Shwab have contributed equally to this work.

*Correspondence:

Gregory E. Crawford
greg.crawford@duke.edu
Allison E. Ashley-Koch
allison.ashleykoch@duke.edu
Ornit Chiba-Falek
o.chibafalek@duke.edu

Full list of author information is available at the end of the article



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

subtypes in which they act to exert their pathogenic effects. Overall, these results suggest that cell subtype-specific *cis-trans* interactions between regulatory elements and TFs, and the genes dysregulated by these networks contribute to the development of LOAD.

Keywords Late onset Alzheimer's disease, Single-nucleus (sn)RNA-seq, snATAC-seq, Regulatory networks, Chromatin accessibility, Transcriptomics, Epigenomics, Multi-omics, Genetic variant

Introduction

Alzheimer's disease (AD) is a complex multifactorial neurodegenerative disorder characterized by extracellular A β deposits, referred to as plaques, as well as intracellular neurofibrillary tangles consisting of hyperphosphorylated tau, followed by synaptic and neuronal loss resulting in progressive cognitive and functional decline. Late onset AD (LOAD) is the common form of the disease with heterogeneous genetic etiologies [1, 2]. The complex genetic architecture of LOAD has been studied over the past 3 decades (see Lambert et al. [3] for review). The e4 allele of the apolipoprotein E gene (*APOE* e4) was the first, strongest, and most firmly established genetic risk factor for LOAD [4–7]. The initial discovery was made nearly 30 years ago by linkage analysis of pedigrees [4] and over the ensuing years it has become the most highly replicated genetic risk factor for LOAD [4–7]. Subsequent large multi-center genome-wide association studies (GWAS) have confirmed the strong association with the *APOE* genomic region and identified associations with numerous additional genomic loci [8–14]. The most recent GWAS meta-analyses reported 75 risk loci for LOAD [15, 16]. However, the precise disease-causing genes, the specific causal genetic variants, and the molecular mechanisms mediating their pathogenic effects have not yet been explained. Most LOAD-GWAS variants are in noncoding genomic regions [17]. Previous studies suggested that some noncoding LOAD SNPs are located in regulatory elements such as enhancers, affecting their functions, and thereby impacting gene expression [18], including expression of distal genes [19]. Thus, to untangle the genetic and genomic architecture of LOAD and to translate LOAD genetic association discoveries to causal mechanisms of disease, LOAD GWAS variants need to be assigned to the correct target genes, rather than merely the nearest gene [19, 20]. Moreover, it is imperative to map these variants and their linked genes to the specific brain cell-type in which they exert their pathogenic effect.

Many functional genomic studies provide evidence for the role of gene dysregulation in LOAD pathogenesis, including those examining specific disease-related genes [21, 22], pathways [23], expression quantitative trait loci (eQTLs) [24–26], differential transcriptome profiles [27],

and the DNA methylation [28–31] and histone mark [32] landscapes in human brain tissues. However, these studies were conducted in bulk brain tissue, and therefore cannot specify the cell type(s) in which gene expression or epigenetic changes occur. Mixed cell subtypes could also mask signals corresponding to a particular cell subtype, especially if the causal cell subtypes comprise a small fraction of the entire sample. Furthermore, bulk tissue studies are confounded by sample-to-sample variation in cell type composition, which could be exacerbated by the neuronal loss and proliferation of glial cells accompanying LOAD [33]. Transcriptomic and epigenomic studies using sorting techniques to separate broad cell types from LOAD brain tissue [34–39] for example, neuronal vs. non-neuronal, have provided new important insights, but even within these categories, there are many different cell types and subtypes in the human brain [40]. Single-cell experimental approaches can circumvent these limitations and inform LOAD-specific epigenomic and transcriptomic changes with unparalleled precision.

The past three years have seen a transition into single-cell multi-omics studies in LOAD functional genomic research. Single-cell transcriptomic studies have achieved previously unattainable resolution in identifying LOAD-associated cell type-specific changes in gene expression [41, 42]. These studies demonstrate the importance of examining gene dysregulation at the cell type-specific level. However, they focused only on single-nucleus (sn) RNA-seq data and therefore the underlying regulatory mechanisms for these gene expression signatures remain to be identified. More recently, integrative multi-omics single-nucleus framework studies profiling chromatin accessibility and gene expression [43, 44] identified cell type-specific, disease-associated candidate *cis*-regulatory elements (cCREs) and their candidate target genes, and demonstrated the utility and potential of this strategy in moving LOAD genetic research forward.

The diagram in Fig. 1a presents an overview of this study pipeline. We used the 10X Genomics platform to perform *parallel* snRNA-seq and snATAC-seq analyses simultaneously from the same pool of nuclei derived from post-mortem temporal cortex tissue of LOAD patients and neuropathologically normal controls. We used these datasets to identify LOAD-associated

differentially expressed genes (DEGs), differentially accessible peaks (DAPs) and *cis* co-accessibility networks (CCANs) at the cell subtype level. The *parallel* experimental design allowed us to integrate these LOAD profiles to promote the mechanistic understanding of gene dysregulation in LOAD. We identified LOAD-associated cell subtype-specific cCREs, their target genes, and the transcription factors (TFs) that may mediate their effects on gene expression changes in LOAD. We found that the expression of a subset of these TFs also changed in LOAD in the same specific cell subtype. Moreover, focusing on LOAD-GWAS regions, we catalogued putative regulatory SNPs positioned in the identified LOAD-associated cell subtype-specific cCREs that change the affinities of TF motifs and thereby potentially affect the expression of the target genes. Collectively, we provided new insights into the relationships between DNA sequence variation, chromatin structure, and transcriptome in LOAD brains at a cell subtype resolution. Furthermore, we identified candidate *cis*- and *trans*- regulatory networks and their target genes for further experimental investigations in model systems relevant to LOAD.

Results

Characterization of cell types and subtypes in the human temporal cortex of healthy aging and Alzheimer's individuals using multi-omics datasets

We isolated nuclei samples from frozen post-mortem human temporal cortex (TC) tissues of 12 LOAD and 12 cognitively normal individuals (Table 1, Additional file 2: Table S1) and performed both snRNA-seq and snATAC-seq in parallel by simultaneously using two aliquots from the same nuclei sample (Fig. 1a). To exclude the APOE ϵ 4 effect, all donors were APOE ϵ 3/3. In both snRNA-seq and snATAC-seq nuclei, we discovered multiple neuronal and glial cell subtypes which were linked across both data sets for downstream analyses (Fig. 1b). This parallel experimental design allowed us to minimize differences

between the snRNA-seq and snATAC-seq assays, including technical variables and variability in cell type composition, and facilitated the multi-omics integrative analyses outlined in Fig. 1.

We first annotated the cell types in our snRNA-seq dataset by the label transfer method [45] using a pre-annotated reference snRNA-seq dataset generated with the same technology [40], and validated these annotations by examining expression of known cell type-specific marker genes [41], and additionally by label transfer from a human prefrontal cortex snRNA-seq dataset [43] (Methods, Additional file 1: Figure S1a and b). After quality control (QC) filtering, we retained a total of 209,518 nuclei for snRNA-seq from all 24 temporal cortex samples (Additional file 2: Table S2). After dimensionality reduction followed by Louvain community detection, we identified 33 distinct cell clusters (Figs. 2a, 3a, Additional file 2: Tables S3–S4) representing cell subtypes of astrocytes (Astro), excitatory neurons (Exc), inhibitory neurons (Inh), microglia (Micro), oligodendrocytes (Oligo), and oligodendrocyte precursor cells (OPC). Cell subtypes were each given a unique label according to their broader cell type (*e.g.*, the 11 clusters of excitatory neurons were labeled Exc1-Exc11, Additional file 2: Table S4). Oligodendrocytes were the most commonly identified cell type in our dataset (85,801 nuclei), with excitatory neurons being the second most common (52,980 nuclei; Additional file 2: Table S4). This highest proportion of oligodendrocytes was consistent with the results reported by Morabito et al. [43], whereas Mathys et al. [41] and Anderson et al. [44] reported excitatory neurons as the most common cell type in similar single-cell experiments. It is possible that the relative proportions of the major cell types recovered during tissue preparation may be affected by the particular brain region of the examined tissue, and/or technical differences between laboratory settings such as methods of tissue preservation, preparation, homogenization, and protocols of nuclei purification. Cells associated with brain vasculature are

(See figure on next page.)

Fig. 1 Experimental approach and integration of snATAC-seq with snRNA-seq clusters. **a**, Schematic of nuclei isolation and parallel snATAC-seq and snRNA-seq library generation. Temporal cortex tissue was collected from 12 Normal and 12 LOAD donors → Nuclei were extracted from tissue samples by homogenization followed by sucrose gradient purification → Gel beads in emulsion (GEM) generation followed by parallel gene expression and chromatin accessibility library generation according 10X Genomics protocol and sequencing → Cell type annotation and subtype clustering for snRNA-seq data based on gene expression profile comparison to reference dataset, followed by matching of subtype clusters between snRNA-seq and snATAC-seq datasets → Identification of differentially expressed genes (DEGs) and differentially accessible peaks (DAPs) via linear mixed effects modeling → Identification of *cis*-coaccessibility networks (CCANs) incorporating DEGs and DAPs, and categorizing directionality of regulation → Identification of candidate coregulatory elements (cCREs), DAPs coaccessible with DEG regulatory regions, and characterization of biological function category enrichment of associated DEGs → Characterization of enriched TF binding motifs within regulatory regions of DEGs in CCANs → Identification of SNPs impacting TF binding to DEG regulatory regions within 0.5 Mb of GWAS loci. **b**, Flow chart of snRNA-seq and snATAC-seq analytical pipeline leading to cluster assignment, and schematic of analytical strategy with label transfer of snRNA-seq clusters onto snATAC-seq data by cell type and subtype

a

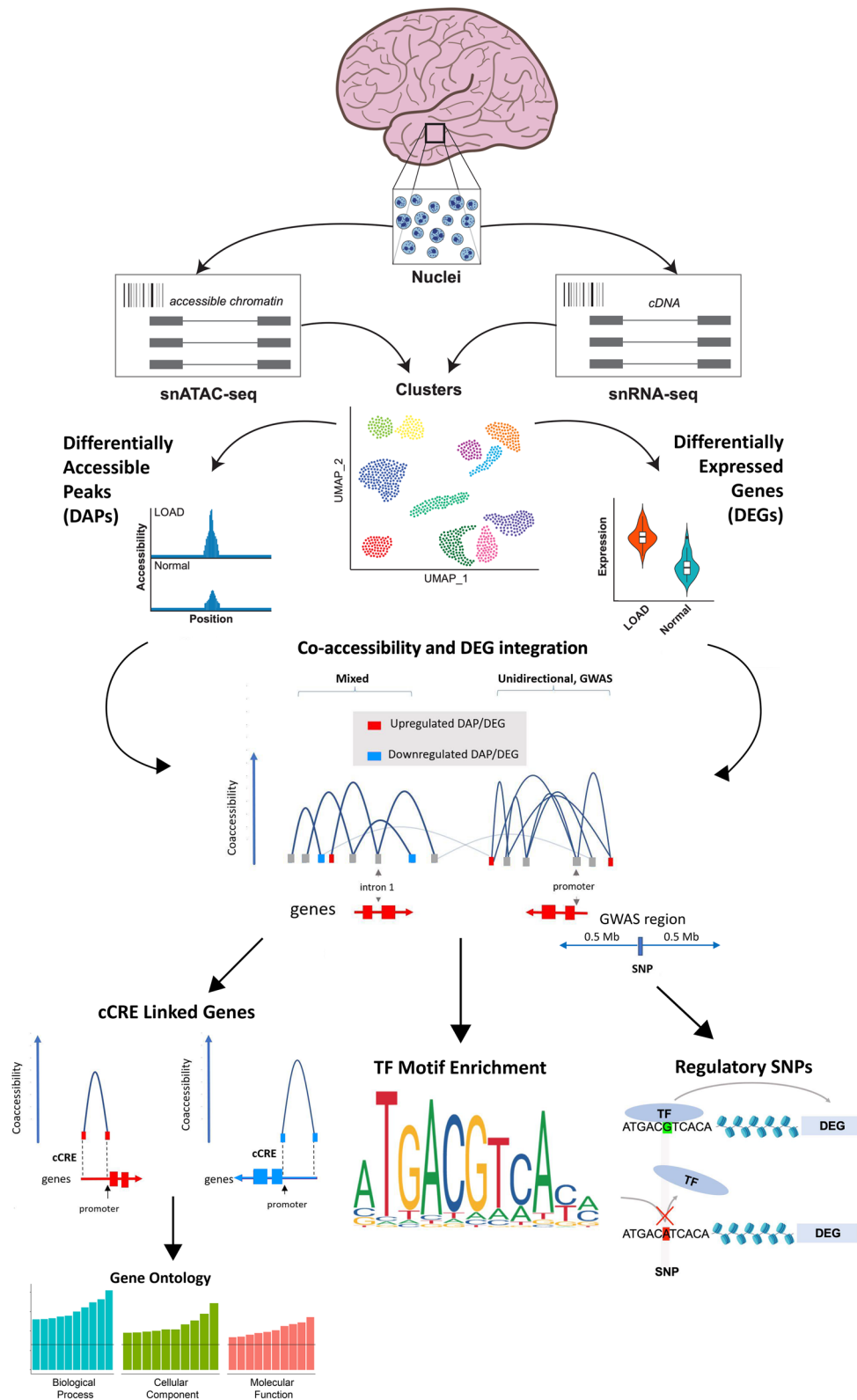


Fig. 1 (See legend on previous page.)

b

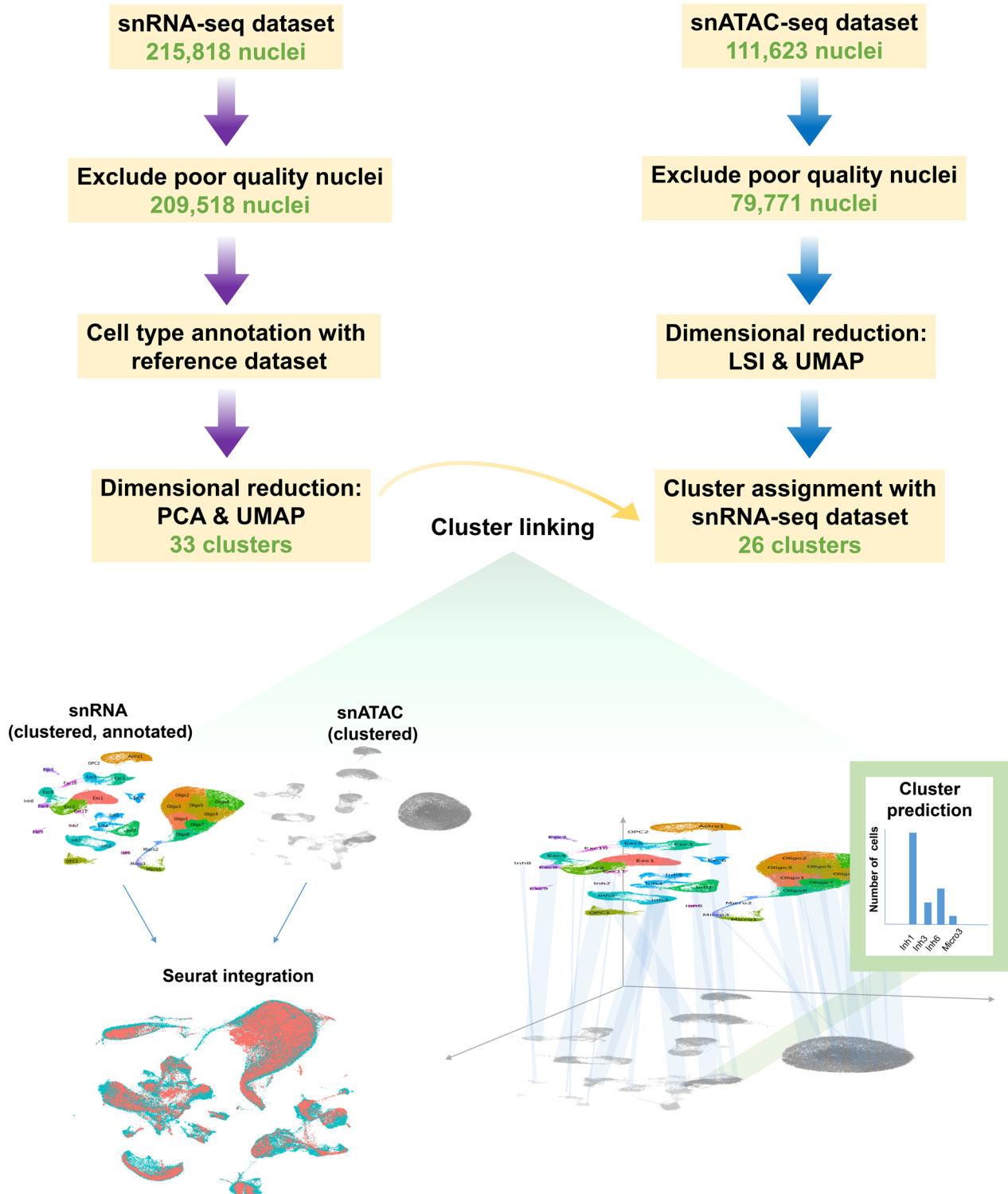


Fig. 1 continued

Table 1 Demographics summary of study cohort

Diagnosis	Sex	n	Age (mean ± SD)	Statistics	PMI (mean ± SD)	Statistics
LOAD	Female	6	84.83 ± 7.52	U = 23, p = 0.48	10.59 ± 4.73	t = 0.40, df = 10, p = 0.70
Normal	Female	6	77.83 ± 10.46		9.36 ± 5.83	
LOAD	Male	6	77.67 ± 7.09	t = 0.32, df = 10, p = 0.76	4.38 ± 6.17	U = 7, p = 0.092
Normal	Male	6	75.83 ± 12.16		7.44 ± 6.23	

PMI post-mortem interval, SD standard deviation, t Student's t-test statistic (for normal data), U Mann–Whitney U-test statistic (for non-normal data), df degrees of freedom, p probability value

challenging to recover due to the necessity of dissociation from the vessel basement membrane and require enrichment procedures or sample pooling to obtain sufficient nuclei quantities for analysis [46, 47]. As such, vascular endothelial cells and vascular and leptomeningeal cells (VLMCs) represented only 0.22% of recovered nuclei in our dataset following QC filtering and were therefore excluded from downstream analyses. Previously known neuronal subtypes such as somatostatin interneurons [48–50] tended to form separate clusters, and a particular subtype of excitatory neurons marked by *LAMP5* was notably reduced (false discovery rate (FDR) adjusted $p < 0.01$) in LOAD tissue (Additional file 1: Figure S1d and e). To check for donor-based batch effects, we examined the distribution of nuclei across subtype clusters for each donor sample. Donor samples overall showed qualitatively even distribution across cell subtype clusters (Additional file 1: Figure S2a).

Similarly, 79,771 nuclei for snATAC-seq were retained after QC filtering (Additional file 2: Table S5) and Louvain community detection yielded 26 cell subtype clusters (Figs. 2b, 4a). We mapped the resulting snATAC-seq clusters to the corresponding snRNA-seq clusters through the identification of shared transfer anchors between the datasets based on observed gene expression profiles for snRNA-seq nuclei and predicted gene expression profiles for snATAC-seq nuclei (Methods, Fig. 1b). snATAC-seq clusters were linked to most snRNA-seq clusters with

the exception of seven cell subtypes: Inh7, Inh8, Micro3, Oligo2, Oligo7, Oligo8, and OPC2. Validation of the cell subtype linking method accuracy based on a test multiome dataset (see Methods) showed an overall 75.6% similarity based on Jaccard index and high concordance with a mean of ~0.89 hybrid score [42] (Additional file 2: Table S6). We also checked the donor-based batch effects in snATAC-seq clustering by examining the distribution of nuclei across subtype clusters for each donor sample, and showed again qualitatively even distribution across subtypes (Additional file 1: Figure S2b).

LOAD brains are pathologically characterized by neuronal loss and gliosis [51]. Thus, we examined differences in mean cell type and subtype proportions (number of nuclei from a particular donor sample belonging to a particular cell type/subtype divided by total number of nuclei for the same sample) between LOAD and normal samples. Using the snRNA-seq dataset from the quality control (QC)-filtered nuclei we observed significant decreases in the proportions of astrocyte and OPC nuclei (FDR < 0.001) and increases in the proportions of inhibitory neurons and oligodendrocytes nuclei (FDR < 0.01) in the LOAD samples (Fig. 2c). Extending the analysis to the full nuclei set, without removing the data from lower quality nuclei (based on QC parameters, Additional file 2: Table S2), replicated the same observations and additionally showed a decrease in the proportion of excitatory neurons (FDR < 0.05, Fig. 2c), as expected in

(See figure on next page.)

Fig. 2 Proportions of LOAD and normal nuclei among cell types and subtype clusters. **a** and **b**, Uniform manifold approximation and projection (UMAP) dimensional reduction plots of **(a)** snRNA-seq and **(b)** snATAC-seq datasets split into LOAD and normal groups following integration and clustering of the total nuclei population. Cell types are labeled in top plots and cell subtype clusters are shown in center plots. Lower plots show distribution of donor samples within subtype clusters (see also Additional file 1: Figure S2). **c**, Box plots quantifying differences in proportion of nuclei from each donor belonging to each cell type (number of nuclei from a particular donor belonging to a particular cell type divided by total number of nuclei for the same donor) split by diagnosis, based on snRNA-seq dataset. As indicated, plots represent cell datasets following QC filtering and without QC filtering. Dots represent means for each sample set. Boxes represent interquartile range, with median values indicated via horizontal line. Mean values are indicated by open diamonds. Whiskers extend to sample means within 1.5 times the interquartile range. Asterisks represent statistical significance of variation between LOAD and normal mean values at FDR-adjusted p-value levels of FDR < 0.05 (*), FDR < 0.01 (**), and FDR < 0.001 (***) based on bootstrapped Wilcoxon test. **d**, Box plots quantifying differences in cell type proportions between LOAD and normal samples based on snATAC-seq dataset. **e** and **f**, Box plots quantifying differences in subcluster proportions between LOAD and normal samples based on **(e)** snRNA-seq and **(f)** snATAC-seq datasets

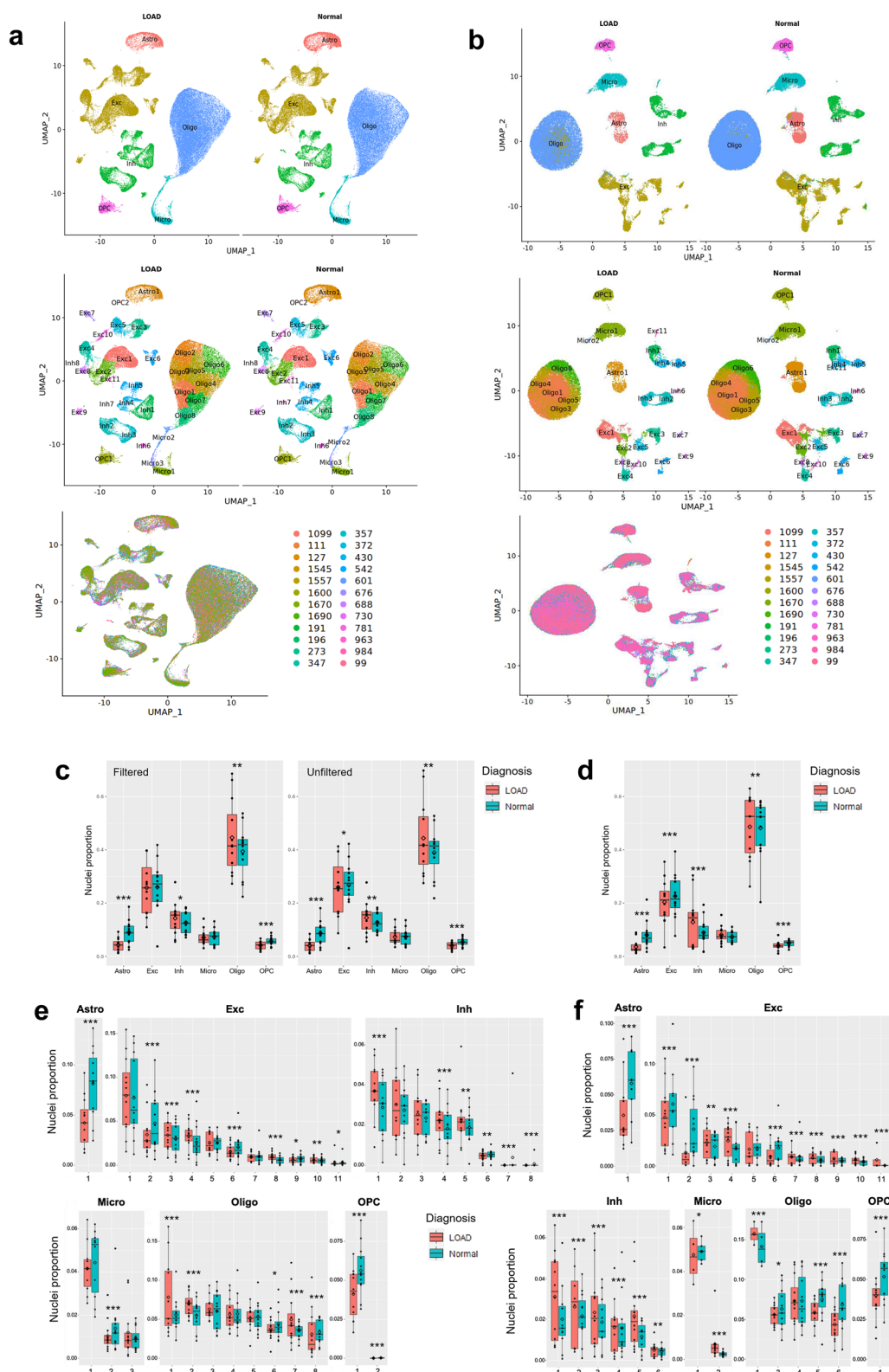


Fig. 2 (See legend on previous page.)

LOAD brains. Supporting these findings, all significant changes in the proportions of LOAD nuclei compared to controls were also detected in QC-filtered snATAC-seq nuclei (FDR < 0.01) (Fig. 2d, Additional file 2: Table S5).

We then performed the nuclei proportion comparison on the more granular level of cell subtypes. For the majority of the cell subtypes, the snRNA-seq and snATAC-seq QC-filtered nuclei demonstrated differences in proportion between LOAD and control, and the directionality of most differences were consistent between these datasets (Fig. 2e and f, respectively). However, we observed some discrepancies in directionality between the snRNA-seq and the snATAC-seq datasets for the smaller subclusters comprising less than 2% of total nuclei (e.g., Exc9, Inh6, Micro2), where even small deviations would be expected to have a greater impact on observed proportions. Several subclusters within particular cell-types showed the same directions of proportional differences between LOAD and control as those reported previously [43]. Furthermore, similarly to the previous observations [43], we found that within the same cell type some subclusters were significantly enriched in LOAD, whereas others were depleted (Fig. 2e and f). Nonetheless, the overall proportions of cell types and subtypes may not reflect true proportions in the brain due to some cell types not being equally amenable to the nuclei preparation procedure.

Cell type- and subtype-specific differential gene expression and chromatin accessibility in LOAD

We used the snRNA-seq and snATAC-seq datasets to characterize LOAD-associated dysregulated genes and changes in chromatin structure within each cell type and subtype. To examine differential gene expression using the snRNA-seq data, we utilized a mixed effects model that incorporates a random effect for donor to avoid pseudo-replication bias [52] (Methods), which is

typically not accounted for in single-cell studies. In addition to diagnosis, age, sex, post-mortem interval (PMI), sequencing saturation, and nuclei proportion within each cluster were incorporated into the model as fixed effects, based on our covariate analysis of 38 metadata variables (Methods, Additional file 1: Figure S3, Additional file 2: Table S2).

This analysis identified numerous differentially-expressed genes (DEGs) in LOAD for each cell type (ranging from 135 in OPCs to 429 in oligodendrocytes; Additional file 2: Table S7) and subtype (ranging from zero in subtype clusters Exc11, Inh7 and 8, and OPC2, to 968 in cluster Oligo4; Additional file 2: Table S8). The DEGs that showed the strongest (based on $|\log_2FC|$) and most significant (FDR-adjusted p -value) upregulation and downregulation effects in LOAD are highlighted in Fig. 3b–d and Additional file 1: Figure S4. In addition, we indicated the top DEGs that are within 500 kb (upstream or downstream) of a LOAD-GWAS tag SNP [15] (Fig. 3b–d, Additional file 1: Figure S4). Some of these DEGs were the most proximate to the tag SNP (e.g., *APOE* in Micro1 and *WWOX* in Oligo4), however, many were more distal within the tag SNP ± 500 kb window (Fig. 3c and d). For example, while *EPHA1* is the nearest gene to LOAD-GWAS SNP rs10808026, *TCAF1* was identified as a DEG in Exc1 within this LOAD region (Fig. 3b).

The differential expression analysis at the cell type level found that the majority of the strongest DEGs did not co-occur across cell types, supporting the need for cell type specificity in genomic analyses of LOAD. The common DEGs showed more subtle changes, and few had opposite directionality between cell types (Fig. 3e).

We used the upregulated and downregulated DEGs in LOAD from each cell type to perform gene ontology (GO) analysis, revealing enrichment of biological processes, cellular components, and molecular functions,

(See figure on next page.)

Fig. 3 Top differentially-expressed genes (DEGs) upregulated and down-regulated in LOAD by cluster and cell type. **a**, UMAP dimensional reduction plot of 33 cell subtype clusters based on snRNA-seq data. Clusters highlighted in panels b–d are circled in red. **b–d**, Unbiased volcano plots for six example clusters representing excitatory neuron (Exc), microglia (Micro), and oligodendrocyte (Oligo) cell types. \log_2 fold change (FC) between LOAD and normal control (NC) samples is plotted against $-\log_{10} p$ -value (FDR). Points representing DEGs with statistically significant ($p < 0.05$) upregulation in LOAD are shown in green while DEGs with significant downregulation are shown in red. Genes without significantly differential expression are shown in blue. The proportion of DEGs to total genes examined is shown above each plot. The six DEGs with the highest absolute fold change ($\log_2FC > 0.2$) in the up- and downregulated categories are labeled in green and red, respectively. The top up- and downregulated DEGs within 500 kb of disease-associated SNPs previously identified in GWAS are labeled in teal and pink, respectively. For each of the labeled genes, dot plots are shown below representing their unscaled expression levels (color) and percent of cells expressing the gene (width) for LOAD and NC samples. **e**, Heatmap showing $\log_2 FC$ for the top three DEGs with the highest $\log_2 FC$ values > 0.2 in at least one cell type in upregulated (green) and down-regulated (blue) categories for each cell type examined, as well as the top three DEGs ($\log_2FC > 0.2$ in at least one cell type) within 500 kb of identified GWAS SNPs (labeled in purple). **f**, The top ten enriched GO terms for downregulated LOAD genes (FDR adjusted $p < 0.5$) for oligodendrocytes and downregulated LOAD genes for microglia. Statistical significance threshold of category enrichment (FDR < 0.05) indicated by vertical lines. **g**, Table and Venn diagrams comparing cell type level DEGs identified in this study to those of Morabito et al. [43] and Anderson et al. [44] Consensus DEGs were identified in 2 or more studies

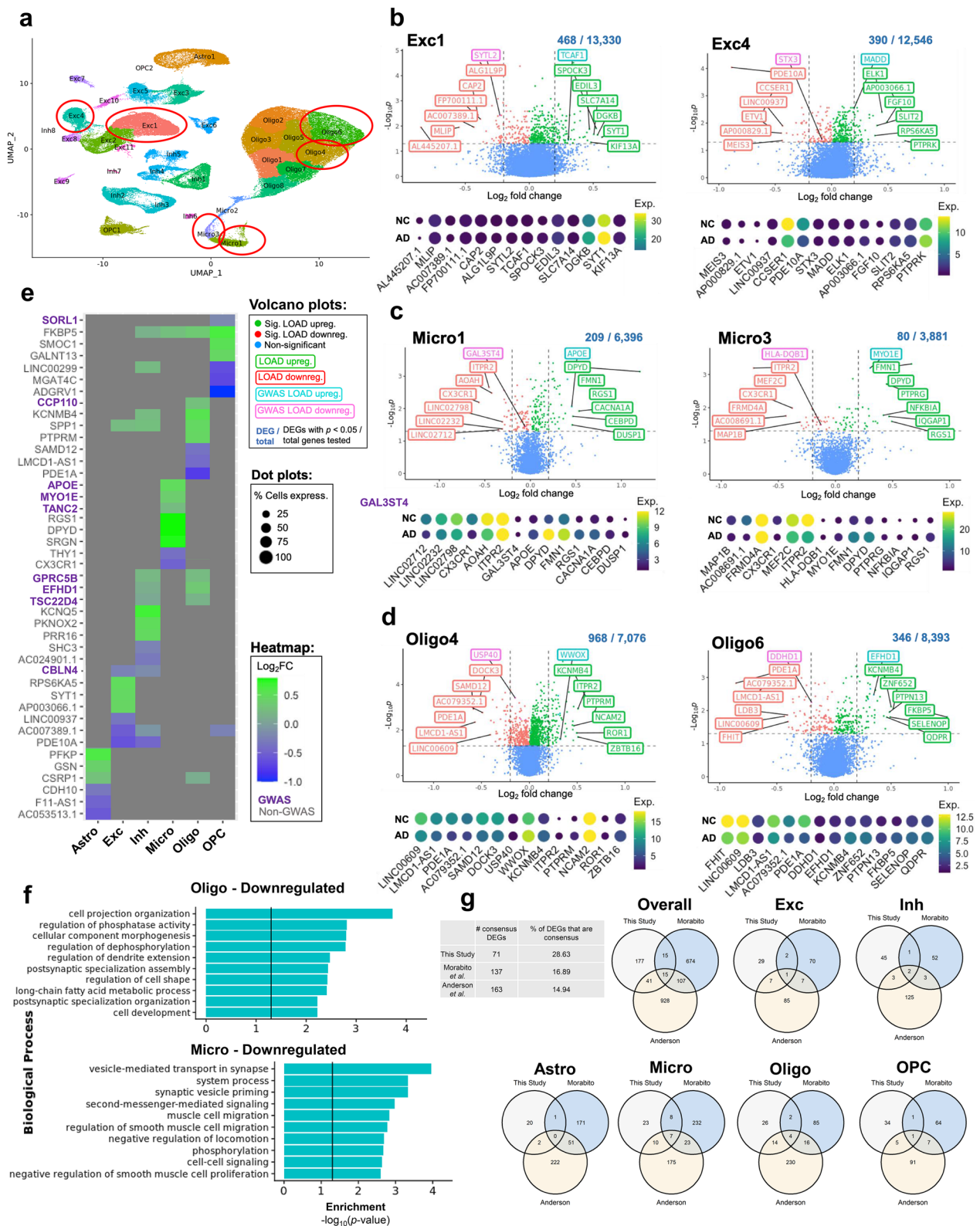


Fig. 3 (See legend on previous page.)

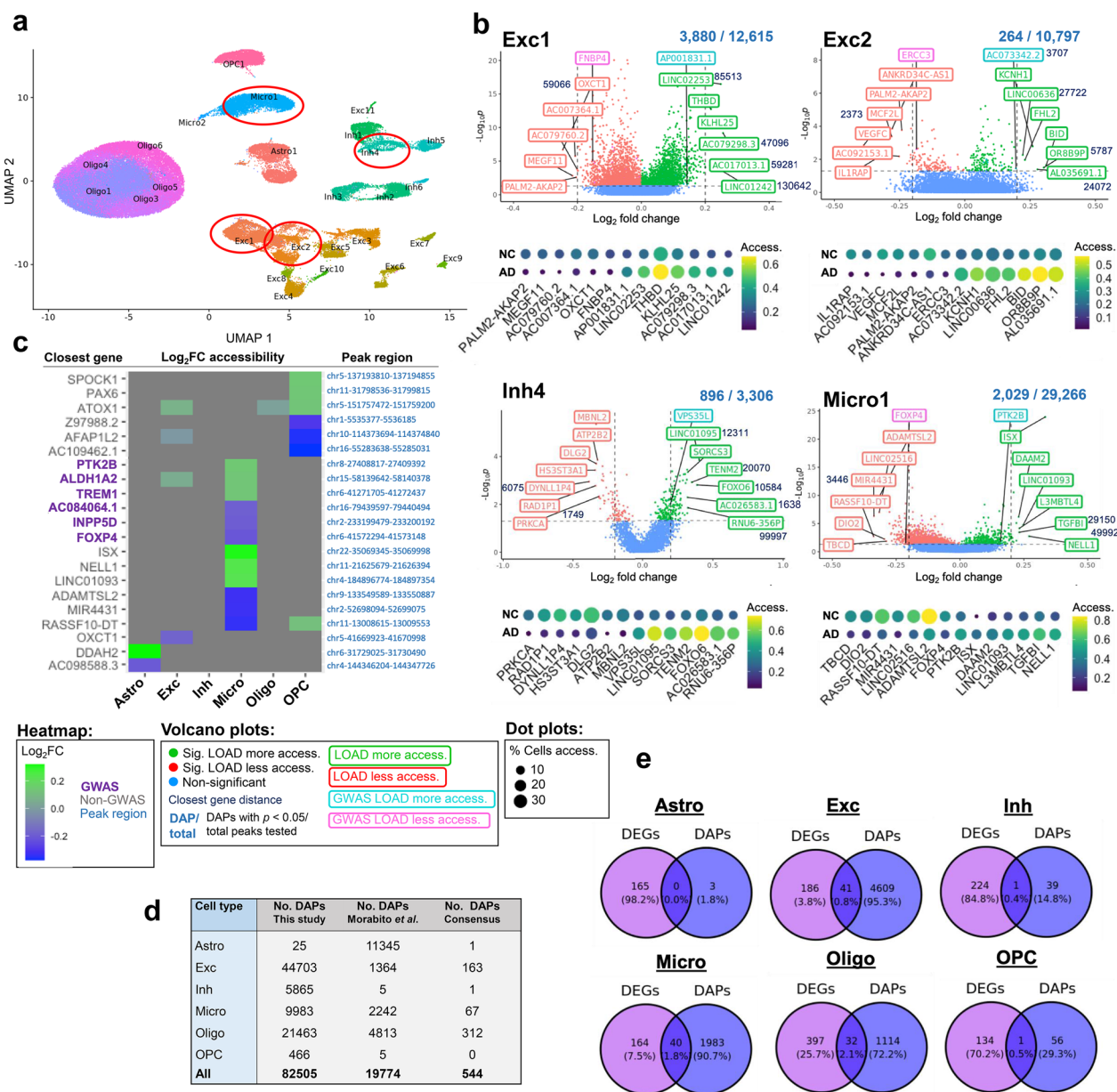


Fig. 4 Top differentially-accessible peaks (DAPs) upregulated and down-regulated in LOAD by cluster and cell type. **a**, UMAP dimensional reduction plot of snATAC-seq cell data indicating both cell type and subtype clusters. Clusters highlighted in panel b are circled in red. **b**, Unbiased volcano plots for four example clusters representing excitatory neuron, inhibitory neuron, and microglia cell types. Log₂ fold change (FC) between LOAD and normal control (NC) samples is plotted against—log₁₀ *p*-value (FDR). Points representing DAPs with statistically significant (*p* < 0.05) upregulation of accessibility in LOAD are shown in green while DAPs with significant downregulation are shown in red. Peaks without significantly differential accessibility in LOAD are shown in blue. The proportion of DAPs to total peaks examined is shown above each plot. The closest genes to the six DAPs with the highest absolute fold change (log₂ FC > 0.2) in the more and less accessible categories are labeled in green and red, respectively. The closest genes to the top more and less accessible DAPs within 500 kb of GWAS SNPs are labeled in teal and pink, respectively. Distances of closest genes from corresponding peaks (in bp) are indicated adjacent to gene labels where such distances are greater than zero. For each of the labeled peaks, dot plots are shown below representing their unscaled accessibility levels (color) and percent of cells with identified peak accessibility (width) for LOAD and NC samples. **c**, Heatmap showing log₂ FC for the top three DAPs (log₂ FC > 0.2 in at least one cell type) with the highest fold changes in upregulated (green) and down-regulated (blue) categories for each cell type examined, as well as the top three DAPs (log₂ FC > 0.2 in at least one cell type) within 500 kb of identified GWAS SNPs (labeled in purple). Closest gene names are indicated to the left of the heatmap while peak ranges are shown to the right. **d**, Table comparing cell type-level DAPs identified in this study to those of Morabito et al. [43] **e**, Venn diagrams showing overlap between DEGs identified via snRNA-seq and DAPs identified via snATAC-seq for each cell type examined

relevant to neuronal function and neurological disease processes in several cell types (Fig. 3f and Additional file 1: Figure S5). For example, the biological processes ‘cell projection organization,’ ‘regulation of dendrite extension’ and ‘postsynaptic specialization assembly’ were enriched for downregulated genes in LOAD for oligodendrocytes, and ‘vesicle mediated transport in synapse’ and ‘synaptic vesicle priming’ were enriched for downregulated genes in LOAD for microglia (Fig. 3f).

Next, we compared the cell type level DEGs identified in our dataset (overall 248 DEGs across all cell subtypes) to those reported by Morabito et al. [43] (811 DEGs) and Anderson et al. [44] (1,091 DEGs), (Fig. 3g). Genes identified in at least two studies within the same cell type and showing the same directionality of regulation were considered consensus DEGs (listed in Additional file 2: Table S9). Only 15 DEGs were identified as consensus DEGs across all three of the studies (*PDE10A* in Exc; *SHC3* and *PDE10A* in Inh; *PTPRG*, *MEF2C*, *FMN1*, *EPS15*, *DPYD*, *CX3CR1* and *ACSL1* in Micro; *PTPN13*, *LDB3*, *GLDN*, and *FKBP5* in Oligo; and *CNTN3* in OPC). Our study identified the fewest DEGs per cell type, possibly due to the regression of a greater number of covariates in our mixed effects model. Nonetheless, our results demonstrated the highest proportion (28.63%) of consensus DEGs, providing support for the reproducibility of our findings. Furthermore, the common cell subtype-specific DEGs between our study and Morabito et al. [43] (*i.e.*, cell type and directionality) included known LOAD-risk genes such as *APOE* in a microglial subtype (Micro1, $\log_2FC=0.454$, $FDR=0.008$), and *BIN1* in a subtype of oligodendrocyte (Oligo4, $\log_2FC=0.153$, $FDR=0.037$).

Using snATAC-seq, we profiled accessible chromatin regions (peaks) and catalogued cell type and subtype-specific differentially accessible peaks (DAPs). A differential analysis approach revealed many DAPs between LOAD and normal for each cell type (ranging from 25 in astrocytes to 44,703 in excitatory neurons) and subtype (ranging from zero in clusters Exc11, Inh7 and 8, Micro3, Oligo2, 7 and 8, and OPC2, to 53,661 in Exc1). We thus identified LOAD-associated DAPs at both cell type and cell subtype resolutions (Fig. 4a–c, Additional file 2: Tables S7 and S8). At the cell subtype level, we show four representative examples, and for each we indicate the top DAPs with the strongest (based on $|\log_2FC|$) and most significant (FDR-adjusted p -value) LOAD-associated effects on chromatin accessibility, as well as the top DAPs within GWAS loci [14] (tag SNP ± 500 kb; Fig. 4b). We also show the top DAPs ($|\log_2FC| < 0.2$) for all cell types and highlight those that overlap with GWAS loci (Fig. 4c). DAPs were labeled by distance to the nearest gene. Several DAPs, such as *ATOX1*, were identified in a few cell types, while others were unique for a specific cell type, such as the DAPs near the *PTK2B* and *INPP5D* genes in microglia (Fig. 4c). Next, we compared our results to the LOAD-associated cell type-specific DAPs reported by Morabito et al. [43] (see Methods) and demonstrated a total of 544 consensus DAPs for all cell types, with the highest number (312) found in oligodendrocytes (Fig. 4d). Of note, comparative analysis between the LOAD DEGs and DAP closest genes for each cell type found a relatively small degree (0.4–2.1%) of overlap between the two gene sets (Fig. 4e), possibly due in part

(See figure on next page.)

Fig. 5 Identification of cis co-accessible networks and associated DEGs in LOAD nuclei. **a** and **b**, Diagrams of example unidirectional CCANs—in which all overlapping DAPs and DEGs are regulated in the same direction in LOAD nuclei (e.g. more accessible peaks and higher gene expression) and mixed CCANs—in which at least one DAP/DEG pair are regulated in the same direction and at least one pair are regulated in opposite directions. Vertical lines at the top of each diagram indicate chromatin peaks and regulatory linkages of accessibility between peak pairs and associated co-accessibility scores are indicated by Bezier curves. Only CCAN-associated peaks are shown. DAPs with greater accessibility in LOAD are shown in salmon, while those with reduced accessibility in LOAD are shown in blue. Non-differentially-accessible peaks are shown in grey. Below the peak linkage plots, gene exons overlapping CCAN regions are depicted as arrows indicating the directionality of transcription. DEGs are labeled with gene names, and LOAD-upregulated DEGs are shown in salmon while non-differentially-expressed genes are shown in grey (no downregulated DEGs depicted). At the bottom of each diagram, a Manhattan plot is shown indicating $-\log_{10}$ of p -values for LOAD-association of chromosome loci within CCAN region as calculated by Kunkle et al. [7] Dotted lines in Manhattan plots indicate statistical significance threshold ($p=0.05$). Panel **a** depicts example CCANs that do not overlap GWAS-identified LOAD-associated SNPs, while panel **b** depicts example CCANs that do overlap LOAD-associated SNP loci, indicated via orange dots and labeled on Manhattan plots. **c**, Table showing total number of CCANs identified for each snATAC-seq nuclei cluster, as well as the number of unidirectional and mixed CCANs, and the mean number of DAPs and DEGs identified per CCAN for each cluster in both the unidirectional and mixed categories, along with the standard deviation (SD) and maximum and minimum values. **d**, Table listing DEGs linked to cCREs in unidirectional and mixed CCANs for each applicable cluster, as well as a functional annotation of specific DEGs with known associations to neurodegenerative disease. Number of cCREs linked to each DEG is shown in parentheses. **e**, Diagram of potential DEG/cCRE associations in which (i) one DAP is coaccessible with one peak overlapping the DEG, (ii) one DAP is coaccessible with multiple DEG peaks, (iii) two DAPs are coaccessible with one DEG peak, and (iv) two DAPs are coaccessible with two DEG peaks. **f**, Gene ontological analysis of biological processes for cCRE-linked DEGs associated with CCANs in the indicated cell subtype clusters. Up to the top ten significantly enriched biological processes involving a minimum of three DEGs are listed. Statistical significance of category enrichment ($p < 0.05$) indicated by vertical lines

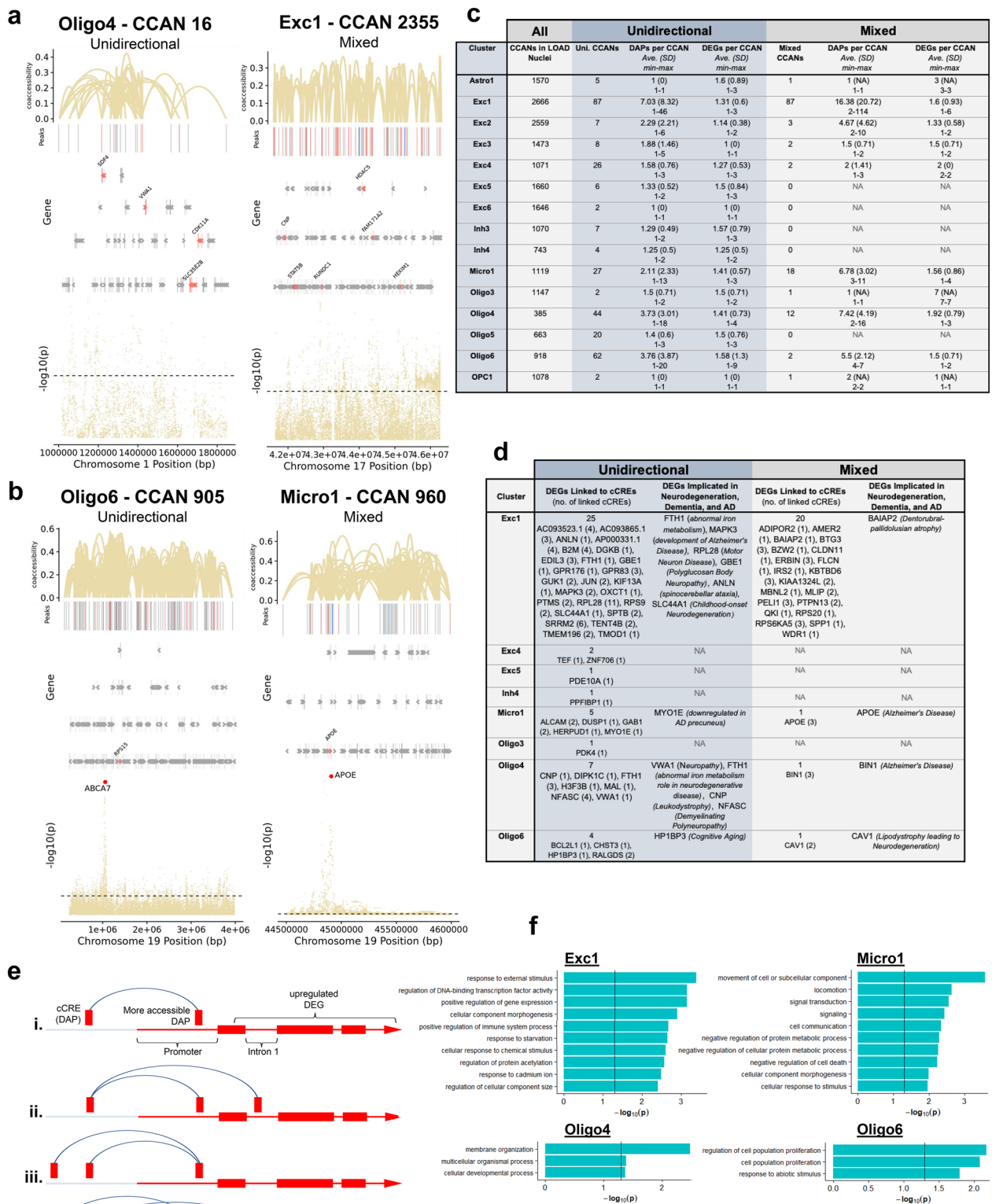


Fig. 5 (See legend on previous page.)

a

Cluster	Unidirectional			Mixed		
	No. CCANs	DEG TFs/ Enriched TFs	DEG (fdr<0.05, log ₂ FC>0.2) TF names [Fold enrichment]	No. CCANs	DEG TFs/ Enriched TFs	DEG (fdr<0.05, log ₂ FC>0.2) TF names [Fold enrichment]
Astro1	5	0	NA	1	0	NA
Exc1	87	7/83	FOSB::JUN [1.67], FOSL1::JUN [1.67], ELK1 [1.59], JUN::JUNB [1.57], FOSL2::JUN [1.49], FOS::JUN [1.46], JUN [1.43]	87	7/124	FOSL1::JUN [1.60], FOSB::JUN [1.57], FOSL2::JUN [1.56], JUN::JUNB [1.47], FOS::JUN [1.43], JUN [1.36], ELK1 [1.22]
Exc2	7	0	NA	3	0	NA
Exc3	8	0	NA	2	0	NA
Exc4	26	1/31	SMAD2::SMAD3::SMAD4 [1.22]	2	0	NA
Exc5	6	0	NA	0	0	NA
Exc6	2	0	NA	0	0	NA
Inh3	7	0	NA	0	0	NA
Inh4	4	0	NA	0	0	NA
Micro1	27	0	NA	18	0	NA
Oligo3	2	0	NA	1	0	NA
Oligo4	44	0	NA	12	0	NA
Oligo5	20	0	NA	0	0	NA
Oligo6	62	0	NA	2	0	NA
OPC1	2	0	NA	1	0	NA

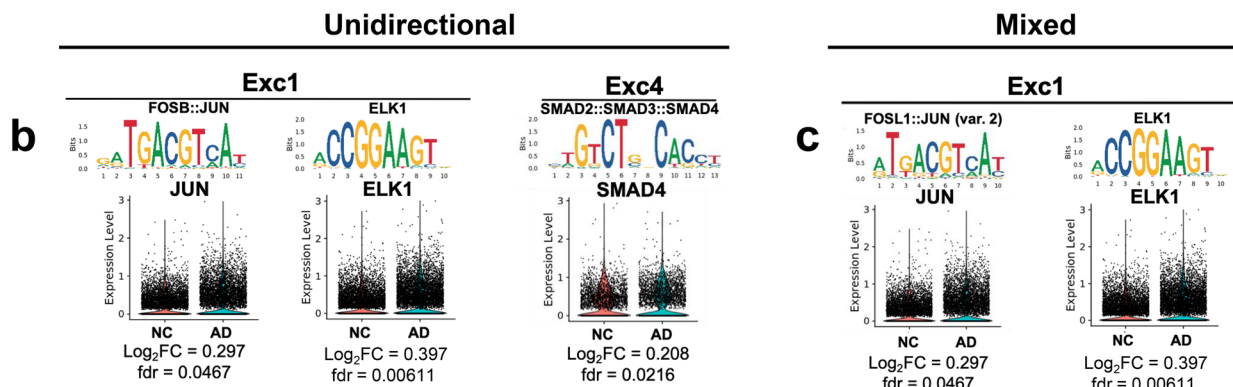


Fig. 6 Differential expression of transcription factors with motifs enriched in cis co-accessible networks. **a**, Table showing number of CCANs identified for each snATAC-seq cluster as well as the number of enriched transcription factor (TF) motifs corresponding to TFs identified as DEGs with $|\log_2FC| > 0.2$ in our snRNA-seq analysis within both unidirectional and mixed CCANs. The specific identified TFs and corresponding motif fold enrichment values are indicated in brackets where applicable. **b** and **c**, Sequence logos for indicated enriched motifs in unidirectional (**b**) and mixed (**c**) CCANs associated with labeled DEG TFs, along with violin plots of normalized snRNA-seq expression data split by normal control (NC) and LOAD groups for the corresponding TFs for each applicable cluster. Also indicated are adjusted *p*-values (FDR) and \log_2FC for TF gene expression data. Motifs are presented from left to right in decreasing order of enrichment for each cluster. Where multiple motifs corresponded to the same TF, the most highly enriched motif is shown

to the challenges in accurately assigning DAPs to target genes.

Cis co-accessibility network analysis in LOAD identified cell subtype-specific DEGs linked to cCREs

Next, we examined the crosstalk between cCREs and gene expression in LOAD TC. We defined a cCRE as a noncoding DAP that is co-accessible with one or more peaks overlapping the promoter/intron1 of a DEG. The small degree of overlap between DEGs and DAP closet genes noted above (Fig. 4e) suggested a complex interaction between LOAD-associated cCREs and target genes, and that cCREs may not simply regulate their nearest

genes. To address this complexity, we performed a multi omics integrative approach using our *parallel* snRNA-seq and snATAC-seq datasets to characterize the *cis*-regulatory landscape in LOAD and to identify target genes of LOAD-associated cCREs. To this end we used the snATAC-seq data from the LOAD nuclei only and utilized the Cicero algorithm [53] (Methods) to construct *cis*-co-accessibility networks (CCANs) for each cell subtype. We defined a total of 25,020 CCANs in LOAD TC across all 26 snATAC-seq cell subtypes and provided a detailed summary of their characteristics (Additional file 2: Table S10). To identify CCANs with likely specific relevance to LOAD, we further refined this set to

include only CCANs that contained at least one LOAD-associated DAP and DEG identified in our differential analyses of the snATAC-seq and snRNA-seq datasets, respectively (described above). This filtering criteria identified a subset of 518 LOAD CCANs in 15 cell subtypes that were categorized into 3 groups based on the directionality (\log_2FC) of the DAP-DEG pair: (i) unidirectional (309 CCANs), all more accessible/upregulated or less accessible/downregulated; (ii) mixed (129 CCANs), both directions for the DAPs and/or DEGs whereas at least one DAP-DEG pair showed the same effect direction; (iii) bidirectional (80 CCANs), all opposite directions (Figs. 1a, 5a–c, Additional file 2: Table S10). Further analyses focused on the unidirectional and mixed LOAD CCANs categories (Fig. 5), several of which (4 and 3, respectively, across Exc1, Micro1, Oligo4 and Oligo6 cell subtypes) also overlapped with LOAD-GWAS loci [14]. To identify the target genes of LOAD-associated cCREs in each cell subtype, we examined the co-accessibility of the peak overlapping the DEG promoter or intron 1 region with the distal DAP that defined the LOAD-associated cCRE. We catalogued the cCRE-linked genes for which their promoter/intron 1 peak was also a DAP and showed the same directionality as the cCRE-DAP. The analysis revealed 69 DEGs linked to cCRE in 8 of the cell subtypes with expression fold change ($|\log_2FC|$) greater than 0.15 (Fig. 5d, Additional file 2: Tables S10 and S11). 32 of the target DEGs were linked to 2–11 cCREs including, *APOE* and *BINI*, regulated by 3 and 2 cCREs, respectively (Fig. 5d, e, Additional file 2: Tables S10 and S11). On the other hand, no LOAD cCREs were linked to more than one DEG (Additional file 2: Table S11). A number of the identified cCRE-targeted DEGs have been implicated in neurodegenerative diseases, including genes involved in AD pathogenesis such as *MAPK3*, *APP*, and *FAM107B* in excitatory neuron cluster 1 (Exc1), *MYO1E* and *APOE* in microglial cluster 1 (Micro1), and *BINI* in oligodendrocyte cluster 4 (Oligo4) (Fig. 5d, Additional

file 2: Table S10). Collectively, our new strategy led to the identification of novel LOAD genes as well as validation of known disease genes and suggested the cell subtype in which they exert their pathogenic effects. In addition, we characterized the *cis*-regulatory elements and networks that govern the dysregulation of these genes in disease.

GO analysis of DEGs targeted by cCREs in each cell subtype revealed a few enriched biological processes (Fig. 5f, Additional file 1: Figure S6). For example, Exc1 showed enrichment for GO terms related to gene transcriptional regulation and response to external stressors. In Micro1, cCRE-linked DEGs involved in biological processes related to signal transduction, metabolic process and movement of cellular components were among the most enriched. Oligo4 and Oligo6 showed enrichment for GO terms involved in neural cell development, and cell proliferation, respectively. Due to the relatively small numbers of genes identified as LOAD associated cCRE-targeted DEGs specific to these Oligo subtypes, significant enrichment scores were found for only three GO terms (Fig. 5f, lower panels).

Cell subtype-specific transcription factors relevant to LOAD

We sought to explore the cell subtype-specific *trans* regulation involved in LOAD-associated changes in gene expression as a complementary approach to our *cis*-regulation analysis. To this end, we searched TFs that may interact with LOAD CCANs and thus contribute to dysregulation of genes in LOAD. The analysis focused on the subset of unidirectional and mixed LOAD CCANs identified in each of the 15 cell subtypes as described above (Fig. 5c) and was inclusive to all peaks (all accessible chromatin sequences, not limited to DAPs). The HOMER software package was used to identify enrichment of TF binding sites (TFBS) within the subset of LOAD CCANs for each cell subtype for TFs that were expressed in $\geq 10\%$ of the corresponding cell subtype (Fig. 6a, Additional file 2: Table S12). In this TFBS analysis we used the fold

(See figure on next page.)

Fig. 7 Identification of SNPs predicted to influence TF binding affinity at GWAS loci in LOAD CCANs. **a**, Summary tables of SNP-TFBS overlaps in unidirectional and mixed LOAD CCANs. **b–e**, Diagrams of specific example SNP-TFBS overlaps. The cell subtype, regulated DEG, TF and SNP ID are shown in bold. The log fold change (\log_2FC) and significance value (FDR) are shown for each DEG and corresponding cCRE. Additionally, functional information for each DEG is provided. The effect of the SNP on the TFBS affinity change and corresponding FDR determined using atSNP (see Methods) are noted. CCAN stacked plots show peak coaccessibility scores, directionality of changes in DAP accessibility in LOAD (red = increased accessibility, blue = reduced accessibility), and degree of LOAD association for GWAS loci. All features are arranged along the same horizontal axis to indicate chromosomal position. cCRE stacked plots are detailed from boxed area of CCAN plots and additionally indicate overlapped gene coding regions, with upregulated DEGs shown in red and downregulated DEGs shown in blue, as well as normalized chromatin accessibility of the genomic region in LOAD and Normal samples. TFBS activity stacked plots are detailed from boxed areas of cCRE plots and indicate aligned chromosomal positions of TFBSs (Reference and disrupted TFBS—dark and light gold horizontal bars, respectively—were determined based on position weight matrix as described in Methods) and SNPs (black lettering). TF Network plots illustrate potential regulatory networks between DEG-overlapping peaks (blue) and TFBS-overlapping peaks (green), with those linkages predicted to be affected by LOAD SNPs shown in red

a

Unidirectional							Mixed							
Cluster	GWAS CCANs	Significant SNP-TFBS Overlaps	Significant TFs	Significant TFBSs	Significant SNPs	Associated DEGs	Cluster	GWAS CCANs	Significant SNP-TFBS Overlaps	Significant TFs	Significant TFBSs	Significant SNPs	Associated DEGs	Associated GWAS SNPs
Exc1	4	26	25	26	17	CUTA, TAPBP, TFEB, TCAF1	Exc1	2	92	81	87	39	RPS15, MBD3, GNA11, FZR1, EFHD1	ABCA7, INPP5D
Micro1	1	39	35	39	4	MYD1E	Micro1	2	70	61	70	21	APOE, LST1	APOE, HLA-DRB1
Oligo4	2	11	11	11	2	SC5D	Oligo4	2	101	83	98	39	PTBP1, NDUJF57, BIN1	ABCA7, BIN1
Oligo6	3	5	4	5	3	RPS15								

b

Exc1, RPS15 // MAZ (rs114520071)

Structural component of ribosome

RPS15 (Promoter)

logFC = 0.19939, FDR = 0.024252

MAZ

Affinity change = -6.0638, FDR = 0.00065477

chr19-1354186-1356119

logFC = 0.098318, FDR = 7.4161e-06

rs114520071

0.000118 (gnomAD), 0.017258 (BRAVO)

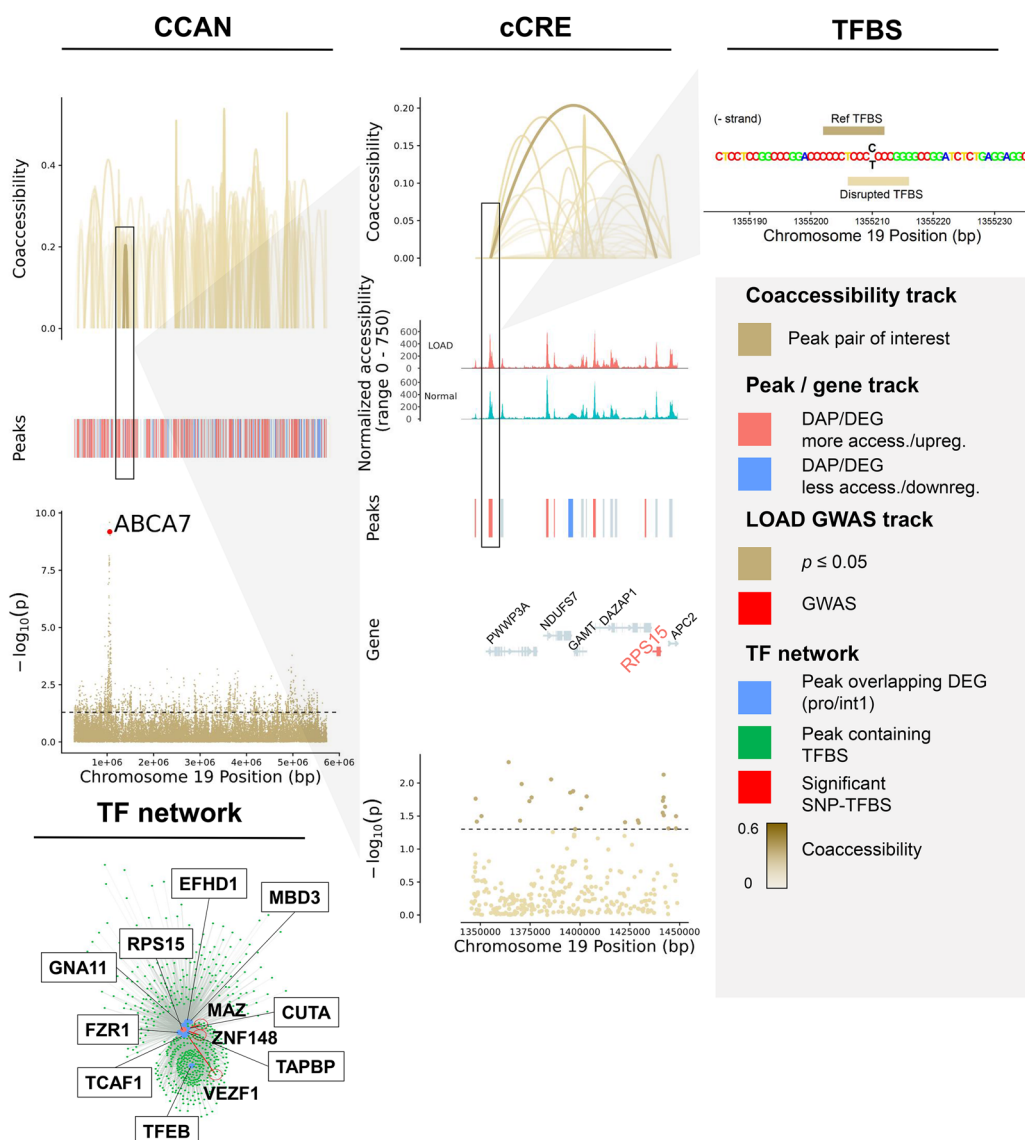


Fig. 7 (See legend on previous page.)

C

Micro1, MYO1E // NR2C2 (rs116657283)

Actin-based molecular motors. Involved in phosphatidylinositol binding and calmodulin binding

MYO1E (Promoter + Intron 1)
logFC = 0.40737, FDR = 0.0012459

NR2C2
Affinity change = 11.219, FDR = 0.0017085

chr15-59169011-59170652
logFC = 0.091773, FDR = 0.0060433

rs116657283
0.00008819 (gnomAD), 0.015176 (BRAVO)

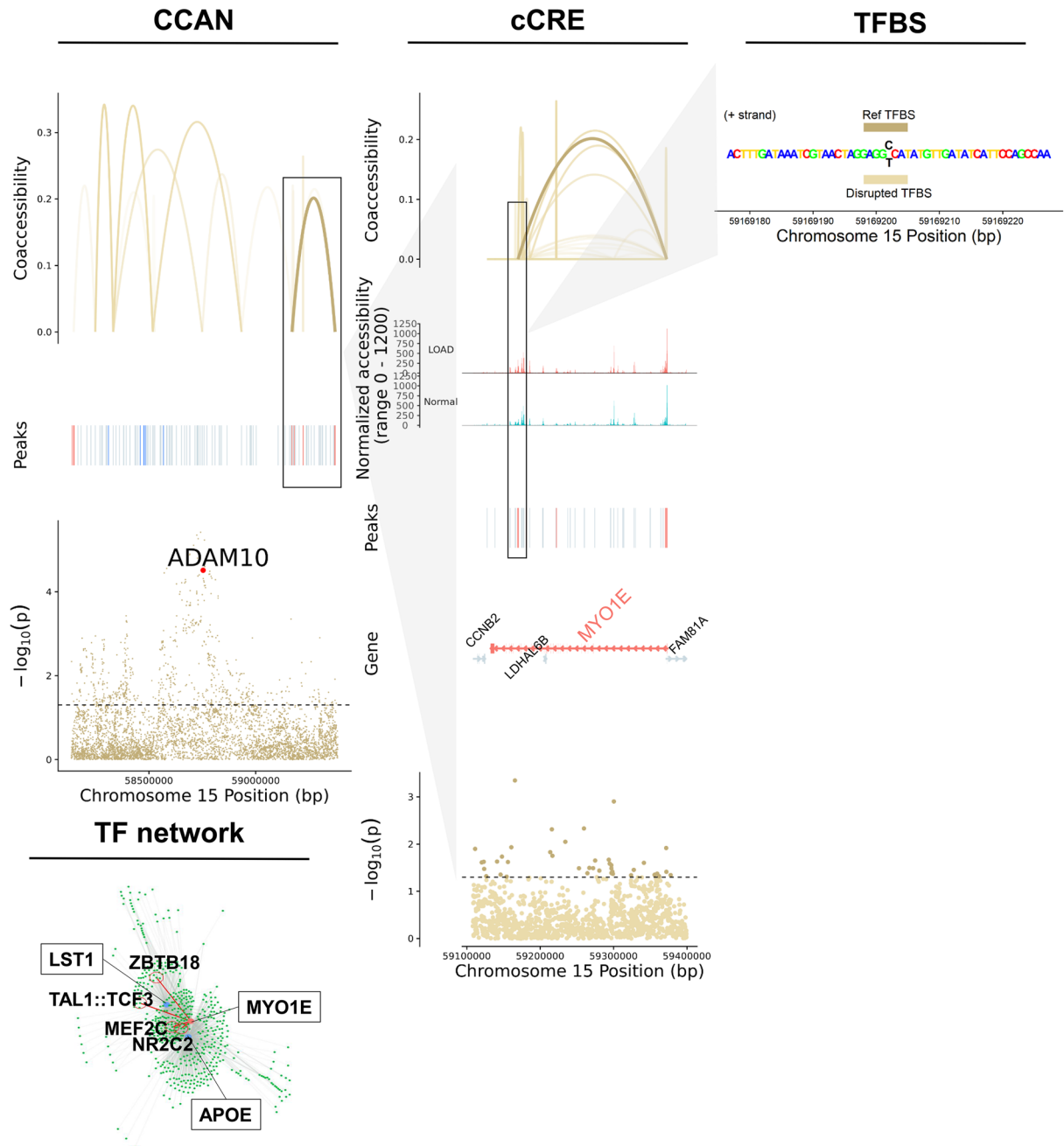


Fig. 7 continued

d

Micro1, APOE // SP1 (rs866159508)

Functions in lipoprotein lipid transport

APOE (Promoter + Intron 1)

$\log_{FC} = 0.45447$, $FDR = 0.0086049$

SP1

$Affinity\ change = -11.805$, $FDR = 0.0004227$

chr19-44845549-44846452

$\log_{FC} = 0.051984$, $FDR = 0.027969$

rs866159508

0.000207 (gnomAD), 0.021474 (BRAVO)

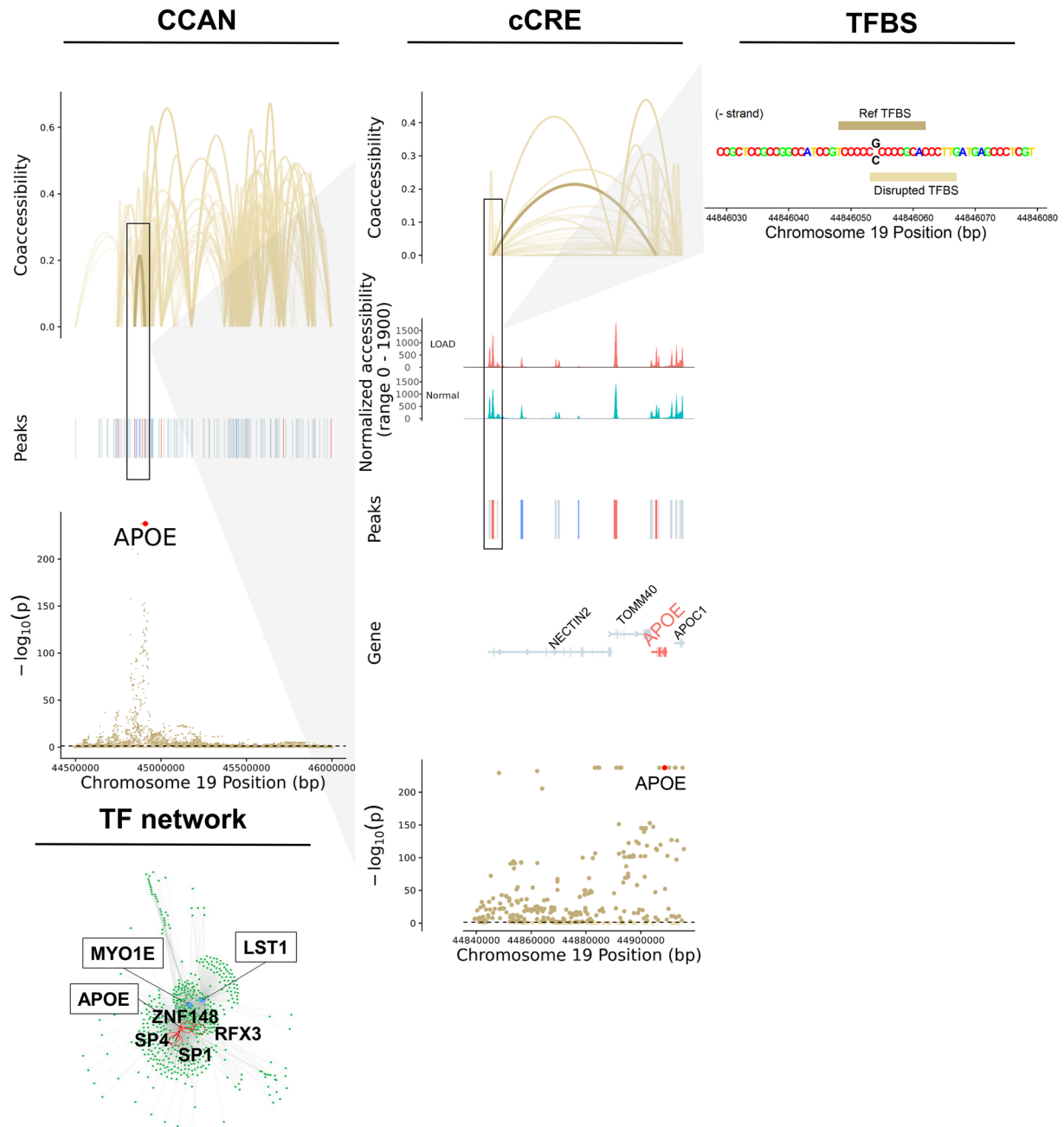


Fig. 7 continued

e

Oligo4, BIN1 // TFAP2E (rs13417187)

Regulates membrane activity. Involved in cell cycle regulation and apoptosis

BIN1 (Intron 1)

TFAP2E

logFC = 0.15359, *FDR* = 0.036918

Affinity change = 10.965, *FDR* = 0.0012483

chr2-127105247-127107901

rs13417187

logFC = 0.082583, *FDR* = 3.3712e-05

0.05649 (*gnomAD*), 0.088401 (*BRAVO*)

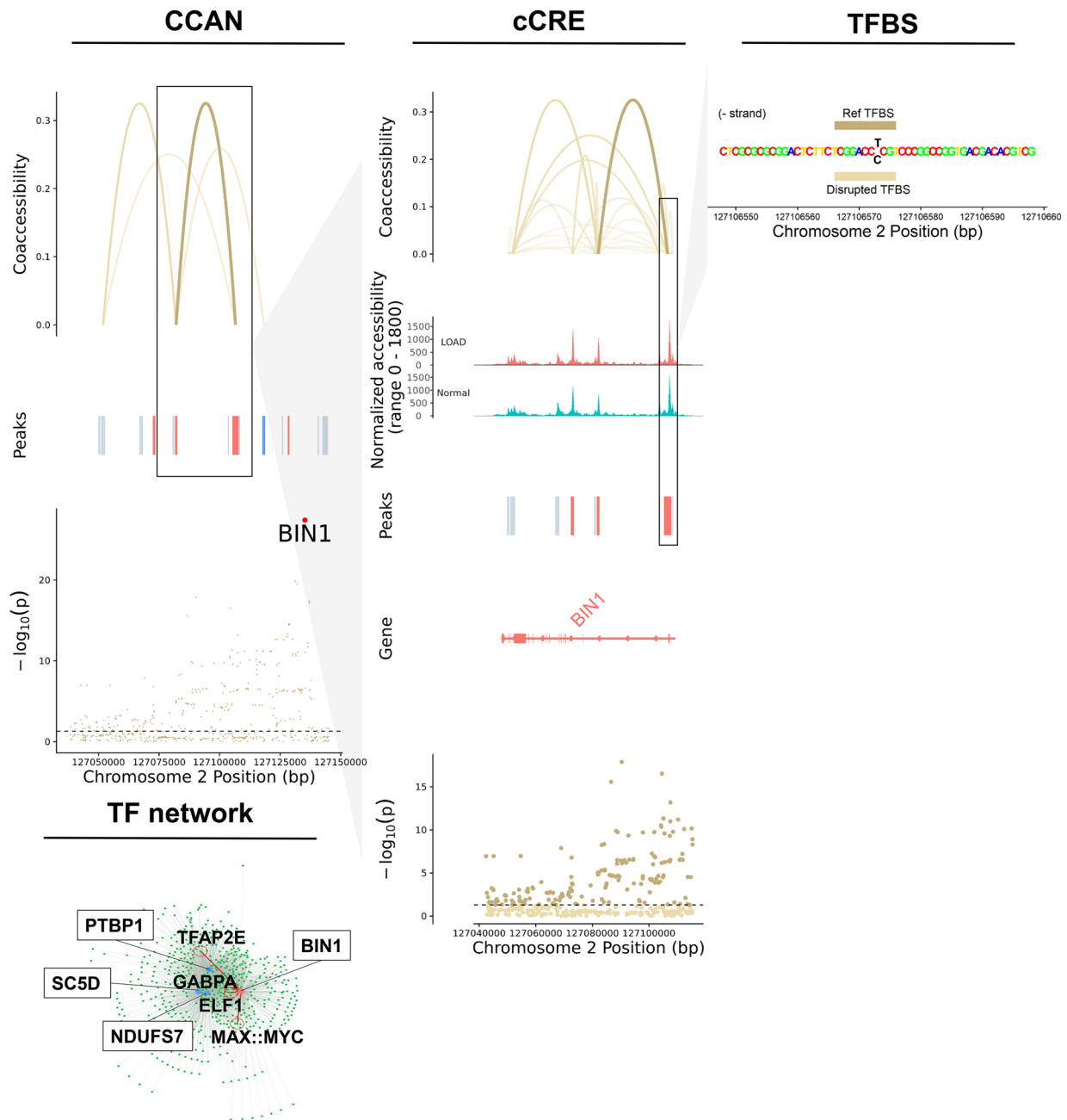


Fig. 7 continued

enrichment cutoff ≥ 1.2 and $FDR \leq 0.05$. The analysis revealed enrichment for 17–124 TFBSs in 7 out of the 15 analyzed cell subtypes (Additional file 2: Table S12). Interestingly, we found that several of these TFs were also LOAD-associated DEGs in five cell subtypes including, Exc1, Exc4, Micro1, Oligo4, and Oligo5 (Additional file 2: Table S12), out of which three TFs in Exc1 and 4 showed LOAD-associated increases in expression of $\log_2FC \geq 0.2$ (Fig. 6). *ELK1*—enriched in Exc1—showed the highest fold increase in expression in LOAD vs. normal cells ($\log_2FC = 0.397$, Fig. 6b–c). *ELK1* has been previously shown to initiate regionalized neuronal death and to associate with inclusions present in Alzheimer's disease, Lewy body disease, and Huntington's disease [54]. Other TF examples included *JUN* and its heterodimers with members of the *FOS* family [55] enriched in Exc1, and a member of *SMAD* in Exc4 (Fig. 6b–c). Recently, Anderson et al. [44] suggested *ZEB1* as a candidate master regulator in LOAD-specific regulatory networks in neurons. Consistently, we found that *ZEB1* binding motifs are significantly enriched within unidirectional CCANs of excitatory neuron cluster Exc1 (fold enrichment = 1.14, $FDR = 0.004$). However, the *ZEB1* results did not meet our criteria, that is to say it was expressed in less than 10% of Exc1 cells and its enrichment was below the cutoff threshold of 1.2. Altogether, these findings point to key TFs, regulatory elements, and *cis-trans* interactions potentially involved in dysregulating gene networks in specific cell subtypes contributing to LOAD pathogenesis.

Identification of SNPs potentially affecting TF binding and expression of DEGs in LOAD-GWAS regions

GWAS have previously identified numerous LOAD-associated SNPs, termed tagging SNPs. However, the majority of these tagging SNPs are based on disease association only, and the actual variants involved in disease risk have yet to be identified. We sought to prioritize candidate functional SNPs within GWAS regions that directly affect LOAD risk, specifically via transcriptional mechanisms. To this end we performed the analysis on 21 LOAD CCANs (unidirectional and mixed) that also overlapped LOAD-GWAS regions (Fig. 7a, Additional file 2: Table S13). We further focused the analysis on SNPs overlapping predicted TF binding sites for the minor or the major allele ($p < 1 \times 10^{-5}$) with $MAF \geq 1\%$. Next, we catalogued 125 SNPs that were predicted to change the affinity of 303 TF binding motifs ($FDR < 0.05$) using the Affinity Test for Identifying Regulatory SNPs (atSNP) R software package [56] (Fig. 7a, Additional file 1: Figure S7a, Additional file 2: Table S13). The minor alleles of the identified candidate regulatory SNPs resulted in gain or loss of TF binding sites in a cell subtype-specific

manner. These identified candidate regulatory SNPs mapped within LOAD-associated cCREs sequences linked to 16 DEGs across 4 cell subtypes (Exc1, Micro1, Oligo4 and 6, Fig. 7a, Additional file 2: Table S13). In the majority of the cCRE-linked DEGs (11 of 16), the peak overlapping the DEG promoter/intron 1 was also a DAP (Additional file 2: Table S13, examples in Fig. 7b–e and Additional file 1: Figure S7). These candidate regulatory SNPs were identified in specific cell subtypes. Many of the SNPs were identified in Exc1 (Fig. 7a, Additional file 2: Table S13), and presumably affect expression of genes including *CUTA*, a negative regulator of beta-amyloid generation [57], *TFEB*, a TF regulator of autophagic dysfunction associated with neurodegenerative pathology [58], *FZRI*, an adapter protein involved in cell cycle regulation that may suppress Cyclin B levels affecting AD-associated aberrant cell cycle re-entry [59], *GNA11*, associated with hypocalcemic hypercalcemia type II [60], *RPS15*, a structural ribosome component linked to Parkinson's disease [61] (Fig. 7b, Additional file 1: Figure S7a–b, Additional file 2: Tables S13 and S14), and *MBD3*, a neuropathy-associated chromatin remodeling complex component [62]. In Micro1, DEGs affected by the catalogued candidate regulatory SNPs included *MYO1E*, an actin-based molecular motor protein previously found to be differentially expressed in a microglial model of AD [63, 64] (Fig. 7c, Additional file 1: Figure S7c–f, Additional file 2: Tables S13 and S14) and *APOE* [14, 15] (Fig. 7d, Additional file 1: Figure S7g–i, Additional file 2: Tables S13 and S14). In Oligo4 we identified SNPs affecting *BINI* (Fig. 7e, Additional file 1: Figure S7j–l, Additional file 2: Tables S13 and S14), a major risk for LOAD suggested to have a role in regulating postsynaptic trafficking, and to accelerate beta-amyloid levels and tau accumulations in the context of LOAD [65–68]. Consistently, Morabito et al. [43] also suggested the potential cell type-specific *cis*-regulatory networks disrupted by causal disease variants in LOAD-GWAS risk loci including both *APOE* and *BINI* in the microglia and oligodendrocytes, respectively. Our data provided additional insights into the candidate regulatory SNPs and the *trans*-interaction with putative TFs. Other DEGs affected by the identified candidate SNPs included *NDUFS7*, a mitochondrial respiratory complex component associated with several neurological disorders [69], *PTBPI*, a regulator of neuronal pre-mRNA splicing associated with frontotemporal dementia and amyotrophic lateral sclerosis [70, 71], and *SC5D*, a cholesterol biosynthesis enzyme associated with lathosterolosis, a congenital disorder affecting central nervous system development [72], all in Oligo4, and *RPS15* [61] in Oligo6 (Additional file 2: Table S13).

Overall, many DEGs for which expression was predicted to be influenced by these identified SNPs were implicated in LOAD and other neurodegenerative disorders. Only 4 DEGs did not have previous evidence related to neurological phenotypes: *EFHD1*, *TAPBP* and *TCAF1* in Exc1, and *LST1* in Micro1. These findings support the biological relevance of this methodology in identifying cell subtype-specific variants with putative roles in LOAD pathogenesis. Noteworthy, many of the candidate regulatory SNPs changed the binding affinities of several TF motifs and each of the 16 cCRE-linked DEGs was associated with multiple SNP-TF interactions (Fig. 7a, Additional file 2: Table S13). However, some of these TFs may not be of biological relevance. Thus, the subsequent analysis was restricted to TFs that were expressed in $\geq 10\%$ of the corresponding cell subtype. In addition, to prioritize the top candidate regulatory SNPs we considered the affected genes and therefore narrowed down to cCRE-linked DEGs with $|\log_2FC| \geq 0.15$. After applying these criteria, the analysis revealed a total of 20 candidate regulatory SNPs and 5 candidate dysregulated genes including 10 SNPs in Micro1 that presumably exert their effects on *MYOIE* or *APOE*, 7 SNPs in Oligo4 potentially affecting *BIN1*, and 3 SNPs in Exc1 for *RPS15* or *TCAF1* (Fig. 7b-e, Additional file 1: Figure S7, Additional file 2: Table S14). Collectively, these outcomes provide a high priority list of candidate regulatory SNPs, TFs and target genes for experimental validation studies using model systems that facilitate investigations in the specific relevant brain cell subtype.

Discussion

In this study, we performed *parallel* snATAC-seq and snRNA-seq using brain samples from European ancestry subjects and identified LOAD-associated transcriptome and chromatin accessibility signatures and their crosstalk at a cell subtype-level resolution. To our knowledge, this is the first study that identified candidate noncoding regulatory variants and their interacting TFs in LOAD by integration of single-nuclei multi-omics datasets, and provided evidence suggesting that LOAD genetic variants exert their putative pathogenic effects in a brain cell subtype-specific manner.

Previously, we applied ATAC-seq on NeuN sorted nuclei and identified multiple LOAD specific neuronal (NeuN+) and non-neuronal (NeuN-) chromatin accessibility sites, several of which overlapped with LOAD-GWAS regions [39]. Furthermore, we identified sex-dependent LOAD changes in chromatin accessibility, particularly that glia-specific sites were found only in females. Last, by integrative analysis of the LOAD-GWAS regions, ATAC-seq on sorted nuclei, and snRNA-seq, we functionally characterized the impact of these

chromatin accessibility differences on gene expression within LOAD loci [73]. The current study extends our previous work in several ways. We now provide granular insight into molecular changes within cell subtypes. Furthermore, we uncovered the crosstalk between epigenetic, genomic, and transcriptomic determinants of LOAD pathogenesis. Our outcomes provide catalogues of candidate genes, cCREs, and variants involved in LOAD genetic etiology and the cell subtypes in which they act to exert their pathogenic effects.

To date, only a few other studies have profiled gene expression in cortex tissue of LOAD patients at single-cell resolution [41–44, 74–76], out of which only two studies have characterized both the transcriptome and chromatin accessibility of LOAD [43, 44]. These pioneering studies of snRNA-seq from cortexes of LOAD patients found that the strongest LOAD-associated changes appeared early in pathological progression and were highly cell type-specific [41], and identified LOAD-associated gene dysregulation in specific cell subpopulations, particularly for *APOE* and transcription factor genes [42]. One study used these datasets to examine the sex-dependent effects of LOAD and found disproportionate representation in disease-associated cell subtype populations as well as differential LOAD expression profiles between the sexes [41], while another characterized markers of selectively vulnerable neuron populations in AD [75]. Nonetheless, these studies focused only on snRNA-seq data. More recently, an integrative multi-omics framework of snRNA-seq and snATAC-seq in late-stage LOAD nuclei identified disease-associated cCREs and analyzed gene coexpression networks [43]. Another study examined gene expression and chromatin accessibility within the same nuclei and identified candidate transcription factors regulating LOAD-associated gene expression in neurons and microglia [44]. Altogether, our and others' studies have demonstrated the importance of cell type and subtype-specific omics profiling of human brain tissues to advance the understanding of the molecular and cellular subtype-specific pathways underlying LOAD. Data sharing across groups will not only provide accessibility to replication cohorts but will also further genetic exploration using meta-analysis in larger sample size to validate discoveries and test different hypotheses.

While GWAS infer genes based on proximity to the most strongly associated SNP, identification of the actual genes and variants involved in disease risk has been a challenge. We examined 1 Mb regions surrounding LOAD tagging SNPs and provided multiple examples for DEGs, *i.e.* candidate genes involved in LOAD, that mapped within the LOAD-associated region but more distal from the tag SNP. Additionally, we performed the first step towards identification of candidate regulatory

SNPs and their linked genes. Using single-nucleus genomics integrative analysis focused on LOAD GWAS loci, we prioritized several SNPs for validation studies. Notably, the associations of these candidate SNPs with LOAD risk remain to be determined. In future studies, a larger sample size may allow conduction of transcriptomic and chromatin accessibility QTL/eQTL mapping to determine colocalization with GWAS loci.

The importance of microglia in AD has been previously established [77–79]. Cells of this type may have a protective function in β -amyloid ($A\beta$) clearance via phagocytosis and protease degradation in the early stages of the disease [80, 81], but may also contribute to $A\beta$ plaque accumulation in later stages via seeding and proinflammatory cytokine production [82–85]. Microglia have also been shown to exacerbate neuronal synapse loss, neurotoxic inflammation, and tau pathology in AD [86–89]. Furthermore, a high number of identified AD risk genes have been found to be preferentially expressed in microglia [90]. Notably, 22% of AD GWAS loci-proximal genes identified by Bellenguez et al. [15] were also microglial signature genes. It was further shown via ChIP-seq and ATAC-seq that AD-associated non-coding variants are specifically enriched in microglial enhancer regions over those of other brain cell types [19]. Both *APOE* and *MYOIE* expression have been previously associated with disease-associated microglia exhibiting neurodegeneration-specific gene expression profiles in mouse models of AD [63]. Expression of human *APOE*e4 also resulted in increased tau pathology and inflammatory microglial response in mice, and increased production of TNF α by microglia in vitro [91]. *APOE* has been further implicated in conjunction with the microglia-associated receptor protein *TREM2* in driving the conversion of microglia to a neurodegenerative phenotype [92]. In this regard we would like to highlight our findings with *APOE* and *MYOIE*. We identified both *APOE* and *MYOIE* as key LOAD-associated genetic factors within microglia in several distinct analyses. First, these genes were the most strongly LOAD-upregulated GWAS DEGs in multiple microglial subtypes (Fig. 3c) as well as microglia overall (Fig. 3e), while genes associated with mediation of synaptic processes were strongly downregulated among microglia generally (Fig. 3f). Furthermore, *APOE* and *MYOIE* were also both linked to cCREs within LOAD CCANs of the Micro1 cluster (Fig. 5b, d), and moreover atSNP analysis revealed multiple SNPs predicted to impact TF binding affinity within cCREs potentially regulating both of these genes within this same cell subtype. Taken together, these findings serve to further underscore the importance of *APOE* and *MYOIE* as LOAD risk factors that exert their pathogenesis in microglia, and also suggest

potential molecular mechanisms for dysregulation in LOAD.

Over the last decade, LOAD GWAS have confirmed strong associations with the *APOE* LD genomic region, and no other LOAD-association remotely approached the same level of significance [8, 10, 14, 93–97]. However, whether the strongest signal is attributed to additional variants and haplotypes within this LD region jointly with e4, as well as the molecular mechanisms underlying the LOAD-association with the *APOE* LD region, is largely unknown. While *APOE*e4 is the first and strongest genetic risk factor for LOAD, accumulating evidence has suggested that the increased overall expression of *APOE* plays an important role in the etiology of LOAD (reviewed by Gottschalk et al. [98] and Yang et al. [99]). Foremost, previously we found significantly higher levels of *APOE*-mRNA in brain tissues obtained from e3/3 LOAD patients compared to e3/3 healthy donors, consistent with other reports showing elevated levels of *APOE*-mRNA in LOAD brains [21–23, 100]. In addition, new snRNA-seq datasets showed LOAD changes in *APOE* expression in glial cell types, in particular upregulation in microglial subpopulations [41, 42, 101]. Moreover, studies using *APP/PS1* transgenic mice showed that lowering the ApoE protein levels ameliorated cognitive dysfunctions and $A\beta$ pathology [102] independent of the *APOE* allele [103–105]. Lastly, studies showed LOAD-associated differential DNA methylation [36, 106–109], further supporting a role for dysregulation of *APOE* expression in the genetic etiology of LOAD. Here, to circumvent the confounding effect of *APOE*e4, we constrained the analysis to *APOE*e3 only. We observed both increased expression and increased chromatin accessibility of *APOE* among microglial populations. Additionally, we identified multiple cCREs linked to both the promoter and intron 1 sequences of *APOE* in the Micro1 subtype cluster, as well as multiple SNPs potentially influencing TF binding to the *APOE* promoter in this same cluster, suggesting a possible mechanism for microglial *APOE* dysregulation in LOAD. Our results thus provide further evidence that there are clear changes in *APOE* expression associated with LOAD and independent of the e4 allele [103], suggesting that regulation of *APOE* expression in specific cell subtypes may impact the risk to develop LOAD, making the modulation of the overall ApoE protein levels useful as a future therapeutic target.

Despite major advances in genome technology and our innovative experimental approach and analytical strategy, there are still a number of limitations associated with the analysis of single cell data. In this study, for cell type annotation and subtype clustering, variance stabilizing transformation (SCT) [110] of the snRNA-seq count data was employed. While SCT offers improvement over

log normalization by regularizing the effect of sequencing depth on transcript counts, it results in data too computationally intensive to use for differential expression analysis. Likewise, the term frequency-inverse document frequency (TF-IDF) normalization [111] used in snATAC-seq clustering addresses some of the pitfalls of log-normalization, but in some cases these corrections may be too harsh. A regularized regression framework may offer more reproducible results, but again would require prohibitively intensive processing power. These methods should be possible in future work with greater computational resources. Another limitation involves the accuracy of cluster linking. Thus, to test the accuracy we applied the pipeline developed for this manuscript using publicly available multiome data and found for each matching snATAC-seq and snRNA-seq cluster an overall 75.6% Jaccard index, where the most accurate cluster in the test dataset had a Jaccard index of 0.95, while the least accurate cluster had a Jaccard index of 0.32 (Methods, Additional file 2: Table S6). Despite this limitation, in all clusters, the cell type of the closest matching snRNA cluster was concordant with the initially predicted cell type of the snATAC cluster. Moreso, the intrinsic structure of the snATAC clusters was not lost. Thus, the downstream analysis remained unaffected while providing loosely informative insight towards epigenetic and transcriptomic interplay. Finally, previous research has estimated that a high proportion of predicted TF binding sites are false positives [112–114]. Raising the stringency of p -value thresholds decreases the rate of false positive motif matches, but increases the rate of false negative matches even more strongly, thus limiting the value of this approach [115]. We attempted to address this limitation by restricting the analysis to TFs that were expressed in $\geq 10\%$ of the corresponding cell subtype, so that our findings were more likely of biological relevance. Currently, single cell methods such as ChIP-seq or ChIA-PET-seq [116] are not available to validate our findings experimentally. Similarly, chromatin confirmation capture (3C) or chromatin confirmation capture carbon copy (5C) datasets on temporal cortex or analogous tissue regions in a single cell-type resolution would provide the most direct evidence for the cCRE-DEG interactions identified in our study, as they become available.

Single cell analyses specifically for disease-involved heterogenous organs such as the brain are imperative. Molecular changes are cell subtype specific, meaning that a particular gene and/or variant may exert its effect on a certain cell subtype but can be neutral in another cell subtype. LOAD exemplifies a genetically complex disease affecting a heterogenous tissue and therefore requires granularity in research approaches to facilitate advancements and new discoveries in the field. Our study

has pioneered the powerful strategy of integration of cell type-specific multi-omics datasets collected from the same sample in parallel to describe *cis-trans* regulatory networks disrupted in LOAD, validate known LOAD loci, and identify new candidate genes. We have furthermore presented the most comprehensive interrogation to date of genetic variants potentially impacting gene regulatory networks in LOAD. Further investigations including sex and ancestry-stratified studies using integrative single cell multi-omics data will advance our understanding of the genetics underpinning LOAD in specific populations. Collectively, our findings provide a rich dataset for future mechanistic experiments, confirm known LOAD-GWAS loci while also identifying novel loci, and highlight the disease-relevant cell types and subtypes for follow-up validation studies in model disease systems and for development of future therapeutic interventions.

Conclusions

Profiling the chromatin accessibility and transcriptomic landscapes from the same pool of nuclei and at the same time is a well-controlled approach to facilitate multi-omic integrative analyses and advance new genetic discoveries in complex disorders such as LOAD. Our study has seven major findings for the field of LOAD genetics. First, we have generated cell subtype-specific profiles of LOAD-associated chromatin accessibility signatures and maps of LOAD cCREs. Second, we leveraged the LOAD-accessible peaks dataset to identify candidate co-accessible networks in each cell subtype. Third, we provided a catalogue of candidate LOAD-associated cell subtype-specific DEGs. Fourth, we identified TFs relevant to LOAD and the cell subtype in which they act. Fifth, we catalogued candidate SNPs involved in dysregulation of key genes in LOAD in a cell subtype-specific manner. Sixth, we have demonstrated that LOAD associations may not be interpreted by the most proximate gene. Seventh, we provided evidence that the genetics underpinning LOAD risk mediates its pathogenic effects in various glial and neuronal cell subtypes. Overall, these results suggest that cell subtype-specific *cis-trans* interactions between regulatory elements, noncoding variants and TFs, and the genes dysregulated by these networks contribute, at least in part, to the development of LOAD.

Methods

Human post-mortem brain tissue samples

Frozen human temporal cortex tissue of LOAD samples ($n=12$) and neurologically healthy control samples (Normal) ($n=12$) was obtained from the Kathleen Price Bryan Brain Bank (KPBBB) at Duke University. The demographics for this cohort are included in Table 1 and detailed in Additional file 2: Table S1. Clinical diagnosis of LOAD

was pathologically confirmed using Braak staging (AT8 immunostaining) and amyloid deposition assessment (4G8 immunostaining) for all LOAD samples. All tissue donors were Caucasians with the *APOE* e3/e3 genotype and Braak & Braak Stage III-V. The project was approved for exemption by the Duke University Health System Institutional Review Board. The methods described were conducted in accordance with the relevant guidelines and regulations.

Cohort statistics

For comparisons of demographic variables, R statistical programming language was used. Age and post-mortem interval (PMI) of female LOAD was compared to female Normal, and age and PMI of male LOAD was compared to male Normal. The Shapiro–Wilk test was used for normality, Bartlett’s test for equal variance of normally distributed data, and Levene’s test for equal variance of non-normally distributed data. If groups were normal and had equal variance, two sample *t*-tests assuming equal variances were used to determine differences between group means. If groups were not normal, a Mann–Whitney’s *U* test was run. No groups had unequal variances.

Nuclei isolation from post-mortem human brain tissue

The nuclei isolation procedure was based on previous studies [117, 118], but has been optimized for single-cell experiments. 100–200 mg of post-mortem human brain tissue samples (gray matter) was thawed over ice in Lysis Buffer (0.32 M Sucrose, 5 mM CaCl₂, 3 mM Magnesium Acetate, 0.1 mM EDTA, 10 mM Tris–HCl pH 7.4, 1 mM DTT, 0.1% Triton X-100) and homogenized using a 7 ml dounce tissue homogenizer (Corning) with pestle A. The homogenate was filtered through a 100 µm cell strainer, transferred to a 14×89 mm polypropylene ultracentrifuge tube, and carefully underlain with sucrose solution (1.8 M Sucrose, 3 mM Magnesium Acetate, 1 mM DTT, 10 mM Tris–HCl, pH 7.4). The nuclei were separated by ultracentrifugation at 4 °C at 107,000 RCF for 15 min. The supernatant was removed by aspiration, and the remaining nuclei were washed with 1 ml Nuclei Wash Buffer (10 mM Tris–HCl pH 7.4, 10 mM NaCl, 3 mM MgCl₂, 0.1% Tween-20, 1% BSA, 0.2 U/µL RNase Inhibitor) and incubated on ice for 5 min. The nuclei were gently resuspended, and 800 µL was transferred to a microcentrifuge tube designated for the 10X Genomics single-cell ATAC assay while 200 µL was transferred to a microcentrifuge tube designated for the 10X Genomics single-cell gene expression assay. The nuclei were centrifuged at 300 RCF for 5 min at 4 °C, and the supernatant was again aspirated. For the ATAC assay, the pellet was resuspended in Diluted Nuclei Buffer (10X Genomics). For the gene

expression assay, the pellet was resuspended in Wash and Resuspension Buffer (1X PBS, 1% BSA, 0.2 U/µL RNase Inhibitor). After a 1-min incubation on ice, the nuclei were filtered through a 35 µm strainer. Nuclei concentrations were determined using a Countess™ II Automated Cell Counter (ThermoFisher) and nuclei quality was assessed at 10X and 40X magnification using an Evos XL Core Cell Imager (ThermoFisher) prior to library construction.

Parallel snATAC-seq/snRNA-seq library preparation and sequencing

Single-nucleus (sn)ATAC-seq libraries were constructed using the Chromium Next GEM Single Cell ATAC Library and Gel Bead v1.1 kit, Chip H Single Cell kit, and Single Index Kit N Set A (10X Genomics) according to manufacturer’s instructions. In parallel, from the same pool of nuclei from each sample, single-nucleus (sn)RNA-seq libraries were constructed using the Chromium Next GEM Single Cell 3’ GEM, Library, and Gel Bead v3.1 kit, Chip G Single Cell kit, and i7 Multiplex kit (10X Genomics) according to manufacturer’s instructions. For each sample, 10,000 nuclei were targeted for both the ATAC and 3’ assays. Library quality control was performed on a Bioanalyzer (Agilent) with the High Sensitivity DNA Kit (Agilent) according to manufacturer’s instructions and the 10X Genomics protocols. Libraries were submitted to the Sequencing and Genomic Technologies Shared Resource at Duke University for quantification using the KAPA Library Quantification Kit for Illumina® Platforms and sequencing. Groups of four snATAC-seq libraries from 1 LOAD female, 1 LOAD male, 1 Normal female, and 1 Normal male were pooled on a NovaSeq 6000 S1 100 bp PE full flow cell to target a sequencing depth of 400 million reads per sample (Read 1N=50, i7 index=8, i5 index=16, and Read 2N=50 cycles). Groups of four snRNA-seq libraries from 1 LOAD female, 1 LOAD male, 1 Normal female, and 1 Normal male were pooled on a NovaSeq 6000 S1 50 bp PE full flow cell to target a sequencing depth of 400 million reads per sample (Read 1=28, i7 index=8, and Read 2=91 cycles). Sequencing was performed blinded to diagnosis, age, and sex.

snRNA-seq data processing

Raw sequencing data from snRNA-seq experiments was converted to fastq format, aligned to a GRCh38 pre-mRNA reference, filtered, and counted using Cell Ranger 4.0.0 (10X Genomics). Subsequent processing was done using Seurat 4.0.1 [119]. Filtered feature-barcode matrices were used to generate Seurat objects for the 24 samples. For quality control filtering, nuclei with less than 200 or greater than 10,000 features were excluded. Nuclei

with greater than 17.4% mitochondrial gene expression were found to cluster together on a uniform manifold approximation and projection (UMAP) feature plot and were also excluded. Because experiments were conducted in nuclei rather than cells, mitochondrial genes were subsequently removed. The 24 Seurat objects were merged into one, and were iteratively normalized using SCTransform [120] with glmGamPoi, which alleviates bias from lowly expressed genes [121]. Batch correction was performed using reference-based integration [45] on the 24 normalized datasets, which improves computational efficiency for integration.

Cell type annotation was conducted using a label transfer method [45] and a previously annotated reference dataset from human M1 (see below). Batch corrected data from both our dataset and the human M1 dataset were used for label transfer. Nuclei with maximum prediction scores of less than 0.5 were filtered out. Nuclei with a percent difference of less than 20% between their first and second highest cell type prediction scores were termed “hybrid” and subsequently removed [42]. Endothelial cells and VLMCs were in low abundance (465 total) and did not form distinct clusters in UMAP analysis and were therefore filtered out of the final dataset. After running a Principal Component Analysis (PCA), dimensionality was examined using an Elbow plot and by calculating the variance explained by each principal component (PC). UMAP analysis was then run with the first 30 PCs, and nuclei were clustered based on the UMAP reduction at a resolution of 0.1. Counts of predicted cell types based on the label transfer were examined for each of the 33 clusters (Additional file 2: Table S4), and clusters were manually annotated based on the majority cell type for each cluster (e.g., ‘Exc1’, ‘Exc2’, etc.). Cell type annotations were further validated by label transfer from snRNA-seq data reported by Morabito et al. [43] processed using the same methodology described below for human M1 primary motor cortex reference data, which did not alter cluster annotations (Additional file 1: Figure S1b).

Human M1 reference data processing

A recent snRNA-seq study on human primary motor cortex (M1) used 10X Genomics technology to characterize 127 transcriptomic cell types [40]. To optimize label transfer, we re-processed the data to map it to GRCh38 Ensembl 80 as we did with our data. Fastq files were obtained from the Neuroscience Multi-omic Data Archive (NeMO: <https://nemoarchive.org/>) and were aligned to the same GRCh38 pre-mRNA reference used for our data, filtered, and counted using CellRanger 4.0.0 (10X Genomics). Filtered feature-barcode matrices were used to generate 14 Seurat objects, and nuclei

that were absent from the annotated metadata from the study were filtered out, leaving 76,519 nuclei in the final re-processed dataset. The Seurat objects were merged and iteratively normalized using SCTransform [120] with glmGamPoi. Batch correction was performed using reference-based integration [45] on the 14 normalized datasets. The 127 transcriptomic cell types were grouped into 8 broad cell types including astrocytes, endothelial cells, excitatory neurons, inhibitory neurons, microglia, oligodendrocytes, OPCs, and VLMCs.

Cell type proportion comparisons

To compare proportions of cell types and subtypes between LOAD and Normal control groups, nuclei of each cell type and subtype were counted and divided by the total nuclei for each sample. Then, a bootstrapped two-sided Wilcoxon rank-sum test was performed in R (v4.0.2) using the *wilcox.test* function with default parameters and Benjamini–Hochberg correction for multiple testing. 20% of nuclei were randomly selected from all samples under comparison in each of 30 iterations [43].

snATAC-seq data processing

DNA fragments acquired from our snATAC-seq experiments were sequenced and converted to fastq format, from which they were mapped to GENCODE’s human release 32 reference [122] and counted using CellRanger-ATAC 1.2.0 (10X Genomics). We screened the remaining nuclei using the following quality control metrics:

- a) Nucleosome signal: defined as the ratio of mononucleosome fragments (147 to 294 bp) to nucleosome free fragments (<147 bp). Nuclei having a nucleosome signal of greater than 4 were removed [123].
- b) Transcription start site (TSS) enrichment: the ratio of aggregated, normalized read signal centered around a reference set of TSS’s compared to the signal in the TSS flanking regions. Nuclei with a TSS enrichment score of less than 2 were removed [123].
- c) Percent reads in peaks: the proportion of fragments in the cell that map to peak regions. Cells with less than 15% of reads in peaks were removed [123].
- d) Total peak region fragments: cells with less than 1000 peak region fragments were discarded due to low sequencing depth. Additionally, cells in the upper 1% in each sample distribution were removed as a precaution against multiplets [123].
- e) Blacklist ratio: the proportion of fragments that map to sequences associated with technical artifacts. Cells with greater than 5% of fragments mapping to blacklisted regions were removed [124].

The remaining preprocessing steps were conducted using R packages Seurat 4.0.1 [45], Signac 1.3.0 [123], and Harmony 0.1.0 [125]. Latent semantic indexing (LSI) was used to create a low-rank approximation of the data [126]. The 24 datasets were term frequency-inverse document frequency (TF-IDF) normalized and aggregated to form a joint peak by cell count matrix. Then, singular value decomposition (SVD) was performed on the joint cell by peak dataset, after which the left singular vectors were standardized, representing the LSI components. Then, the correlation of each component with sequencing depth was measured. Due to high correlation with sequencing depth, the first dimension was removed from downstream analysis ($\rho=0.7$) [123]. The remaining LSI components were then adjusted with the function RunHarmony to remove batch effects before clustering the data. In alignment with snRNA dimensionality reduction methods, we used dimensions 2 through 30 for clustering and downstream analysis.

Clusters were constructed from the adjusted LSI embeddings of the integrated dataset using the Seurat functions *FindNeighbors* and *FindClusters*, with k -nearest neighbors set to 20, and a cluster resolution of 1. The data was then projected onto a 2D surface using the uniform manifold approximation (UMAP) algorithm and inspected to ensure ample cluster resolution.

Cell type annotation of snATAC-seq nuclei

Cell type annotation of the snATAC cells was done using the integrated snRNA dataset as a reference [45]. First, “gene activity” matrices were constructed from each snATAC sample by counting the fragments mapping to promoter regions (between 2000 bp upstream and 200 bp downstream of TSS) of each cell. After quantifying promoter region fragments, the matrices were log-normalized using Seurat.

The Seurat function *FindTransferAnchors* was used to annotate the snATAC data against our snRNA data, utilizing canonical correlation analysis as the reduction method. Seurat computes the cross correlation between variable features of snATAC and snRNA cells. After L2 normalization, the left and right singular vectors from the SVD of this matrix are taken as the canonical correlation vectors. Seurat then uses a mutual nearest neighbor approach to find anchors between the datasets, representing biologically similar cell states across modalities. For each cell, the weighted combination of the k -nearest anchors was used to calculate prediction scores for each of the major cell types. For each cell, the predicted cell identity was the cell type with the maximum prediction score. Nuclei with maximum prediction scores of less than 0.5 were filtered out. “Hybrid” nuclei were identified

using the same metric as for snRNA-seq data above, and subsequently removed.

Doublet/Multiplet detection in snRNA-seq and snATAC-seq data

“Heterotypic” multiplets (i.e. composed of different cell types) were removed from snRNA and snATAC data by considering the “hybrid score”. The score metric was originally used by Grubman et al. [42] to identify intermediate cell states, and defined as $(x_1 - x_2)/x_1$, where x_1 is the highest prediction score, and x_2 is the second highest prediction score. We reasoned that heterotypic multiplets would have a transcriptomic/epigenomic profile from multiple cell types and thus exhibit competing cell type prediction scores. “Homotypic” multiplets (i.e. composed of one cell type) were removed by considering the number of features detected in cells (snRNA multiplets) and the total number of fragments in peaks (snATAC multiplets). snRNA cells with > 10,000 features and snATAC cells with total fragments in peaks above the 95th percentile were removed. Methods for removing homotypic multiplets that are based on fragment/UMI-counts also help to filter out heterotypic multiplets.

Linking snATAC and snRNA datasets

The Seurat functions *FindTransferAnchors* and *TransferData* were used in a comparable manner to link snATAC clusters to snRNA clusters. As in cell type annotation, the anchors were used to transfer the snRNA cluster information to the snATAC cells. Each cell in the snATAC data was given 33 prediction scores, corresponding to each of the snRNA clusters. Directly clustering snATAC cells by using the snRNA cluster prediction scores did not fully align with the intrinsic structure of the snATAC data found from the initial Louvain clustering. As such, the original snATAC clusters were left unchanged, and the snRNA cluster prediction scores in each cell were summed across all nuclei belonging to the same snATAC cluster. The maximum prediction score in each snATAC cluster was used to designate a closest matching snRNA cluster. Lastly, the cell type of the linked snRNA cluster was matched against the original cell type of the snATAC cluster to ensure concordance.

To assess the accuracy of cluster linking, we used PBMC granulocyte multiome data, freely available on the 10X Genomics website. The snATAC and snRNA assays were processed separately using the same pipelines outlined in this manuscript (quality control, normalization, and community detection). Each snATAC cluster was designated a closest matching snRNA cluster by summing

prediction scores of each cell as described previously. For each matching snATAC and snRNA cluster pair, the Jaccard index of the cluster barcode compositions was calculated. Across all clusters the average Jaccard index was 0.756 (Additional file 2: Table S6). Additionally, the hybrid score for each snATAC cluster was calculated as described in cell type annotation, using the overall cluster prediction scores as input. Cluster hybrid scores provided a measure of the consensus amongst cells in each snATAC cluster. The overall cluster concordance hybrid score was 88.56% (Additional file 2: Table S6).

Peak calling

As different cell types exhibit epigenetic heterogeneity, it was necessary to perform peak calling on each cluster before differential accessibility analysis to reduce noise and computational burden. Peak regions were predicted empirically using the MACS2 algorithm on each cluster within each sample separately [127]. MACS2 assumes a dynamic Poisson distribution, with λ dependent on the read signal of the surrounding genomic region and is robust to technical and biological bias. Peaks were designated as regions having a p value of $\leq 10e-5$. To combine peaks into a consensus set for each cluster, Multi Sample Peak Calling (MSPC) software package was used [128], which employs Fisher's method to evaluate overlapping peaks across samples. Peaks that occurred in at least 2 samples, with an FDR of ≤ 0.05 from Fisher's combined probability test were used as the consensus set for downstream analysis.

Covariate selection for differential analyses

Prior to differential analysis, we assessed the potential impact of several technical variables from each of the snRNA-seq and snATAC-seq experiments separately such as number of nuclei, sequencing saturation, and reads mapped to the genome, as well as donor-level characteristics such as age, sex, and PMI. Several processing steps were taken prior to association testing on the snRNA-seq and snATAC-seq data separately. For each experiment, read counts were summed for all nuclei per donor sample, resulting in only one expression or accessibility peak value per donor sample per gene or chromatin peak, respectively. This down-coding was done for the covariate selection analysis to address the fact that all nuclei from each donor would have identical donor characteristics. Subsequently, genes with no expression or peaks with zero accessibility for $> 20\%$ of samples were removed, and all values were mean centered and scaled prior to analysis.

Principal component (PC) analysis was performed using `prcomp` in R for all genes and peaks passing our

pre-processing steps, separately. We then performed linear regression using `glm` in R of PCs explaining $> 10\%$ of the variability in global expression or chromatin accessibility on both nuclei- and donor-specific metadata variables to identify factors that should be included as covariates in differential analysis. Specifically, we selected the variable most associated (surpassing Bonferroni correction for multiple testing, $q < 0.05$) with PC1 (or alternatively, the PC explaining the most variability) and regressed all genes or peaks on the associated variable to obtain gene or peak residuals that are adjusted for its effect. We then performed PC analysis on the gene or peak residuals, and in an iterative process, repeating the above steps until no additional metadata variables were associated with global expression or chromatin accessibility ($q < 0.05$). For the snRNA-seq analysis, sex, age, PMI, sequencing saturation, and cluster proportions (calculated for each donor by dividing the number of nuclei of each cell type or subtype by the donor's total nuclei count) were selected as covariates for the differential analysis, and for the snATAC-seq analysis, sex, age, PMI, peak region fragments, cluster proportions, and percent fragments overlapping any targeted region were selected.

Differential expression analysis

To avoid pseudoreplication bias, we used MAST [129] with a random effect for donor, as in a recent publication [52]. For each cell type and cluster, raw counts from the snRNA-seq assay were $\log_2(x+1)$ transformed, and genes expressed in less than 10% of cells in either group (LOAD or Normal) were filtered out. For differential expression testing, a two-part hurdle model was run, consisting of a zero-inflated regression fitting a generalized linear mixed-effects model followed by a likelihood ratio test comparing the model with and without the group factor. The reference level was set to 'Normal' such that the results for \log_2FC coefficients would be positive if up-regulated in LOAD and negative if down-regulated in LOAD. Cellular detection rate (number of genes expressed) was calculated for each nucleus, centered, and scaled, and added to the model as a covariate to control for nucleus size. The proportion of nuclei for each cell type and cluster was calculated for each donor (e.g., for a given cell type, the number of nuclei for a donor divided by that donor's total nuclei count) and added to the model to control for sample-to-sample variation in cell type composition. P values were adjusted for FDR to correct for multiple comparisons. The percentage of nuclei expressing each gene was calculated for both groups and added to the results.

Differential accessibility analysis

Seurat's "LR" test was used for differential accessibility testing between LOAD and normal samples was performed on each cell type and each cluster [123, 130]. The model predicted diagnosis using a binomial regression, with peak region fragments, percent fragments overlapping any targeted region, nuclei proportion, age, sex, PMI, and fragment counts as the predictors. A likelihood ratio test was conducted to compare the null and experimental models, using fragment counts as the parameter of interest. Corresponding p values for each peak were adjusted for FDR on a per cluster basis. As with snRNA-seq data, positive \log_2 fold change corresponded to increased accessibility, and vice versa.

Cell-type-specific DEG/DAP comparative analysis across studies

To quantify the consensus DEGs across studies, for each study [43, 44] we counted (a) the total number of DEGs ($|\log_2\text{FC}| > 0.2$), (b) the total number of DEGs found in 1 other study and (c) the total number of DEGs found across all three studies. Consensus DEGs perturbed in different directions (opposite sign \log_2 fold change) were not counted as part of the consensus sets.

Additional metrics for consensus DAPs were needed because there was not a guaranteed 1 to 1 matching of DAPs between datasets. This is due to differences in peak calling across the datasets and the level of noise present in snATAC-seq data. To link DAPs across this study and Morabito et al. [43], for all pairs of overlapping peaks, we first considered the width of the intersecting region in base pairs. To limit size discrepancy between overlapping peaks, we also considered the jaccard similarity index. For any two overlapping peaks, the jaccard similarity was calculated as the width of the intersection divided by the width of the merged regions. Overlapping peaks with < 200 base pair overlap and a jaccard index of < 0.25 were removed. A DAP was deemed consensus if it had at least one overlap meeting the previous criteria, and the same sign \log_2 fold change with the overlapped DAP.

Assessment of cis co-accessibility networks

Next, we sought to characterize chromatin interactions in the data using the R package Cicero [53]. The Cicero pipeline was conducted on a per-cluster basis using LOAD cells only. We passed our integrated LSI embeddings to Cicero's bootstrap aggregation procedure, in which highly similar cells are aggregated by summing the raw counts in groups of 50 k-nearest neighbors. The fragment sums are then normalized to account for within-group sequencing depth. Cicero then uses a graphical LASSO to estimate the partial correlation structure of each peak with its neighboring peaks. A penalty term

dependent on the genomic distance between peak pairs is used in GLASSO, and the resulting regularized correlations derived from the precision matrix are termed "co-accessibility scores." We defined the maximum peak-peak distance, at which regularized correlations are assigned 0, as 500 Kbp. We then specified a minimum co-accessibility score of 0.2 before extracting *cis* co-accessibility networks (CCANs) from the resulting data using Louvain community detection.

LOAD cCRE's and GO analysis

To probe candidate *cis*-regulatory elements within CCANs, we first isolated CCANs that (a) contained ≥ 1 DAP and (b) contained ≥ 1 peak that overlaps the promoter or intron1 of a DEG in the closest matching snRNA cluster. An additional factor to consider was whether the direction (i.e., the sign of DAP and DEG \log_2 fold change) aligned within DAP-DEG pairs. This assessment was done on each CCAN individually. First, all DAPs in the CCAN were extracted. Second, all DEGs in which the promoter or intron 1 was located within a peak belonging to the CCAN were extracted. Third, all pairs of DAPs and DEGs were screened as to whether they were perturbed in the same direction. Based on this screening, the CCANs were then given the following designations:

- a) *Unidirectional*: Of all possible DAP-DEG pairs in the CCAN, all had the same sign for \log_2 fold change.
- b) *Mixed*: Of all possible DAP-DEG pairs in the CCAN, there was ≥ 1 pair that had the same sign for \log_2 fold change, but not all pairs had the same sign for \log_2 fold change.
- c) *Bidirectional*: of all possible DAP-DEG pairs in the CCAN, there were no pairs that had the same sign for \log_2 fold change.

As we were primarily interested in enhancer promoter interactions, only unidirectional and mixed CCANs were considered. Within unidirectional and mixed CCANs separately, we looked for DAP-DEG overlap peak pairs that met the following criteria:

- a) DAP was highly coaccessible (coaccessibility score of ≥ 0.2) with a peak overlapping the promoter or intron 1 of a DEG.
- b) The peak overlapping the promoter or intron 1 of a DEG was also a DAP.
- c) Both DAPs and the overlapping DEG had the same direction of effect.
- d) DEG \log_2 fold change ≥ 0.15 .

For each cluster, we pooled the overlapping DEGs from both unidirectional and mixed CCANs that met these criteria and performed GO analyses with the R package topGO [131]. Genes expressed in at least 10% of cells in their corresponding cluster were used as the background gene sets. We used Fisher's hypergeometric test statistic in evaluating enrichment of GO terms. Terms of interest were defined as having a $p \leq 0.05$ and mapping to at least 3 genes in the test gene set.

Motif detection and enrichment analysis

To detect transcription factor binding sites (TFBS) in the data, position weight matrices (PWMs) from JASPAR 2020 were used to scan the genome for motifs within the CCANs of each cluster [132, 133]. The p value threshold for a motif match was $5e-5$. For overlapping motif matches of the same transcription factor, only the highest scoring match was used. Given the high rate of type I error associated with in silico motif discovery algorithms, motif enrichment was also performed to rule out unlikely candidates. We used HOMER to detect motif enrichment in CCANs [134]. HOMER first quantifies the GC content and n -mer composition of both the background and target regions and applies weights to eliminate sequence bias before using a binomial test to compute enrichment p values. Peaks within CCANs were used as target sequences, whereas all other cluster-specific peaks were used as background regions. For downstream analysis, we took motifs that were enriched with $FDR \leq 0.05$ and fold enrichment ≥ 1.2 .

In silico identification of regulatory variants

The R package atSNP was used to quantify potential regulatory variants in silico [56]. atSNP takes as input a list of position probability matrices and a list of SNP loci and uses importance sampling to detect motif positions and to assess the significance of an observed TFBS affinity change when the SNP allele is introduced. To strengthen the reliability and to ensure the tested motifs were real TFBSs, we modified the atSNP pipeline as follows: Position weight matrices using a \log_2 probability ratio were created, assuming empirical base frequencies from the regions tested as background. Then, the regions were scanned, and TFMPvalue R package was used to generate an exact score threshold from which to detect TFBSs to be tested. Our input set of TFBSs and SNPs passed to atSNP was based on the following criteria:

Open chromatin region criteria:

- a) TFBS-SNP pair was located within a DAP.
- b) TFBS-SNP pair was located within a unidirectional or mixed CCAN.

- c) CCAN contained ≥ 1 peak that was within 500 Kbp from a GWAS SNP.
- d) DAP containing the TFBS was highly coaccessible (coaccessibility score of ≥ 0.2) with a peak overlapping the promoter or intron 1 of a DEG in the closest matching snRNA cluster.
- e) DAP and associated DEG had the same sign \log_2 fold change.

TFBS criteria:

- f) Motif p value at the TFBS was $\leq 5E-5$ (for major or minor allele sequence).
- g) TFBS overlapped ≥ 1 SNP.

SNP criteria:

- h) Overlapping SNP had a minor allele frequency $\geq 1\%$. For multiallelic SNPs, only alleles with frequency $\geq 1\%$ were included.

atSNP assumes the underlying nucleotide distribution follows a first order Markov model. The sequences in the proposal distribution are of length $2L-1$, and contain a subsequence of length L , which matches the motif in question. The sequences are weighted based on the affinity score of the matching subsequence and the expected affinity change resulting from a single nucleotide alteration at the SNP position, L . atSNP outputs a "1" in the event that a selected sequence has an expected affinity change greater than or equal to that which was observed, and a "0" otherwise. This number is multiplied by the sample weights, i.e. the null and proposal distributions' likelihood ratio. The mean value after N Monte Carlo samples is taken as the estimated p value, where N is determined by the length of the motif. In addition, we ran 1000 iterations of atSNP for every SNP-TFBS overlap to further reduce the variance estimate from the Monte-Carlo sampling procedure. We reported the mean p value, variance, minimum and maximum p value for each atSNP test across all 1000 iterations. In reporting key results, we controlled the FDR at 0.05 using the Benjamini-Hochberg procedure.

Our flagship results included in the main/supplemental figures passed the following additional criteria:

- a) Dysregulated TF showed expression in $\geq 10\%$ of cells in the corresponding snRNA cluster
- b) Target DEG associated with the regulatory SNP had \log fold change magnitude of ≥ 0.15
- c) $FDR \leq 0.01$

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13578-023-01120-5>.

Additional file 1: Figure S1. Confirmation of cluster annotation. a, Feature plots of log-normalized, corrected count data from SCTransform output showing cell type-specific markers for astrocytes (*SLC1A2*, *AQP4*, *GFAP*), neurons (*RBFOX3*), excitatory neurons (*SLC17A7*), inhibitory neurons (*GAD1*, *GAD2*, *SLC6A1*), microglia (*APBB1IP*, *C3*, *CD74*, *CSF1R*), oligodendrocytes (*MOBP*), and OPCs (*MEGF11*), and endothelial cell markers (*FLT1*, *CLDN5*) as negative controls. b, Chromatin accessibility tracks for gene promoter and coding regions for cell type markers for oligodendrocytes (*MOBP*), microglia (*C3*), astrocytes (*AQP4*), excitatory neurons (*SLC17A7*), OPCs (*MEGF11*), and inhibitory neurons (*GAD2*). c, UMAP plots showing cell type annotation based on snRNA-seq of human primary motor cortex [33] (left) and snRNA-seq of human prefrontal cortex [36] (right), d, UMAP plot of re-annotated snRNA-seq dataset with known excitatory and inhibitory neuron subtypes. e, Proportions of nuclei of each of the five subtypes of excitatory and inhibitory neurons for each sample. Average proportions for the 12 LOAD and 12 Normal samples are shown. **Figure S2.** Distribution of nuclei among cell subtype clusters by donor sample ID. a, UMAP dimensional reduction plots of cell subtype clusters for snRNA-seq dataset split by donor sample ID. b, UMAP plots of cell subtype clusters for snATAC-seq dataset split by donor sample ID. Cell subtype clusters are color coded. **Figure S3.** Correlation of metadata covariates in snRNA-seq and snATAC-seq data. **Figure S4.** Top differentially-expressed genes (DEGs) upregulated and down-regulated in LOAD by cluster. Unbiased volcano plots for all clusters containing DEGs not shown in Fig. 3, representing astrocyte (Astro), excitatory neuron (Exc), inhibitory neuron (Inh), microglia (Micro), oligodendrocyte (Oligo), and oligodendrocyte precursor (OPC) cell types. \log_2 fold change (FC) between LOAD and normal control samples is plotted against $-\log_{10}$ *p*-value (FDR). Points representing DEGs with statistically significant ($p < 0.05$) upregulation in LOAD are shown in green while DEGs with significant downregulation are shown in red. Genes without significantly differential expression are shown in blue. The proportion of DEGs to total genes examined is shown above each plot. The six DEGs with the highest absolute fold change (\log_2 FC > 0.2) in the up- and downregulated categories are labeled in green and red, respectively. The top up- and downregulated DEGs within 500 kb of disease-associated SNPs previously identified in GWAS are labeled in teal and pink, respectively. The Inh8 and Inh9 clusters did not contain DEGs and are not shown. **Figure S5.** Gene ontology analysis of DEGs identified in snRNA-seq data for cell types. Gene ontological analysis of biological processes, cellular components, and molecular functions for DEGs associated with the indicated cell types. Up to the top ten enriched terms involving a minimum of three DEGs are listed. Statistical significance threshold ($p < 0.05$) is indicated by vertical lines. **Figure S6.** Additional gene ontology analysis of cCRE-linked DEGs. Gene ontological analysis of biological processes, molecular functions and cellular components for cCRE-linked DEGs associated with CCANs in the indicated cell subtype clusters. Up to the top ten enriched terms involving a minimum of three DEGs are listed. Statistical significance threshold ($p < 0.05$) is indicated by vertical lines. **Figure S7.** Identification of SNPs predicted to influence TF binding affinity at GWAS loci in LOAD CCANs. a-l, Diagrams of specific example SNP-TFBS overlaps. The cell subtype, regulated DEG, TF and SNP ID are shown in bold. The log fold change (\log_2 FC) and significance value (FDR) are shown for each DEG and corresponding cCRE. Additionally, functional information for each DEG is provided. The effect of the SNP on the TFBS affinity change and corresponding FDR determined using atSNP (see Methods) are noted. CCAN stacked plots show peak coaccessibility scores, directionality of changes in DAP accessibility in LOAD (red = increased accessibility, blue = reduced accessibility), and degree of LOAD association for GWAS loci. All features are arranged along the same horizontal axis to indicate chromosomal position. cCRE stacked plots are detailed from boxed area of CCAN plots and additionally indicate overlapped gene coding regions, with upregulated DEGs shown in red and downregulated DEGs shown in blue, as well as normalized chromatin accessibility of the genomic region in LOAD and Normal samples. TFBS activity stacked plots are detailed from boxed areas of cCRE plots and indicate aligned

chromosomal positions of TFBSs (Ref and disrupted TFBS—dark and light gold horizontal bars, respectively—were determined based on position weight matrix as described in Methods) and SNPs (black lettering). TF Network plots illustrate potential regulatory networks between DEG-overlapping peaks (blue) and TFBS-overlapping peaks (green), with those linkages predicted to be affected by LOAD SNPs shown in red.

Additional file 2: Table S1. Demographic information. **Table S2.** snRNA-seq QC data. **Table S3.** Cluster markers. **Table S4.** Cluster cell types. **Table S5.** snATAC-seq QC data. **Table S6.** Jaccard index. **Table S7.** DAP/DEG Summary—Cell Type. **Table S8.** DAP/DEG Summary – Cluster. **Table S10.** CCAN Summary. **Table S11.** Linked cCREs + DEGs. **Table S12.** Motif Enrichment Summary. **Table S13.** atSNP Summary. **Table S14.** atSNP Results Annotated

Acknowledgements

We thank the Kathleen Price Bryan Brain Bank at Duke University (funded by NIA AG028377) for providing us with the brain tissues, and the Duke Sequencing and Genomic Technologies Shared Resource for sequencing. We thank Dr. M. Lutz for generously providing his advice and guidance on bioinformatics resources. This work used a high-performance computing facility partially supported by grant 2016-IDG-1013 ("HARDAC+: Reproducible HPC for Next-generation Genomics") from the North Carolina Biotechnology Center.

Author contributions

OC-F conceived of the presented idea. JG and OC-F acquired brain samples and designed the study sample. JC, DG and JB performed tissue processing, nuclei extraction, and the 10x experiments for snATAC-seq/snRNA-seq library preps. JG, DG and EKS performed all data analyses. MEG performed covariate analysis. CH performed DEG annotation. OC-F planned and supervised the work. GEC and AAK provided guidance. JG, DG, EKS, MEG, GEC, AAK, and OC-F discussed and interpreted the results. JG, DG and EKS prepared figures and tables. JG, EKS, DG and OC-F wrote the manuscript. DG, MEG, GEC, and AAK reviewed and edited the manuscript. All authors read and approved the final manuscript. OC-F obtained funding support.

Funding

This work was funded in part by the National Institutes of Health/National Institute on Aging (NIH/NIA) [R01 AG057522 and RF1 AG077695 to OC-F].

Availability of data and materials

The single-nucleus (sn)RNA-Sequencing and snATAC-Sequencing data are available at the Synapse data repository (<https://synapse.org>, ProjectSynID: syn50996869). Access will be granted upon request under controlled use conditions. In addition, the snRNA-seq and snATAC-seq raw and normalized count data generated in this study are available at the Duke Research Data Repository (<https://research.repository.duke.edu>). All computer code used for this study has been deposited to GitHub (<https://www.github.com>) and will be made available upon request.

Declarations

Ethics approval and consent to participate

The project was approved by the Duke Institutional Review Board (IRB). The study does not involve living human subjects. All samples were obtained from autopsies, and all are de-identified.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Division of Translational Brain Sciences, Department of Neurology, Duke University Medical Center, DUMC Box 2900, Durham, NC 27710, USA. ²Center for Genomic and Computational Biology, Duke University Medical Center, Durham, NC 27708, USA. ³Duke Molecular Physiology Institute, Duke University Medical Center, DUMC Box 104775, Durham, NC 27701, USA. ⁴Department

of Medicine, Duke University Medical Center, Durham, NC 27708, USA.

⁵Department of Pediatrics, Division of Medical Genetics, Duke University Medical Center, DUMC Box 3382, Durham, NC 27708, USA. ⁶Center for Advanced Genomic Technologies, Duke University Medical Center, Durham, NC 27708, USA.

Received: 26 April 2023 Accepted: 30 August 2023

Published online: 03 October 2023

References

- Lo MT, Kauppi K, Fan CC, Sanyal N, Reas ET, Sundar VS, Lee WC, Desikan RS, McEvoy LK, Chen CH, Alzheimer's Disease Genetics, C. Identification of genetic heterogeneity of Alzheimer's disease across age. *Neurobiol Aging*. 2019;84(243):e241-243. <https://doi.org/10.1016/j.neurobiolaging.2019.02.022>.
- Nacmias B, Bagnoli S, Piaceri I, Sorbi S. Genetic heterogeneity of Alzheimer's disease: embracing research partnerships. *J Alzheimers Dis*. 2018;62:903-11. <https://doi.org/10.3233/JAD-170570>.
- Lambert JC, Ramirez A, Grenier-Boley B, Bellenguez C. Step by step: towards a better understanding of the genetic architecture of Alzheimer's disease. *Mol Psychiatry*. 2023. <https://doi.org/10.1038/s41380-023-02076-1>.
- Corder EH, Saunders AM, Strittmatter WJ, Schmechel DE, Gaskell PC, Small GW, Roses AD, Haines JL, Pericak-Vance MA. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science* (New York, NY). 1993;261:921-3.
- Liu N, Zhang K, Zhao H. Haplotype-association analysis. *Adv Genet*. 2008;60:335-405. [https://doi.org/10.1016/S0065-2660\(07\)00414-2](https://doi.org/10.1016/S0065-2660(07)00414-2).
- Schmechel DE, Saunders AM, Strittmatter WJ, Crain BJ, Hulette CM, Joo SH, Pericak-Vance MA, Goldgaber D, Roses AD. Increased amyloid beta-peptide deposition in cerebral cortex as a consequence of apolipoprotein E genotype in late-onset Alzheimer disease. *Proc Natl Acad Sci U S A*. 1993;90:9649-53. <https://doi.org/10.1073/pnas.90.20.9649>.
- Saunders AM, Strittmatter WJ, Schmechel D, George-Hyslop PH, Pericak-Vance MA, Joo SH, Rosi BL, Gusella JF, Crapper-MacLachlan DR, Alberts MJ, et al. Association of apolipoprotein E allele epsilon 4 with late-onset familial and sporadic Alzheimer's disease. *Neurology*. 1993;43:1467-72. <https://doi.org/10.1212/wnl.43.8.1467>.
- Lambert J-C, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C, Jun G, DeStefano AL, Bis JC, Beecham GW, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet*. 2013;45:1452-8. <https://doi.org/10.1038/ng.2802>.
- Naj AC, Jun G, Beecham GW, Wang LS, Vardarajan BN, Buron J, Gallins PJ, Buxbaum JD, Jarvik GP, Crane PK, et al. Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. *Nat Genet*. 2011;43:436-41. <https://doi.org/10.1038/ng.801>.
- Harold D, Abraham R, Hollingworth P, Sims R, Gerrish A, Hamshere ML, Pahwa JS, Moskva V, Dowzell K, Williams A, et al. Genome-wide association study identifies variants at CLU and PICCALM associated with Alzheimer's disease. *Nat Genet*. 2009;41:1088-93. <https://doi.org/10.1038/ng.440>.
- Hollingworth P, Harold D, Sims R, Gerrish A, Lambert JC, Carrasquillo MM, Abraham R, Hamshere ML, Pahwa JS, Moskva V, et al. Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. *Nat Genet*. 2011;43:429-35. <https://doi.org/10.1038/ng.803>.
- Jansen I, Savage J, Watanabe K, Bryois J, Williams D, Steinberg S, Sealock J, Karlsson I, Hagg S, Athanasias L, et al. Genetic meta-analysis identifies 9 novel loci and functional pathways for Alzheimer's disease risk. *bioRxiv*. 2018. <https://doi.org/10.1101/258533>.
- Marioni R, Harris SE, McRae AF, Zhang Q, Hagenaars SP, Hill WD, Davies G, Ritchie CW, Gale C, Starr JM, et al. GWAS on family history of Alzheimer's disease. *bioRxiv*. 2018. <https://doi.org/10.1101/246223>.
- Kunkle BW, Grenier-Boley B, Sims R, Bis JC, Damotte V, Naj AC, Boland A, Vronskaya M, van der Lee SJ, Amlie-Wolf A, et al. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Abeta, tau, immunity and lipid processing. *Nat Genet*. 2019;51:414-30. <https://doi.org/10.1038/s41588-019-0358-2>.
- Bellenguez C, Küçükali F, Jansen I, Andrade V, Moreno-Grau S, Amin N, Naj AC, Grenier-Boley B, Campos-Martin R, Holmans PA, et al. New insights on the genetic etiology of Alzheimer's and related dementia. *medRxiv*. 2020. <https://doi.org/10.1101/2020.10.01.20200659>.
- Kunkle BW, Grenier-Boley B, Sims R, Bis JC, Damotte V, Naj AC, Boland A, Vronskaya M, van der Lee SJ, Amlie-Wolf A, et al. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Abeta, tau, immunity and lipid processing. *Nat Genet*. 2019;51:414-30. <https://doi.org/10.1038/s41588-019-0358-2>.
- Jansen IE, Savage JE, Watanabe K, Bryois J, Williams DM, Steinberg S, Sealock J, Karlsson IK, Hagg S, Athanasias L, et al. Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nat Genet*. 2019;51:404-13. <https://doi.org/10.1038/s41588-018-0311-9>.
- Kikuchi M, Hara N, Hasegawa M, Miyashita A, Kuwano R, Ikeuchi T, Nakaya A. Enhancer variants associated with Alzheimer's disease affect gene expression via chromatin looping. *BMC Med Genomics*. 2019;12:128. <https://doi.org/10.1186/s12920-019-0574-8>.
- Nott A, Holtman IR, Coufal NG, Schlachetzki JCM, Yu M, Hu R, Han CZ, Pena M, Xiao J, Wu Y, et al. Brain cell type-specific enhancer-promoter interactome maps and disease-risk association. *Science*. 2019;366:1134-9. <https://doi.org/10.1126/science.aay0793>.
- Lutz MW, Sprague D, Chiba-Falek O. Bioinformatics strategy to advance the interpretation of Alzheimer's disease GWAS discoveries: the roads from association to causation. *Alzheimers Dement*. 2019;15:1048-58. <https://doi.org/10.1016/j.jalz.2019.04.014>.
- Linnertz C, Anderson L, Gottschalk W, Crenshaw D, Lutz MW, Allen J, Saith S, Mihovilovic M, Burke JR, Welsh-Bohmer KA, et al. The cis-regulatory effect of an Alzheimer's disease-associated poly-T locus on expression of TOMM40 and apolipoprotein E genes. *Alzheimers Dement*. 2014;10:541-51. <https://doi.org/10.1016/j.jalz.2013.08.280>.
- Zarow C, Victoroff J. Increased apolipoprotein E mRNA in the hippocampus in Alzheimer disease and in rats after entorhinal cortex lesioning. *Exp Neurol*. 1998;149:79-86. <https://doi.org/10.1006/exnr.1997.6709>.
- Matsui T, Ingelsson M, Fukumoto H, Ramasamy K, Kowa H, Frosch MP, Irizarry MC, Hyman BT. Expression of APP pathway mRNAs and proteins in Alzheimer's disease. *Brain Res*. 2007;1161:116-23. <https://doi.org/10.1016/j.brainres.2007.05.050>.
- Zou F, Chai HS, Younkin CS, Allen M, Crook J, Pankratz VS, Carrasquillo MM, Rowley CN, Nair AA, Middha S, et al. Brain expression genome-wide association study (eGWAS) identifies human disease-associated variants. *PLoS Genet*. 2012;8:e1002707. <https://doi.org/10.1371/journal.pgen.1002707>.
- Karch CM, Jeng AT, Nowotny P, Cady J, Cruchaga C, Goate AM. Expression of novel Alzheimer's disease risk genes in control and Alzheimer's disease brains. *PLoS One*. 2012;7:e50976. <https://doi.org/10.1371/journal.pone.0050976>.
- Karch CM, Ezerskiy LA, Bertelsen S, Goate AM, Alzheimer's Disease Genetics C. Alzheimer's disease risk polymorphisms regulate gene expression in the ZCWPW1 and the CELF1 Loci. *PLoS One*. 2016;11:e0148717. <https://doi.org/10.1371/journal.pone.0148717>.
- Mills JD, Nalpathamkalam T, Jacobs HI, Janitz C, Merico D, Hu P, Janitz M. RNA-Seq analysis of the parietal cortex in Alzheimer's disease reveals alternatively spliced isoforms related to lipid metabolism. *Neurosci Lett*. 2013;536:90-5. <https://doi.org/10.1016/j.neulet.2012.12.042>.
- Zhao J, Zhu Y, Yang J, Li L, Wu H, De Jager PL, Jin P, Bennett DA. A genome-wide profiling of brain DNA hydroxymethylation in Alzheimer's disease. *Alzheimers Dement*. 2017;13:674-88. <https://doi.org/10.1016/j.jalz.2016.10.004>.
- De Jager PL, Srivastava G, Lunnon K, Burgess J, Schalkwyk LC, Yu L, Eaton ML, Keenan BT, Ernst J, McCabe C, et al. Alzheimer's disease: early alterations in brain DNA methylation at ANK1, BIN1, RHBDLF2 and other loci. *Nat Neurosci*. 2014;17:1156-63. <https://doi.org/10.1038/nn.3786>.
- Yu L, Chibnik LB, Srivastava GP, Pochet N, Yang J, Xu J, Kozubek J, Obholzer N, Leurgans SE, Schneider JA, et al. Association of Brain DNA methylation in SORL1, ABCA7, HLA-DRB5, SLC24A4, and BIN1 with pathological diagnosis of Alzheimer disease. *JAMA Neurol*. 2015;72:15-24. <https://doi.org/10.1001/jamaneurol.2014.3049>.
- Watson CT, Roussos P, Garg P, Ho DJ, Azam N, Katsel PL, Haroutunian V, Sharp AJ. Genome-wide DNA methylation profiling in the

- superior temporal gyrus reveals epigenetic signatures associated with Alzheimer's disease. *Genome Med.* 2016;8:5. <https://doi.org/10.1186/s13073-015-0258-8>.
32. Natívio R, Donahue G, Berson A, Lan Y, Amlie-Wolf A, Tuzer F, Toledo JB, Gosai SJ, Gregory BD, Torres C, et al. Dysregulation of the epigenetic landscape of normal aging in Alzheimer's disease. *Nat Neurosci.* 2018;21:497–505. <https://doi.org/10.1038/s41593-018-0101-9>.
 33. Dillman AA, Majounie E, Ding J, Gibbs JR, Hernandez D, Arepalli S, Traynor BJ, Singleton AB, Galter D, Cookson MR. Transcriptomic profiling of the human brain reveals that altered synaptic gene expression is associated with chronological aging. *Sci Rep.* 2017;7:16890. <https://doi.org/10.1038/s41598-017-17322-0>.
 34. Gasparoni G, Bultmann S, Lutsik P, Kraus TFJ, Sordon S, Vlcek J, Dietinger V, Steinmaurer M, Haider M, Mulholland CB, et al. DNA methylation analysis on purified neurons and glia dissects age and Alzheimer's disease-specific changes in the human cortex. *Epigenet Chromatin.* 2018;11:41. <https://doi.org/10.1186/s13072-018-0211-3>.
 35. Li P, Marshall L, Oh G, Jakubowski JL, Groot D, He Y, Wang T, Petronis A, Labrie V. Epigenetic dysregulation of enhancers in neurons is associated with Alzheimer's disease pathology and cognitive symptoms. *Nat Commun.* 2019;10:2246. <https://doi.org/10.1038/s41467-019-10101-7>.
 36. Tulloch J, Leong L, Thomson Z, Chen S, Lee EG, Keene CD, Millard SP, Yu CE. Glia-specific APOE epigenetic changes in the Alzheimer's disease brain. *Brain Res.* 2018;1698:179–86. <https://doi.org/10.1016/j.brainres.2018.08.006>.
 37. Srinivasan K, Friedman BA, Etxeberria A, Huntley MA, van der Brug MP, Foreman O, Paw JS, Modrusan Z, Beach TG, Serrano GE, Hansen DV. Alzheimer's patient microglia exhibit enhanced aging and unique transcriptional activation. *Cell Rep.* 2020;31:107843. <https://doi.org/10.1016/j.celrep.2020.107843>.
 38. de Paiva Lopes K, Snijders GJL, Humphrey J, Allan A, Sneebouer M, Navarro E, Schilder BM, Vialle RA, Parks M, Missall R, et al. Atlas of genetic effects in human microglia transcriptome across brain regions, aging and disease pathologies. *bioRxiv.* 2020. <https://doi.org/10.1101/2020.10.27.356113>.
 39. Barrera J, Song L, Gamache JE, Garrett ME, Safi A, Yun Y, Premasinghe I, Sprague D, Chipman D, Li J, et al. Sex dependent glial-specific changes in the chromatin accessibility landscape in late-onset Alzheimer's disease brains. *Mol Neurodegener.* 2021;16:58. <https://doi.org/10.1186/s13024-021-00481-0>.
 40. Bakken TE, Jorstad NL, Hu Q, Lake BB, Tian W, Kalmbach BE, Crow M, Hodge RD, Krienen FM, Sorensen SA, et al. Evolution of cellular diversity in primary motor cortex of human, marmoset monkey, and mouse. *bioRxiv.* 2020. <https://doi.org/10.1101/2020.03.31.016972>.
 41. Mathys H, Davila-Velderrain J, Peng Z, Gao F, Mohammadi S, Young JZ, Menon M, He L, Abdurrob F, Jiang X, et al. Single-cell transcriptomic analysis of Alzheimer's disease. *Nature.* 2019;570:332–7. <https://doi.org/10.1038/s41586-019-1195-2>.
 42. Grubman A, Chew G, Ouyang JF, Sun G, Choo XY, McLean C, Simmons RK, Buckberry S, Vargas-Landin DB, Poppe D, et al. A single-cell atlas of entorhinal cortex from individuals with Alzheimer's disease reveals cell-type-specific gene expression regulation. *Nat Neurosci.* 2019;22:2087–97. <https://doi.org/10.1038/s41593-019-0539-4>.
 43. Morabito S, Miyoshi E, Michael N, Shahin S, Martini AC, Head E, Silva J, Leavy K, Perez-Rosendahl M, Swarup V. Single-nucleus chromatin accessibility and transcriptomic characterization of Alzheimer's disease. *Nat Genet.* 2021;53:1143–55. <https://doi.org/10.1038/s41588-021-00894-z>.
 44. Anderson AG, Rogers BB, Loupe JM, Rodriguez-Nunez I, Roberts SC, White LM, Brazell JN, Bunney WE, Bunney BG, Watson SJ, et al. Single nucleus multiomics identifies ZEB1 and MAFB as candidate regulators of Alzheimer's disease-specific cis regulatory elements. *Cell Genomics.* 2023. <https://doi.org/10.1016/j.xgen.2023.100263>.
 45. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM 3rd, Hao Y, Stoeckius M, Smibert P, Satija R. Comprehensive integration of single-cell data. *Cell.* 2019;177:1888–902. <https://doi.org/10.1016/j.cell.2019.05.031>.
 46. Yang AC, Vest RT, Kern F. A human brain vascular atlas reveals diverse mediators of Alzheimer's risk. *Nature.* 2022;603:885–92.
 47. Sun N, Akay LA, Murdock MH, Park Y, Galiana-Melendez F, Bubnys A, Galani K, Mathys H, Jiang X, Ng AP, et al. Single-nucleus multi-region transcriptomic analysis of brain vasculature in Alzheimer's disease. *Nat Neurosci.* 2023;26:970–82. <https://doi.org/10.1038/s41593-023-01334-3>.
 48. Yuan M, Meyer T, Benkowitz C, Savanthrapadian S, Ansel-Bollepalli L, Foggetti A, Wulff P, Alcamí P, Elgueta C, Bartos M. Somatostatin-positive interneurons in the dentate gyrus of mice provide local- and long-range septal synaptic inhibition. *eLife.* 2017;6:e211105.
 49. Jinnó S, Klausberger T, Marton LF, Dalezios Y, Roberts JD, Fuentealba P, Bushong EA, Henze D, Buzsáki G, Somogyi P. Neuronal diversity in GABAergic long-range projections from the hippocampus. *J Neurosci.* 2007;27:8790–804.
 50. Eyre MD, Bartos M. Somatostatin-expressing interneurons form axonal projections to the contralateral hippocampus. *Front Neural Circuits.* 2019;13:56.
 51. Johnson TS, Xiang S, Dong T, Huang Z, Cheng M, Wang T, Yang K, Ni D, Huang K, Zhang J. Combinatorial analyses reveal cellular composition changes have different impacts on transcriptomic changes of cell type specific genes in Alzheimer's disease. *Sci Rep.* 2021;11:353. <https://doi.org/10.1038/s41598-020-79740-x>.
 52. Zimmerman KD, Espeland MA, Langefeld CD. A practical solution to pseudoreplication bias in single-cell studies. *Nat Commun.* 2021;12:738. <https://doi.org/10.1038/s41467-021-21038-1>.
 53. Pliner HA, Packer JS, McFaline-Figueroa JL, Cusanovich DA, Daza RM, Aghamirzaie D, Srivatsan S, Qiu X, Jackson D, Minkina A, et al. Cicero predicts cis-Regulatory DNA interactions from single-cell chromatin accessibility data. *Mol Cell.* 2018;71:858–871.e858.
 54. Sharma A, Callahan LM, Sul JY, Kim TK, Barrett L. A neurotoxic phosphoform of Elk-1 associates with inclusions from multiple neurodegenerative diseases. *PLoS ONE.* 2010;5:e9002.
 55. Abate C, Luk D, Curran T. Transcriptional regulation by Fos and Jun in vitro: interaction among multiple activator and regulatory domains. *Mol Cell Biol.* 1991;11:3624–32.
 56. Zuo C, Shin S, Keles S. atSNP: transcription factor binding affinity testing for regulatory SNP detection. *Bioinformatics.* 2015. <https://doi.org/10.1093/bioinformatics/btv328>.
 57. Hou P, Liu G, Zhao Y, Shi Z, Zheng Q, Bu G, Xu H, Zhang YW. Role of copper and the copper-related protein CUTA in mediating APP processing and A β generation. *Neurobiol Aging.* 2015;36:1310–5.
 58. Cortes CJ, La Spada AR. TFEB dysregulation as a driver of autophagy dysfunction in neurodegenerative disease: molecular mechanisms, cellular processes, and emerging therapeutic opportunities. *Neurobiol Dis.* 2019;122:83–93.
 59. Ishizawa J, Sugihara E, Kuninaka S, Mogushi K, Kojima K, Benton CB, Zhao R, Chachad D, Hashimoto N, Jacamo RO, et al. FZR1 loss increases sensitivity to DNA damage and consequently promotes murine and human B-cell acute leukemia. *Blood.* 2017;129:1958–68.
 60. Nesbit MA, Hannan FM, Howles SA, Babinsky VN, Head RA, Cranston T, Rust N, Hobbs MR, Heath HR, Thakker RV. Mutations affecting G-protein subunit $\alpha 11$ in hypercalcemia and hypocalcemia. *N Engl J Med.* 2013;368:2476–86.
 61. Martin I, Kim JW, Lee BD, Kang HC, Xu JC, Jia H, Stankowski J, Kim MS, Zhong J, Kumar M, et al. Ribosomal protein s15 phosphorylation mediates LRRK2 neurodegeneration in Parkinson's disease. *Cell.* 2014;157:472–85.
 62. Ishibashi T, Kajihara I, Mizuhashi S, Kuriyama H, Kimura T, Kanemaru H, Makino K, Miyashita A, Aoi J, Makino T, et al. Methyl-CpG binding domain protein 3: a new diagnostic marker and potential therapeutic target of melanoma. *Biosci Trends.* 2020;14:390–5.
 63. Rangaraju S, Dammer EB, Raza SA, Rathakrishnan P, Xiao H, Gao T, Duong DM, Pennington MW, Lah JJ, Seyfried NT, Levey AI. Identification and therapeutic modulation of a pro-inflammatory subset of disease-associated-microglia in Alzheimer's disease. *Mol Neurodegener.* 2018;13:24.
 64. Gerrits E, Brouwer N, Kooistra SM, Woodbury ME, Vermeiren Y, Lambourne M, Mulder J, Kummer M, Möller T, Biber K, et al. Distinct amyloid- β and tau-associated microglia profiles in Alzheimer's disease. *Acta Neuropathol.* 2021;141:681–96.
 65. Schurmann B, Bermingham DP, Kopeikina KJ, Myczek K, Yoon S, Horan KE, Kelly CJ, Martin-de-Saavedra MD, Forrest MP, Fawcett-Patel JM, et al.

- A novel role for the late-onset Alzheimer's disease (LOAD)-associated protein Bin1 in regulating postsynaptic trafficking and glutamatergic signaling. *Mol Psychiatry*. 2020;25:2000–16. <https://doi.org/10.1038/s41380-019-0407-3>.
66. Perdigao C, Barata MA, Burrinha T, Guimas Almeida C. Alzheimer's disease BIN1 coding variants increase intracellular Abeta levels by interfering with BACE1 recycling. *J Biol Chem*. 2021;297:101056. <https://doi.org/10.1016/j.jbc.2021.101056>.
 67. Lambert E, Saha O, Soares Landeira B, Melo de Farias AR, Hermant X, Carrier A, Pelletier A, Gadaud J, Davoine L, Dupont C, et al. The Alzheimer susceptibility gene BIN1 induces isoform-dependent neurotoxicity through early endosome defects. *Acta Neuropathol Commun*. 2022;10:4. <https://doi.org/10.1186/s40478-021-01285-5>.
 68. Franzmeier N, Ossenkoppeler R, Brendel M, Rubinski A, Smith R, Kumar A, Mattsson-Carlgrén N, Strandberg O, Düring M, Buerger K, et al. The BIN1 rs744373 Alzheimer's disease risk SNP is associated with faster Abeta-associated tau accumulation and cognitive decline. *Alzheimers Dement*. 2022;18:103–15. <https://doi.org/10.1002/alz.12371>.
 69. Lebon S, Minal L, Chretien D, Corcos J, Serre V, Kadhom N, Steffann J, Pauchard JY, Munnich A, Bonnefont JP, Rötig A. A novel mutation of the NDUFS7 gene leads to activation of a cryptic exon and impaired assembly of mitochondrial complex I in a patient with Leigh syndrome. *Mol Genet Metab*. 2007;92:104–8.
 70. Wang J, Gao QS, Wang Y, Lafyatis R, Stamm S, Andreadis A. Tau exon 10, whose missplicing causes frontotemporal dementia, is regulated by an intricate interplay of cis elements and trans factors. *J Neurochem*. 2004;88:1078–90.
 71. Bampton A, Gittings LM, Fratta P, Lashley T, Gatt A. The role of hnRNPs in frontotemporal dementia and amyotrophic lateral sclerosis. *Acta Neuropathol*. 2020;140:599–623.
 72. Krakowiak PA, Wassif CA, Kratz L, Cozma D, Kovárová M, Harris G, Grinberg A, Yang Y, Hunter AG, Tsokos M, et al. Lathosterolosis: an inborn error of human and murine cholesterol synthesis due to lathosterol 5-desaturase deficiency. *Hum Mol Genet*. 2003;12:1631–41.
 73. Barrera J, Song L, Safi A, Yun Y, Garrett ME, Gamache J, Premasinghe I, Sprague D, Chipman D, Li J, et al. Sex dependent glial-specific changes in the chromatin accessibility landscape in late-onset Alzheimer's disease brains. *bioRxiv*. 2021. <https://doi.org/10.1101/2021.04.07.438835>.
 74. Zhou Y, Song WM, Andhey PS, Swain A, Levy T, Miller KR, Poliani PL, Cominelli M, Grover S, Gilfillan S, et al. Human and mouse single-nucleus transcriptomics reveal TREM2-dependent and TREM2-independent cellular responses in Alzheimer's disease. *Nat Med*. 2020;26:131–42.
 75. Leng K, Li E, Eser R, Piergies A, Sit R, Tan M, Neff N, Li SH, Rodriguez RD, Suemoto CK, et al. Molecular characterization of selectively vulnerable neurons in Alzheimer's disease. *Nat Neurosci*. 2021;24:276–87.
 76. Del-Aguila JL, Li Z, Dube U, Mihindukulasuriya KA, Budde JP, Fernandez MV, Ibanez L, Bradley J, Wang F, Bergmann K, et al. A single-nuclei RNA sequencing study of Mendelian and sporadic AD in the human brain. *Alzheimer's Res Ther*. 2019;11:71.
 77. Hansen DV, Hanson JE, Sheng M. Microglia in Alzheimer's disease. *J Cell Biol*. 2018;217:459–72.
 78. Qin Q, Teng Z, Liu C, Li Q, Yin Y, Tang Y. TREM2, microglia, and Alzheimer's disease. *Mech Ageing Dev*. 2021;195:111438.
 79. Romero-Molina C, Garretti F, Andrews SJ, Marcora E, Goate AM. Microglial efferocytosis: diving into the Alzheimer's disease gene pool. *Neuron*. 2022;110:3513–33.
 80. Bolmont T, Haiss F, Eicke D, Radde R, Mathis CA, Klunk WE, Kohsaka S, Jucker M, Calhoun ME. Dynamics of the microglial/amyloid interaction indicate a role in plaque maintenance. *J Neurosci*. 2008;28:4283–92.
 81. El Khoury J, Luster AD. Mechanisms of microglia accumulation in Alzheimer's disease: therapeutic implications. *Trends Pharmacol Sci*. 2008;29:626–32.
 82. Wegiel J, Wang KC, Tarnawski M, Lach B. Microglia cells are the driving force in fibrillar plaque formation, whereas astrocytes are a leading factor in plaque degradation. *Acta Neuropathol*. 2000;100:356–64.
 83. Hellwig S, Masuch A, Nestel S, Katzmarski N, Meyer-Luehmann M, Biber K. Forebrain microglia from wild-type but not adult 5xFAD mice prevent amyloid- β plaque formation in organotypic hippocampal slice cultures. *Sci Rep*. 2015;5:14624.
 84. Sastre M, Dewachter I, Landreth GE, Willson TM, Klockgether T, van Leuven F, Heneka MT. Nonsteroidal anti-inflammatory drugs and peroxisome proliferator-activated receptor-gamma agonists modulate immunostimulated processing of amyloid precursor protein through regulation of beta-secretase. *J Neurosci*. 2003;23:9796–804.
 85. Yan Q, Zhang J, Liu H, Babu-Khan S, Vassar R, Biere AL, Citron M, Landreth G. Anti-inflammatory drug therapy alters beta-amyloid processing and deposition in an animal model of Alzheimer's disease. *J Neurosci*. 2003;23:7504–9.
 86. Wu Z, Nakanishi H. Lessons from microglia aging for the link between inflammatory bone disorders and Alzheimer's disease. *J Immunol Res*. 2015;2015:471342.
 87. Spangenberg EE, Green KN. Inflammation in Alzheimer's disease: lessons learned from microglia-depletion models. *Brain Behav Immun*. 2017;61:1–11.
 88. Asai H, Ikezu S, Tsunoda S, Medalla M, Luebke J, Haydar T, Wolozin B, Butovsky O, Kügler S, Ikezu T. Depletion of microglia and inhibition of exosome synthesis halt tau propagation. *Nat Neurosci*. 2015;18:1584–93.
 89. Leyns CEG, Holtzman DM. Glial contributions to neurodegeneration in tauopathies. *Mol Neurodegener*. 2017;12:50.
 90. Zhang YW, Sloan SA, Clarke LE, Caneda C, Plaza CA, Blumenthal PD, Vogel H, Steinberg GK, Edwards MS, Li G, et al. Purification and characterization of progenitor and mature human astrocytes reveals transcriptional and functional differences with mouse. *Neuron*. 2016;89:37–53.
 91. Shi Y, Yamada K, Liddel SA, Smith ST, Zhao L, Luo W, Tsai RM, Spina S, Grinberg LT, Rojas JC, et al. ApoE4 markedly exacerbates tau-mediated neurodegeneration in a mouse model of tauopathy. *Nature*. 2017;549:523–7.
 92. Krasemann S, Madore C, Cialic R, Baufeld C, Calcagno N, El Fatimy R, Beckers L, O'Loughlin E, Xu Y, Fanek Z, et al. The TREM2-APOE pathway drives the transcriptional phenotype of dysfunctional microglia in neurodegenerative diseases. *Immunity*. 2017;47:566–81.
 93. Lambert JC, Heath S, Even G, Campion D, Sleegers K, Hiltunen M, Combarros O, Zelenika D, Bullido MJ, Tavernier B. Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. *Nat Genet*. 2009;41:1094–9.
 94. Heinzen EL, Need AC, Hayden KM, Chiba-Falek O, Roses AD, Strittmatter WJ, Burke JR, Hulette CM, Welsh-Bohmer KA, Goldstein DB. Genome-wide scan of copy number variation in late-onset Alzheimer's disease. *J Alzheimers Dis*. 2010;19:69–77.
 95. Kamboh MI, Barmada MM, Demirci FY, Minster RL, Carrasquillo MM, Pankratz VS, Younkin SG, Saykin AJ, Sweet RA, Alzheimer's Disease Neuroimaging Initiative, et al. Genome-wide association analysis of age-at-onset in Alzheimer's disease. *Mol Psychiatry*. 2012;17:1340–6.
 96. Kamboh MI, Demirci FY, Wang X, Minster RL, Carrasquillo MM, Pankratz VS, Younkin SG, Saykin AJ, Jun G, Initiative ASDN. Genome-wide association study of Alzheimer's disease. *Transl Psychiatry*. 2012;2:e117.
 97. Coon KD, Myers AJ, Craig DW, Webster JA, Pearson JV, Lince DH, Zismann VL, Beach TG, Leung D, Bryden L. A high-density whole-genome association study reveals that APOE is the major susceptibility gene for sporadic late-onset Alzheimer's disease. *J Clin Psychiatry*. 2007;68:613–8.
 98. Gottschalk WK, Mihovilovic M, Roses AD, Chiba-Falek O. The role of upregulated APOE in Alzheimer's disease etiology. *J Alzheimers Dis Parkinsonism*. 2016;6:209.
 99. Yang A, Kantor B, Chiba-Falek O. APOE: the new frontier in the development of a therapeutic target towards precision medicine in late-onset Alzheimer's. *Int J Mol Sci*. 2021;22:1244.
 100. Akram A, Schmeidler J, Katsel P, Hof PR, Haroutunian V. Association of ApoE and LRP mRNA levels with dementia and AD neuropathology. *Neurobiol Aging*. 2012;33:628.
 101. Hashemiaghdam A, Mroczek M. Microglia heterogeneity and neurodegeneration: the emerging paradigm of the role of immunity in Alzheimer's disease. *J Neuroimmunol*. 2020;341:577185.
 102. Zheng JY, Sun J, Ji CM, Shen L, Chen ZJ, Xie P, Sun YZ, Yu RT. Selective deletion of apolipoprotein E in astrocytes ameliorates the spatial learning and memory deficits in Alzheimer's disease (APP/PS1) mice by inhibiting TGF-beta/Smad2/STAT3 signaling. *Neurobiol Aging*. 2017;54:112–32.

103. Huynh TP, Liao F, Francis CM, Robinson GO, Serrano JR, Jiang H, Roh J, Finn MB, Sullivan PM, Esparza TJ. Age-dependent effects of apoE reduction using antisense oligonucleotides in a model of beta-amyloidosis. *Neuron*. 2017;96:1013–23.
104. Bien-Ly N, Gillespie AK, Walker D, Yoon SY, Huang Y. Reducing human apolipoprotein E levels attenuates age-dependent Abeta accumulation in mutant human amyloid precursor protein transgenic mice. *J Neurosci*. 2012;32:4803–11.
105. Kim J, Jiang H, Park S, Eltorai AE, Stewart FR, Yoon H, Basak JM, Finn MB, Holtzman DM. Haploinsufficiency of human APOE reduces amyloid deposition in a mouse model of amyloid-beta amyloidosis. *J Neurosci*. 2011;31:18007–12.
106. Foraker J, Millard SP, Leong L, Thomson Z, Chen S, Keene CD, Bekris LM, Yu CE. The APOE gene is differentially methylated in Alzheimer's disease. *J Alzheimers Dis Parkinsonism*. 2015;48:745–55.
107. Shao Y, Shaw M, Todd K, Khrestian M, D'Aleo G, Barnard PJ, Zahratka J, Pillai J, Yu CE, Keene CD. DNA methylation of TOMM40-APOE-APOC2 in Alzheimer's disease. *J Hum Genet*. 2018;63:459–71.
108. Mancera-Paez O, Estrada-Orozco K, Mahecha MF, Cruz F, Bonilla-Vargas K, Sandoval N, Guerrero E, Salcedo-Tacuma D, Melgarejo JD, Vega E. Differential methylation in APOE (Chr19; Exon Four; from 44,909,188 to 44,909,373/hg38) and increased apolipoprotein E plasma levels in subjects with mild cognitive impairment. *Int J Mol Sci*. 2019;20:1394.
109. Babenko VN, Afonnikov DA, Ignatieva EV, Klimov AV, Gusev FE, Rogaev EI. Haplotype analysis of APOE intragenic SNPs. *BMC Neurosci*. 2018;19:16.
110. Hafemeister C, Satija R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol*. 2019;20:20.
111. Cusanovich DA, Reddington JP, Garfield DA, Daza RM, Aghamirzaie D, Marco-Ferreres R, Pliner HA, Christiansen L, Qiu X, Steemers FJ, et al. The cis-regulatory dynamics of embryonic development at single-cell resolution. *Nature*. 2018;555:538–42.
112. Bi Y, Kim H, Gupta R, Davuluri RV. Tree-based position weight matrix approach to model transcription factor binding site profiles. *PLoS One*. 2011;6:e24210.
113. Khamis AM, Motwalli O, Oliva R, Jankovic BR, Medvedeva YA, Ashoor H, Essack M, Gao X, Bajic VB. A novel method for improved accuracy of transcription factor binding site prediction. *Nucleic Acids Res*. 2018;46:e72.
114. Tompa M, Li N, Bailey TL, Church GM, De Moor B, Eskin E, Favorov AV, Frith MC, Fu Y, Kent WJ, et al. Assessing computational tools for the discovery of transcription factor binding sites. *Nat Biotechnol*. 2005;23:137–44.
115. Wasserman W, Sandelin A. Applied bioinformatics for the identification of regulatory elements. *Nat Rev Genet*. 2004;5:276–87.
116. Li G, Cai L, Chang H, Hong P, Zhou Q, Kulakova EV, Kolchanov NA, Ruan Y. Chromatin interaction analysis with Paired-End Tag (ChIA-PET) sequencing technology and application. *BMC Genomics*. 2014;15:S11.
117. Jiang Y, Matevossian A, Huang HS, Straubhaar J, Akbarian S. Isolation of neuronal chromatin from brain tissue. *BMC Neurosci*. 2008;9:42. <https://doi.org/10.1186/1471-2202-9-42>.
118. Marzluff WF. Preparation of active nuclei. *Methods Enzymol*. 1990;181:30–6. [https://doi.org/10.1016/0076-6879\(90\)81109-8](https://doi.org/10.1016/0076-6879(90)81109-8).
119. Hao Y, Hao S, Andersen-Nissen E, Mauck WM, Zheng S, Butler A, Lee MJ, Wilk AJ, Darby C, Zagar M, et al. Integrated analysis of multimodal single-cell data. *bioRxiv*. 2020. <https://doi.org/10.1101/2020.10.12.335331>.
120. Hafemeister C, Satija R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol*. 2019;20:296. <https://doi.org/10.1186/s13059-019-1874-1>.
121. Lause J, Berens P, Kobak D. Analytic Pearson residuals for normalization of single-cell RNA-seq UMI data. *bioRxiv*. 2020. <https://doi.org/10.1101/2020.12.01.405886>.
122. Frankish A, Diekhans M, Ferreira AM, Johnson R, Jungreis I, Loveland J, Mudge JM, Sisu C, Wright J, Armstrong J, et al. GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res*. 2019;47:D766–73. <https://doi.org/10.1093/nar/gky955>.
123. Stuart T, Srivastava A, Madad S, Lareau CA, Satija R. Single-cell chromatin state analysis with Signac. *Nat Methods*. 2021;18:1333–41.
124. Amemiya HM, Kundaje A, Boyle AP. The ENCODE blacklist: identification of problematic regions of the genome. *Sci Rep*. 2019;9:9354.
125. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, Baglaenko Y, Brenner M, Loh PR, Raychaudhuri S. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods*. 2019;16:1289–96.
126. Cusanovich DA, Hill AJ, Aghamirzaie D, Daza RM, Pliner HA, Berletch JB, Filippova GN, Huang X, Christiansen L, DeWitt WS, et al. A single-cell atlas of in vivo mammalian chromatin accessibility. *Cell*. 2018;174:1309–1324.e1318.
127. Zhang Y, Liu T, Meyer CA, Eeckhoutte J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS. Model-based analysis of ChIP-Seq (MACS). *Genome Biol*. 2008;9:R137.
128. Jalili V, Matteucci M, Masseroli M, Morelli MJ. Using combined evidence from replicates to evaluate ChIP-seq peaks. *Bioinformatics (Oxford, England)*. 2015;31:2761–9.
129. Finak G, McDavid A, Yajima M, Deng J, Gersuk V, Shalek AK, Slichter CK, Miller HW, McElrath MJ, Prlic M, et al. MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol*. 2015;16:278. <https://doi.org/10.1186/s13059-015-0844-5>.
130. Ntranos V, Yi L, Melsted P, Pachter L. Identification of transcriptional signatures for cell types from single-cell RNA-Seq. Unpublished. 2018
131. Alexa A, Rahnenfuhrer J. (2022). topGO: Enrichment analysis for gene ontology. R package version 2.50.0.
132. Fornes O, Castro-Mondragon JA, Khan A, van der Lee R, Zhang X, Richmond PA, Modi BP, Correard S, Gheorghe M, Baranašić D, et al. JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res*. 2020;48:D87–92.
133. Schep, A. (2022). motifmatcher: Fast Motif Matching in R. R package version 1.20.0.
134. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell*. 2010;38:576–89.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.