Single-nucleus multiomic atlas of frontal cortex in amyotrophic lateral sclerosis with a deep learning-based decoding of alternative polyadenylation mechanisms

Paul M. McKeever[*,1]; Aiden M. Sababi[*,2,3]; Raghav Sharma[1,4]; Nicholas Khuu[5]; Zhiyu Xu[5]; Shu Yi Shen[4]; Shangxi Xiao[1]; Philip McGoldrick[1]; Elias Orouji[5]; Troy Ketela[5]; Christine Sato[1]; Danielle Moreno[1]; Naomi Visanji[1]; Gabor G. Kovacs[1,4]; Julia Keith[7], Lorne Zinman[8]; Ekaterina Rogaeva[1]; Hani Goodarzi[6]; Gary D. Bader[2,3,9,10,11]; Janice Robertson[1,4]

*. These authors contributed equally to the work
1. Tanz Centre for Research in Neurodegenerative Diseases, University of Toronto, Toronto, ON, Canada
2. Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada
3. The Donnelly Centre, University of Toronto, ON, Canada
4. Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, ON, Canada
5. Princess Margaret Cancer Centre, University Health Network, Toronto, ON, Canada.
6. Department of Biophysics & Biochemistry, University of California San Francisco, United States of America
7. Laboratory Medicine and Molecular Diagnostics, Division of Anatomical Pathology, Sunnybrook Health Sciences Centre, University of Toronto
8. Department of Medicine, Division of Neurology, Sunnybrook Health Sciences Centre, University of Toronto
9. Lunenfeld-Tanenbaum Research Institute, Sinai Health System, Toronto, ON, Canada
10. Department of Computer Science, University of Toronto, ON, Canada
11. Princess Margaret Research Institute, University Health Network, Toronto, ON, Canada

Correspondence and request for materials should be sent to either J.R. (jan.robertson@utoronto.ca) or G.D.B (gary.bader@utoronto.ca)

Abstract
The understanding of how different cell types contribute to amyotrophic lateral sclerosis (ALS) pathogenesis is limited. Here we generated a single-nucleus transcriptomic and epigenomic atlas of the frontal cortex of ALS cases with C9orf72 (C9) hexanucleotide repeat expansions and sporadic ALS (sALS). Our findings reveal shared pathways in C9-ALS and sALS, characterized by synaptic dysfunction in excitatory neurons and a disease-associated state in microglia. The disease subtypes diverge with loss of astrocyte homeostasis in C9-ALS, and a more substantial disturbance of inhibitory neurons in sALS. Leveraging high depth 3'-end sequencing, we found a widespread switch towards distal polyadenylation (PA) site usage across ALS subtypes relative to controls. To explore this differential alternative PA (APA), we developed APA-Net, a deep neural network model that uses transcript sequence and expression levels of RNA-binding proteins (RBPs) to predict cell-type specific APA usage and RBP interactions likely to regulate APA across disease subtypes.

Introduction
The advent of transcriptomic and epigenomic interrogation at the single cell level facilitates a deeper understanding of the mechanisms underpinning disease, especially in complex tissues such as the brain. Amyotrophic lateral sclerosis (ALS) is an adult-onset neurodegenerative disease caused by degeneration of motor neurons in the brain, brainstem and spinal cord. Over 50% of ALS patients also exhibit progressive cognitive and/or behavioral symptoms caused by frontotemporal lobar degeneration (FTLD), with 10-15% of patients fulfilling the diagnostic criteria of frontotemporal dementia (FTD). Although the vast majority of ALS cases are sporadic (sALS), with no family history of disease, over 30 genes have been associated with disease causation, including SOD1, TARDBP (TDP-43) and FUS. The most common genetic cause linking ALS and FTLD is a G4C2 hexanucleotide repeat expansion within intron 1 of C9orf72 (C9)[1,2], a gene with unclear function but with emerging roles in nucleocytoplasmic transport[3] and synaptic homeostasis[4–6].

The molecular underpinnings of cortical dysfunction associated with FTLD in ALS remain unclear but has been linked with changes in expression and/or subcellular localizations of various RNA binding proteins (RBPs), including TDP-43[7–10], the most prevalent disease pathology in ALS/FTLD, as well as FUS[11], SFPQ[12,13], TIA1[14], and heterogeneous ribonucleoproteins (HNRNPs)[15,16]. RBPs regulate diverse functions across cell types, including chromatin dynamics[10,17–19], transcription[7–10], auto-regulation[20,21], alternative splicing (AS)[8], and alternative polyadenylation (APA)[20,22–25]. Evidence supporting disruption of RBP functions has been reported in ALS cases[8,26]. Brain region-specific APA is seen in postmortem bulk RNA-seq analysis of ALS patients[27] and in iPSC-derived motor neurons from patients with *TARDBP* and *VCP* mutations[28]. However, the cell type-specific mechanisms underlying APA in postmortem tissue from patients with ALS have not been explored.

Here, we generated a single nucleus RNA-seq (snRNA-seq) and Assay for Transposase Accessible Chromatin Sequencing (snATAC-seq) atlas of 196,896 nuclei to uncover transcriptomic and epigenomic changes in post-mortem orbitofrontal cortex from C9-ALS and sALS cases. A comparison of C9-ALS and sALS cases with non-neurological controls was performed to assess mutation- and disease-specific effects, and provide novel insight into cellular changes in this ALS-affected brain region. We find cell type- and disease subtype-specific alterations in cis-regulatory elements, transcription factor (TF) motif enrichment, and pathways in C9-ALS and sALS cases. We also analyzed cell type-specific dysregulation of APA in ALS, identifying thousands of significant APA events, including distal, intronic and internal exonic APAs in both diseases compared to non-neurological controls. To further decode this APA dysregulation, we developed an interpretable state-of-the-art deep learning method called Alternative Polyadenylation Network (APA-Net). APA-Net enabled us to identify a range of *cis/trans* regulators correlated with the observed APA events in the disease. By interpreting APA-Net, we identify potential RBP interactions in dysregulating APA in ALS, shedding more light on the regulatory programs likely to induce APA in ALS. These findings improve our understanding of cell-type specific transcriptomic and epigenetic changes occurring in the frontal cortex in ALS, and provide mechanistic insights into the potential coordinated interaction of RBPs in regulating APA in disease.

Results

Transcriptomic and epigenomic profiling of single nuclei from the frontal cortex of ALS patients

We generated and analyzed a single cell atlas containing 196,896 nuclei derived from 24 snRNA-seq (103,076 nuclei) and 21 snATAC-seq (93,820 nuclei) samples from orbitofrontal cortex (Brodmann area 11; referred to here as frontal cortex) of C9-ALS (n=10 individuals) and sALS (n=8) cases compared to non-neurological controls (n=6) (Figure 1, Supplementary Data Fig. 1a, Supplementary Table 1). Standard quality control (QC) guidelines were applied to the snRNA-seq[29] and snATAC-seq[30–33] data. We used an established snRNA-seq data integration method[34], ensuring consistent alignment of cell clusters across technologies for sex (Supplementary Data Fig. 1b), disease subtype (Supplementary Data Fig. 1c), and samples (Supplementary Data Fig. 1d).

To delineate cell types, a consistent clustering algorithm[35] and cell annotation approach[36,37] was used for both snRNA-seq and snATAC-seq datasets (Supplementary Data Fig. 1e-f) (Methods). We uncovered 23 frontal cortex cell subtypes for both snRNA-seq and snATAC-seq datasets, including oligodendrocytes, oligodendrocyte precursor cells (OPCs), astrocytes, endothelial cells, vascular leptomeningeal cells (VLMCs), microglia and perivascular macrophages (PVM), ten excitatory neuron subtypes, and seven inhibitory neuron subtypes (Supplementary Data Fig. 1g). A catalog of specific gene expression markers for each cell type is presented in Supplementary Data Fig. 2a. To facilitate downstream analysis, we grouped cell subtypes into seven major cell types: oligodendrocytes, OPCs, astrocytes, endothelial-VLMC, microglia-PVM, excitatory neurons, and inhibitory neurons, based on canonical markers (Supplementary Data Fig. 2b-e). The nuclei from endothelial-VLMC cells were excluded from all downstream analyses due to poor yield across samples and technologies.

All major expected cortical cell types were consistently found in both snRNA-seq and snATAC-seq samples, resulting in similar distributions of cell types when categorized by disease subtype (Supplementary Data Fig. 3a-c). Gene activity levels, measured by the open chromatin peaks from the snATAC-seq data[32,38], show

significant correlation with gene expression levels across cell types (Pearson R > 0.65 and p-values < 1e-6). Collectively, these two technologies exhibit robust concordance (Supplementary Data Fig. 3f), supporting their quality and use in subsequent analyses.

Converging and diverging transcriptomic and epigenomic changes across ALS subtypes

To uncover the transcriptomic and epigenomic cell-states in frontal cortex of C9-ALS and sALS cases compared with controls, we carried out differential expression analysis[39] and computed differential accessible chromatin regions (DARs) (Methods). We observe that a majority of large and small ribosomal machinery subunits (*RPL* and *RPS* prefix) are dysregulated across cell types in C9-ALS and sALS (Supplementary Data Figure 4a-f). These findings agree with recent reports from dorsolateral prefrontal and primary motor cortex neurons from C9-ALS[40,41] and sALS[42]. This highlights dysregulation of translation machinery as a significant cellular state in ALS. To alleviate the overrepresentation of translational machinery pathways in the analysis, we removed all ribosomal subunit genes from the count matrix which enables the identification of other cell-type specific changes[43].

The subsequent differential expressed genes (DEG) analysis revealed distinct changes in major cell types between C9-ALS and sALS relative to controls (all adjusted *p*-value < 0.05; Fig. 2a-c, Extended Data Fig. 1a-f, Supplementary Data Table 2). This analysis shows significant changes in DEGs and DARs in excitatory neurons in both ALS subtypes (Fig. 2c-d). However, inhibitory neurons in sALS show higher DEG and DAR dysregulation than in C9-ALS (Fig. 2b, Fig. 2e-f).

Pathway enrichment analysis on DEGs shows that signal transduction terms are enriched in excitatory neurons in C9-ALS, with up-regulated genes including *WNT9A*, *EDNRB*, *PRKCH*, *AR*, and *EGF* (Fig. 2g). Dysregulation of androgen receptors (AR) has been associated with ALS motor neurons and impaired motor neuron/cranial nerve function[44,45]. Excitatory neurons in sALS show enrichment for lipid metabolism and neuronal system genes. No significantly altered pathways were detected in C9-ALS inhibitory neurons. In contrast, genes involved in neuronal system and synaptic pathways such as *GRIN3A, KCNG1*, *KCNG2*, *SYT10*, and *CAMK2D* (Fig. 2h) are enriched in sALS inhibitory neurons, differentiating sALS from C9-ALS.

Altered cellular states in microglia are important for the pathogenesis of ALS, and with neurodegeneration more broadly[46–48]. In microglia in C9-ALS, we observe up-regulation of *IL1B* and *CD83,* and enrichment of innate immune response pathways (Extended Data Fig. 1b), suggesting a potential shift of these cells toward a proinflammatory state[49,50]. To further characterize microglia in ALS, we scored microglial cell states based on those typically observed in neurodegenerative disease[48] by ranking cell state markers according to their gene expression levels[51]. We observe a decrease in homeostatic microglial markers such as *CSF1R*, *CX3CR1*, and *P2RY12* in both ALS subtypes with C9-ALS showing the lowest scores (Extended Data Fig. 1c). In both ALS subtypes, we observe an increase in disease-associated microglia (DAM) stage 1 markers, including TYROBP, APOE, and B2M. DAM stage 2 markers, however, show only a marginal increase (Extended Data Fig. 1d-e) with SPP1 the most significant (Supplementary Data Table 3). These results indicate a transcriptional shift from homeostatic microglia towards a disease associated state.

Astrocytes are linked with the pathophysiology of ALS through reactive toxicity and loss of homeostasis[52,53]. Here, we observe up-regulation of reactive astrocyte genes *ACTN1*, *MYH1*, and *OLFM1*[54] (Extended Data Fig. 1f). Upregulated GFAP is a common feature in reactive astrocytes[55]. While *GFAP* is not among the DEGs for astrocytes in C9-ALS or sALS, we observe significant increase in chromatin accessibility (Supplementary Data Table 3) and higher gene expression of *GFAP* (Extended Data Fig. 1g) in both ALS subtypes. These changes are concordant with astrocytic changes reported in C9-ALS dorsolateral prefrontal and primary motor cortex[40]. The differential expression and accessibility analyses reveal that astrocytes in C9-ALS are substantially more affected than in sALS, as is evident by the larger number of detected DEGs and DARs (Extended Data Fig. 1h-i).

To study how DARs are distributed across genomic regions, we annotated open chromatin peak locations[56]. Genomic regions for excitatory neurons in C9-ALS and sALS are highly enriched in proximal promoter, intronic, and distal intergenic loci (Fig. 2i). DARs annotate to proximal promoter regions in inhibitory neurons, oligodendrocytes, OPCs, and astrocytes (Fig. 2j, Extended Data Fig. 1j-m). Oligodendrocytes, OPCs, and

astrocytes in sALS show a larger proportion of proximal peaks compared with C9-ALS (Extended Data Fig. 1j-l), suggesting promoter over distal element level regulation by TFs in these cells.

To evaluate the biological implications of DARs in C9-ALS and sALS, we conducted a pathway enrichment analysis across cell types, focusing on enriched peaks in either proximal promoter and intronic or distal regions (Supplementary Data Table 4). For excitatory neurons in C9-ALS, proximal peaks are associated with genes involved in signal transduction and protein localization, aligning with the corresponding DEGs observed in excitatory neurons (Fig. 2). Distal peaks in excitatory neurons in C9-ALS did not yield significant pathways. In excitatory neurons in sALS, both proximal and distal peaks are related to neuronal development, whereas distal peaks relate to potassium ion transport. For inhibitory neurons from C9-ALS, the enrichment is exclusively in proximal peaks, pointing to processes such as transcriptional regulation, development, and growth. By contrast, proximal peaks in inhibitory neurons in sALS are associated with lipid metabolism, nervous system development, and transcriptional regulation, and these overlap with pathways identified from the DEG analysis (Fig. 2). Microglia in C9-ALS show proximal peaks focused on tumor necrosis factor regulation and leukocyte adhesion, while distal peaks are associated with JAK-STAT signaling and leukocyte differentiation. Overlapping up-regulated gene examples include *JAK3*, *STAT3*, and *LY96* in C9-ALS microglia (Supplementary Data Table 2). In sALS microglia, only the distal peaks show enrichment for JAK-STAT signaling, with proximal peaks displaying no significant enrichment. Up-regulation of JAK-STAT signaling has been observed in spinal cords[57] and cortices[58] from ALS patients and C9orf72 negatively regulates JAK-STAT signaling *in vivo*[59]. These findings indicate the presence of epigenetic regulatory mechanisms involved in the cell type-specific dysregulated pathways in ALS frontal cortex.

ALS risk gene expression across frontal cortex cell types

To investigate the cell type expression level of ALS-associated genes, we analyzed the expression patterns of a comprehensive set of genes (n=30) implicated in familial ALS or linked to heightened risk in sporadic cases[60,61]. Hierarchical clustering of ALS risk gene expression patterns by cell types and conditions (columns) primarily reveals a segregation pattern based on cell type, with disease subtype as the next most distinguishing factor (Extended Data Fig. 2). Gene clustering was conducted to examine how ALS risk gene expression patterns are partitioned. A total of seven gene clusters are observed: Cluster 1 (*NEK1*, *SQSTM1*, *ATXN2*) and cluster 6 (*FIG4*, *C9orf72*, *VAPB*, *TBK1*, *SETX*, *GRN*, *FUS*, *SPG11*, *PFN1*, *ANG*) show high expression across all major cell types, with particularly high expression in microglial cells. Cluster 2 containing *KIF5A*, *OPTN*, *ALS2*, *TUBA4A*, and *CCNF* is highly enriched in neurons; Cluster 3, characterized by the mitochondrial genes *SOD1* and *CHCHD10*, is predominantly associated with excitatory and inhibitory neurons (Extended Data Fig. 1). The observed increase in SOD1 expression across different cell types in ALS suggests that it might play a significant role in the disease and could be explored as a therapeutic target not only for the mutant SOD1-linked ALS[62] but for additional ALS subtypes as well. Cluster 4 (*SIGMAR1*, *TIA1*, *MATR3*) and Cluster 5 (*SPAST*, *TARDBP*, *CHMP2B*, *UBQLN2*) show high expression across all cell types except microglia. Overall, these findings agree with other single cell studies on ALS exploring a different frontal cortex region and the motor cortex[40,41] suggesting that these gene expression levels are not specific to a cortical subregion in patients nor to a particular disease subtype.

Previous studies have found haploinsufficiency of *C9orf72* in C9-ALS as indicated by decreased transcript and protein levels[1,2,63–65]. We evaluated chromatin accessibility and RNA expression of C9orf72 across disease subtypes and cell types revealing accessible peaks in the 5′UTR region of *C9orf72* (Extended Data Fig. 3a). Microglial cells alone show intronic peaks between exons 3 and 4 of *C9orf72* which have high co-accessibility with 5′UTR end peaks (Extended Data Fig. 3a). A pseudo-bulk analysis of chromatin peaks uncovered decreased peak height near the location of the repeat expansion in C9-ALS cases relative to controls and sALS cases (Extended Data Fig. 3a). This chromatin accessibility decrease in the 5′UTR, the known site of *C9orf72* repeat expansions[1,2], points to a possible contributory role in *C9orf72* haploinsufficiency observed in these cases. The expression profile of *C9orf72* shows similar cell type enrichment in the transcriptome and epigenome (Extended Data Fig. 3b-c). Further, we observe a concomitant trend towards decreased *C9orf72* in the gene expression and gene activity scores in C9-ALS across all measured cell types (Extended Data Fig. 3d). Overall, *C9orf72* shows altered chromatin accessibility in C9-ALS, but a decrease in RNA expression levels is observed in both C9-ALS and sALS across surveyed cell types.

Cell type-enriched TF motif analysis in ALS subtypes

To establish the cell type TF motif landscape in accessible DNA, we annotated potential TF-binding sites in open chromatin regions across the ALS and control datasets (Extended Data Fig. 4a)[66]. The pseudo-bulk analysis of our snATAC data reveals TF-binding within a range of approximately 10-100kb from the transcriptional start site (TSS) in both directions (Extended Data Fig. 4b) suggesting cis-regulatory activity extending both up and downstream of the TSS.

To elucidate the cell type-specific activity of TF motifs, we performed a differential analysis across major cell types (Fig. 3a, Supplementary Data Table 5). For oligodendrocytes, this identifies motifs from the SOX family, which have established roles in oligodendroglia development[67]. The NFI family of TFs[68] are enriched in both OPCs and astrocytes, however, astrocytes show further enrichment for GSX1 and PAX4 motifs which are known to be active in astrocytes[69]. For microglia, SPI family motifs show highest enrichment and have been shown to regulate microglial lineage from myeloid cells[70,71]. TF motifs EGR2, NEUROD2, and OLIG3 motifs are enriched in excitatory neurons[72,73], whereas inhibitory neurons are characterized by ASCL family and TCF12 motifs which are involved in neuronal migration during cortical development[74].

The differentially accessible TF motifs across cell types demonstrate that TF motif enrichment differs between C9-ALS and sALS in neurons and glial cells relative to controls (Fig. 3b, Extended Data Fig. 4c, Supplementary Data Table 6). Overall, C9-ALS shows more altered TF motifs than in sALS, with ZSCAN11, ZNF621, and SMAD4 being the most significant in excitatory neurons in C9-ALS (Fig. 3b). In both disease subtypes, excitatory neuron-enriched TF motifs NEUROD2 and OLIG3 show altered activity (Fig. 3b). Inhibitory neurons show more differentially accessible motifs in sALS than C9-ALS (Fig. 3b), underscoring the DEG findings which suggest that this cell type is more affected in sALS (Fig. 2b). Both excitatory and inhibitory cells show alterations in MECP2, DNMT1, and FOXG1 in C9-ALS and sALS (Fig. 3b). For glial cells, we found that C9-ALS microglia and OPCs show more changes than in sALS, which agrees with the DEG results (Extended Data Fig. 2a, 4c). In contrast, sALS astrocytes show more TF motif variability than in C9-ALS (Extended Data Fig. 4c). Oligodendrocytes show few TF motif changes in either C9-ALS or sALS (Extended Data Fig. 4c).

We characterized the genomic regions where the differentially active TF motifs bind relative to the TSS across major cell types in C9-ALS and sALS (Fig. 3c-e, Extended Data Fig. 4d-f). The landscape of TF binding in ALS suggests cis-regulatory influences extending both upstream and downstream of the TSS, mirroring the distribution of genomic loci observed in DARs (Fig. 2, Extended Data Fig. 1). Particularly, TF motifs from inhibitory neurons are predominantly located proximal to the TSS across ALS subtypes as seen in DARs altered in the same cell type (Fig. 3d, Fig. 1j). For all glial cell types, the sALS cohort exhibits higher preference for TF motifs proximal to the TSS, suggesting regionally distinct regulatory patterns compared to C9-ALS (Fig. 3e, Extended Data Fig. 4d-f).

We next visualized the cell type and ALS subtype levels of open chromatin around TF binding site motifs. A TF footprinting analysis[75] uncovers a high Tn5 insertion density and drop-off around cell type-enriched TF motifs (Fig. 3f-h). The footprinting analysis reveals a notable reduction in Tn5 insertion up and downstream of NEUROD2 motifs in excitatory neurons from both ALS subtypes relative to controls (Fig. 3f). Considering the critical role of NEUROD2 in maintaining excitatory neuron homeostasis[76], this finding implicates a dysregulation of excitatory neuron homeostasis in ALS. FOXG1, which promotes neuronal survival[77], also shows increased Tn5 insertion in both excitatory and inhibitory neurons in C9-ALS and sALS (Extended Data Fig. 4g). Astrocytes from C9-ALS and sALS show increased Tn5 insertion around STAT1 (Fig. 3h) and BACH2 motifs (Extended Data Fig. 4i) suggesting a neuroinflammatory response[78]. Insertion near JUNB motifs in microglia shows a higher accessibility in C9-ALS relative to controls, whereas sALS demonstrates lower accessibility (Extended Data Fig. 4j). This indicates a potential divergent dysregulation of immediate early response factors[79] in microglia in C9-ALS versus sALS. Taken together, these results provide a foundation for TF motif variability by cell type in the frontal cortex and delineate TF motif changes in C9-ALS and sALS.

Dysregulation of APA landscape in ALS

Despite previous studies showing that APA is dysregulated in ALS[27], the underlying mechanisms have remained largely unexplored. We identified known polyadenylation (PA) sites in transcripts across our ALS subtypes and non-neurological controls and performed differential PA analysis comparing ALS with control[80,81] (Methods). We find many significant PA differences between ALS subtypes and controls across all major cell types (Fig. 4a,b, Supplementary Data Figure 5). In both ALS subtypes, we observe larger peaks and tails for positive APA usage (Fig. 4c) indicating a global trend towards distal PA sites.

We next examined the location of APA events in transcripts. While most events are found in the 3′ most exons, a significant number are also in intronic and exonic regions. This pattern was consistent across major cell types in both C9-ALS and sALS (Fig. 4d). The abundance of intronic APAs in ALS compared to controls suggests a potential increase in truncated gene transcripts, which could have a significant impact on gene function. The shift towards distal PA sites and longer 3′ UTRs could impact function through altering mRNA stability[82,83]. We observe lengthening and shortening of many ALS risk genes across major cell types in C9-ALS and sALS in 3′ UTR, introns, and internal exons (Extended Data Fig. 5a-j). For instance, in excitatory neurons, genes such as *TARDBP*, *MATR3*, *SETX*, and *ANXA11* exhibit lengthening, while *TIA1* shows 3′ UTR shortening (Fig. 4e,f). Increased intronic PAs usage for *UNC13A* is observed in excitatory and inhibitory neurons across ALS subtypes, indicating truncated *UNC13A* transcripts in ALS (Extended Data Fig. 5e-f,h). Depletion of *UNC13A* as a result of mis-splicing of the transcript has been reported in ALS[84,85] suggesting that *UNC13A* transcript structure modifications are involved in ALS.

To explore the biological significance of APA dysregulation in ALS, we performed a pathway analysis for each cell type (Supplementary Data Table 8). We found that lengthened genes in excitatory neurons from C9-ALS and sALS are enriched in pathways such as cell cycle[86], cell proliferation[87], vesicle organization[88], ribonucleoprotein complexes[40], and apoptosis[89] (Fig. 4g). In C9-ALS excitatory neurons, lengthened transcripts are enriched in pathways such as cell motility, organelle assembly, and protein localization, whereas shortened transcripts are enriched in transcription regulation and protein complex assembly. In sALS excitatory neurons, lengthened transcripts are related to mitosis and programmed cell death pathways (Fig. 4g), whereas no terms are enriched for shortened transcripts. These findings highlight shared and distinct pathways in C9-ALS and sALS subtypes. Taken together, our data indicate a complex dysregulation of the APA landscape in ALS, involving both increased shorter and longer gene transcripts that differ across C9-ALS and sALS cell types.

Deep learning model to map APA dysregulation in ALS

To decode the complex grammar of APA dysregulation in our ALS cohorts, we developed a deep learning model called APA-Net. APA-Net is trained to predict the APA log-fold change (LFC) values from our ALS datasets vs. control comparison (Fig. 4a,b) using the RNA sequences surrounding proximal and distal PA sites for each APA event per transcript and the RBP expression profiles per cell type (Fig. 5a) as input. To assess APA-Net performance, we compared predicted to known APA LFC values using Pearson correlation. APA-Net achieved a robust Pearson correlation coefficient of over 0.61 on a previously unseen C9-ALS test data set and 0.64 on an unseen sALS data set, across the entire data set (Fig. 5b,c, Extended Data Fig. 6a,b). APA profiles are highly specific for each cell type in C9-ALS and sALS (Fig. 5c, Extended Data Fig. 6b). Thus, APA-Net successfully learns cell type-specific APA profiles across disease subtypes.

Using a Convolutional Neural Network (CNN) architecture augmented with multi-head attention, our model was designed to identify cis-regulatory elements that influence PA site selection across different cell types. We optimized the model architecture, including kernel size and max pooling steps, to capture relevant genomic motifs (Fig.5a)[90,91]. To identify potential cis-regulatory elements, we used the filter weights from the CNN module of APA-Net, which represent learned sequence motifs. We scanned every sequence within the test dataset using the model's filters to identify regional subsequences where the filters showed the highest activation. We employed a position-weight matrix technique to capture the patterns each filter had learned for C9-ALS and sALS (Supplementary Data Fig. 6-7). We next aligned RBPs from the compendium of RNA-binding motifs[92] with the motifs found by APA-Net. We observe ALS risk genes FUS[93] and TDP43[94], as well as APA and AS factors such as HNRNPC, SFPQ, and SRSF7[95–99] among the aligned RBPs (Fig. 5d,e). A

differential expression analysis reveals significant dysregulation of these RBPs across the major cell types in both C9-ALS and sALS relative to control (Fig. 5e,f, Extended Data Fig. 6c).

To corroborate APA-Net findings, we used a CLIP-seq dataset to investigate the enrichment of HNRNPC binding sites in excitatory neuron transcripts. HNRNPC binds to regions surrounding proximal PA sites, and its knockdown can promote distal PA site usage[98,100]. Considering the observed downregulation of HNRNPC in excitatory neurons in C9-ALS (Fig. 5e), we hypothesized that HNRNPC binding sites are enriched in distal APA sites. To explore this, we performed an enrichment analysis of HNRNPC binding sites near proximal PAs. We observe a significant enrichment of binding sites in distal APA events and a lower enrichment in proximal APA events (Fig. 5g). Overall, our findings highlight the value of APA-Net's interpretability, enabling us to discern the underlying dysregulation of APA in ALS and deepening our molecular understanding of the disease.

RBP interactions reveal cell type-specific mechanisms and dysregulation of APA in ALS

Leveraging APA-Net's interpretability, we investigated potential RBPs involved in APA dysregulation in ALS. We first identified co-occurring sequence motifs that were automatically identified by APA-Net with our data and then identified RBPs that are known to bind these motifs. This resulted in RBP co-occurrence profiles that were clustered to identify major modules grouping RBPs with similar motif co-occurrence indicating potential interactions. C9-ALS and sALS had different module structures indicating a disease subtype-specific dysregulation (Fig. 6a, Extended Data Fig. 8).

In C9-ALS, Module 1 (Fig. 6a) highlights interactions with RBPs like MBNL1 and SRSF7, both involved in transcript shortening[99,101]. This module also features other APA-related factors such as HNRNPA1, as well as TDP-43 (*TARDBP*). Module 3 includes RBPs, such as CNOT4, SAMD4A and RBM24, which are associated with the coordinated regulation of transcription, AS, and APA[102]. Module 4, with the highest number of RBPs, highlights HNRNPC and PAPBN1, both involved in transcript lengthening[103,104]. This module also includes ALS-related RBPs like FUS, MATR3, SFPQ, and TIA1[11,93,105–107], alongside various other RBPs involved in transcript processing and APA. In sALS (Extended Data Fig. 8), Module 3 of the RBP interaction profile includes TDP-43, SRSF7, ENOX1, and FXR1. While SRSF7 can physically interact with TDP-43[108], ENOX1 demonstrates FUS and TDP-43-like properties in functional yeast screens[109]. The precise interactions between FXR1, a member of the FMRP family, and TDP-43 are not known; however, both FMRP family proteins and TDP-43 are concurrently detected in RNA granules in post-mortem brains of FTLD patients[110,111]. Taken together, our findings indicate that dysfunctional interactions among these RBPs could be a contributing factor to the complex interplay between AS and APA, thereby shaping the dysregulated APA landscape observed across ALS subtypes.

We next explored the RBP interaction patterns across various cell types. Distinct RBP interactions are found among different cell types in both C9-ALS and sALS (Extended Data Fig. 9a,b). We also identified eight distinctive RBP expression patterns (Fig. 6b), suggesting that specific sets of RBPs might be important in regulating APA events within individual cell types. This cell type-specificity in RBP interactions and expression profiles hints at the presence of distinct cis and trans regulatory elements that may influence APA. Collectively, our findings help map cell type-specific APA regulation in ALS and the complex molecular landscape underlying disease.

Discussion

In this study, we explored the molecular landscape of ALS through post-mortem tissue analysis of the orbitofrontal cortex, a region associated with behavioural impairments in ALS ALS/FTLD spectrum[112–118]. Using single nucleus transcriptomic and epigenomic profiling, and a state-of-the-art deep learning model, we shed light on the cell type-specific dysregulation of gene expression, open chromatin, and APA in C9-ALS and sALS. Our findings reveal that excitatory neurons in both C9-ALS and sALS exhibit altered cell states, characterized by changes in genes associated with neuronal and synaptic functions. Gene expression and APA changes are more pronounced in inhibitory neurons in sALS than C9-ALS when compared with control cases. Microglia from both C9-ALS and sALS similarly shift from a homeostatic to a disease-like phenotype.

ALS-related risk genes change expression across cell types in C9-ALS and sALS. Notably, there is a consistent upregulation of *SOD1* and *CHCHD10*, highlighting potential significance of these genes in disease subtypes other than mutant *SOD1*- and *CHCHD10*-linked ALS[119,62].

To explore the APA regulation in ALS, we introduce APA-Net, an interpretable deep learning model that can predict disease and cell-type specific APA events and their associated regulatory mechanisms. Compared to previous work, APA-Net is more interpretable and better performing, particularly as it provides cell type-specific APA insights[120,121]. The strength of APA-Net lies in its ability to handle identification of cis regulatory elements without relying on additional methods for post-analysis interpretation and to uncover cell type-specific RBP interactions, offering a window into their roles in APA dysregulation and their links to disease. This capability has led to the identification of novel RBP interaction candidates in both C9-ALS and sALS, shedding light on the cell type-specific APA dysregulation within ALS. Such insights deepen our understanding of ALS at the molecular level and extend the potential of APA-Net to investigate other diseases marked by APA dysregulation.

Our single-nucleus transcriptomic and epigenomic profiling from the frontal cortex provides a granular understanding of the cellular heterogeneity in ALS. The atlas presented here complements existing single cell mapping papers[40,122], which profiled dorsolateral prefrontal, occipital, and motor cortex regions of the brain in C9-ALS and sALS. We expand on these studies by providing an evaluation of the orbitofrontal cortex. Moreover, we introduce APA-Net to enable the cell type-specific probing of APA dysregulation. Our findings help map ALS disease mechanisms and targets useful for future therapeutic research.

Methods

Human brain samples

Informed consent was obtained from all participants in accordance with the Ethics Review Boards at Sunnybrook Health Sciences Centre and University of Toronto. ALS clinical diagnosis was determined based on the El Escorial revisited clinical criteria[123]. Fresh frozen orbitofrontal cortex tissues were collected from ALS cases with pathologically confirmed FTLD or no FTLD, and six non-neurological control cases (detailed information in Supplementary Data 1). C9-genotypes were determined as described previously[63,124]. Postmortem patient samples comprised C9-ALS/FTLD (n=7), C9-ALS no FTLD (n=3), and sALS no FTLD (n=8). Postmortem non-neurological control samples (n=6) were obtained from the Douglas-Bell Canada Brain Bank (DBCBB, Montreal, Canada) (n=4) or University Health Network–Neurodegenerative Brain Collection (UHN–NBC, Toronto, Canada) (n=2). Expert neuroanatomists ensured that each orbitofrontal cortex tissue section contained the entire laminar structure of the cortex (layers I-VI and white matter). A subset of samples were processed using the 10X v2 chemistry for C9-ALS/FTLD (n=3), C9-ALS no FTLD (n=1), and sALS no FTLD (n=4). These v2 samples were used solely for an in-depth cell type annotation process across disease subtypes. For all downstream analyses on snRNA-seq, exclusively v3 chemistry samples were used. Here, the C9-ALS/FTLD (n=4) and C9-ALS without FTLD (n=2) were combined to create the C9-ALS cohort (n=6). The sALS without FTLD (n=4) samples constituted the sALS cohort. These disease subtypes (C9-ALS and sALS) were then compared to the non-neurological control group (n=6).

Single nucleus RNA-seq by fluorescence activated cell sorting

Frozen orbitofrontal cortex (~50mg per sample) was dounce homogenized on ice in lysis buffer (0.32mM sucrose, 5mM $CaCl_2$, 3mM $Mg(Ac)_2$, 20mM Tris-HCl [pH 7.5], 0.1% Triton X-100, 0.5M EDTA [pH 8.0], 40U/mL RNase inhibitor in H_2O), centrifuged at 800×*g* for 10 min at 4°C. The supernatant was removed and the pellet was washed twice and resuspended in a resuspension buffer (1x PBS, 1% BSA, 0.2U/µL RNase inhibitor). Resuspended nuclei were sorted by fluorescence activated cell sorting with DAPI (Roche) labeling, removing any debris and nuclei aggregates within DAPI-positive gating, capturing ~6000 nuclei per sample. Library preparation was performed using either the 10X Chromium Single Cell 3' v2 or v3 platform following the manufacturer protocols. The QC of cDNA libraries was conducted on a 2100 Bioanalyzer (Agilent). The cDNA libraries were 100-bp paired end sequenced on either an Illumina NovaSeq 6000 SP XP or NovaSeq6000 S2 standard flow cell at the Princess Margaret Genomic Centre (Toronto, Ontario). Raw

Illumina base call files from each sample were demultiplexed to produce FASTQ files with the *cellranger mkfastq* pipeline (10X Genomics). Reads were aligned to the pre-mRNA GRCh38-2020-A genome and quantified using *cellranger count* command on Cell Ranger v5.0.0 (10X Genomics).

Single nucleus ATAC-seq by density gradient

Frozen orbitofrontal cortex (~50mg per sample) was dounce homogenized on ice in homogenization buffer (HB, 260mM sucrose, 30mM KCl, 10mM MgCl2, 20mM Tricine-KOH [pH 7.8], 1mM DTT, 0.5mM Spermidine, 0.15mM Spermine, 0.3% NP40, 1 tablet cOmplete Protease inhibitor [Sigma-Aldrich]) and then filtered using a 40µm Flowmi cell strainer. Filtered nuclei were spun down at 500×*g* for 5 min at 4°C. Following removal of supernatant, nuclei were resuspended in 400µL of HB and then mixed with 50% iodixanol (Sigma-Aldrich). The final nuclei suspension at 25% iodixanol was loaded onto a 30-40% iodixanol density gradient and then centrifuged at 3200×*g* for 20 min at 4°C. Nuclei collected from the 30-40% iodixanol interface were resuspended in HB and centrifuged at 500×*g* for 10 min at 4°C. Supernatant was removed and the gradient purified nuclei were resuspended in diluted nuclei buffer (10X Genomics). Library preparation was performed following manufacturer instructions on the 10X Chromium Single Cell ATAC v1.1 platform. The QC of cDNA libraries was conducted on a 2100 Bioanalyzer (Agilent). The cDNA libraries were 100-bp paired end sequenced on either an Illumina NovaSeq 6000 SP XP or NovaSeq6000 S2 standard flow cell at the Princess Margaret Genomic Centre (Toronto, Ontario). Raw Illumina base call files from each sample were demultiplexed to produce FASTQ files with the *cellranger-atac mkfastq* pipeline (10X Genomics). Reads were aligned to the GRCh38-2020-A genome and quantified using *cellranger-atac count* command on Cell Ranger ATAC v2.0 (10X Genomics).

snRNA-seq sample processing, QC, and clustering

All processing and QC of snRNA-seq samples was performed using Seurat (v4.0), and custom R and Python scripts. Seurat objects were created for each sample using the filtered feature-barcode matrices obtained from Cell Ranger (v5.0.0). For each sample, nuclei containing mitochondrial reads with a threshold greater than three mean absolute deviations from the median number of mitochondrial reads with a maximum cut-off of 5% were removed. Next, nuclei with fewer than 200 and greater than 12000 detected genes were removed. Reads pertaining to the cell cycle were scored using the scran R package[125]. Potential doublets were estimated and removed using scDblFinder with default parameters[126]. The remaining singlet transcriptomes were merged and batch corrected with Harmony[34] on *log1p* normalized counts. Dimensionality reduction was performed using principal component analysis (PCA) on 50 PCs and then visualized using UMAP[127]. Seurat clustering was approximated using a resolution of 0.6. Clusters with fewer than 200 cells were filtered out as background as shown previously[128] based on poor representation across samples and disease subtypes.

snATAC-seq sample filtering, quality control, and clustering

Filtering and QC was performed on snATAC-seq samples using the "Merging Objects" vignette from Signac (v1.7.0)[32]. Peak sets were read from BED files produced by Cell Ranger ATAC (v2.0) followed by conversion to genomic ranges and then reduced to a common peak set shared by all samples. Fragment files were loaded and filtered using a 500 count cut-off to filter out low quality barcodes. The *CreateChromatinAssay* step was performed prior to creating a Seurat object for each sample. All samples were then merged and an additional filtering step was performed to remove barcodes from nuclei containing fewer than 2000 and greater than 25000 peak fragments in a given region, less than 15% reads in a given peak, less than a nucleosomal signal of 4, and less than 2 or greater than 20 TSS enrichment score[32]. On the merged object, matrix normalization was performed with the *RunTFIDF* function followed by singular value decomposition using *RunSVD* prior to visualization by UMAP. Doublets and multiplets were then filtered out from the merged object using the ATAC-seq MULtiplet Estimation Tool (AMULET)[129]. We removed clusters with fewer than 250 cells as background and a doublet cluster which was identified during annotation, as performed previously[128] (Morabito 2021). Batch correction on the dataset was performed using Harmony[34]. Re-clustering was then performed with Seurat at resolution of 0.6 and visualized again by UMAP. From here, peaks on the merged object were called using MACS2[130]. A gene activity matrix was added to the object using the *GeneActivity* command in Signac. This approach estimates gene expression by counting ATAC-seq peaks within the gene body and 2 kilobases upstream of the TSS. The resulting gene activity matrix was normalized by log1p

transformation. QC visualization was performed using the ArchR package[131] and any additional visualization was performed using the scCustomize[132] or dittoSeq[133] R packages.

Cluster annotation by machine learning and reference-based approaches

For both the snRNA-seq and snATAC-seq datasets, clustering was performed using the Leiden algorithm[35]. To uncover orbitofrontal cell types from the identified clusters, we employed marker discovery by machine learning (NSForest v3.9.1)[37] and reference-based annotation[36] using the Allen Brain Atlas. Default parameters in NSForest were applied to identify binary markers for each cluster. The machine learning-based marker classification approach identified binary markers for each cluster using random forest feature selection and expression scoring. These binary markers for clusters were then used to confirm cortical cell type identity using reference-based annotation. In the case of the snATAC-seq dataset, the gene activity matrix was used as input for NSForest. Markers from each dataset were then compared with reference 10X Genomics experiments from two cortical regions in humans, including the primary motor cortex and medial temporal gyrus from the Seattle Alzheimer Disease Cell Atlas (https://portal.brain-map.org/atlases-and-data/rnaseq). Markers were additionally compared to the results of a rank-sum Wilcoxon and auROC analysis with the R package Presto[134]. Final confirmation of cell type identity was confirmed for each cell subtype in the snRNA-seq and snATAC-seq (both the open chromatin peaks and gene activity matrix data) datasets by manual annotation[36] using canonical cortical cell markers.

Differential gene expression analysis

DEGs were uncovered using DESeq2[39]. Based on a comprehensive comparison of most single cell differential expression analyses, pseudobulk approaches were shown to outperform other methods[135,136]. DESeq2 has been shown to preserve biological variability and expected effect sizes in a single cell context[137]. For the DEG analysis, only the 10X v3 samples were included in the analysis to account for differences in the number of detected genes and UMI distribution. The counts matrix for each of these samples was aggregated for a given cell type, creating a pseudobulk RNA count for each cell type by biological replicate. First, a Single Cell Experiment (SCE) object[138] was created using the counts matrix and metadata from the snRNA-seq object from above created with Seurat. Any clusters with fewer than 10 cells in a given sample were removed from the SCE object. The resulting counts matrix was aggregated using the *aggregate.Matrix* command from the Matrix.utils package (v0.9.8). A DESeq2 object was created with the design ~ *diagnosis* + *sex*. Dispersion estimates and coefficients were all inspected manually prior to shrinking log2 fold changes to accurately estimate effect size using the Apeglm package[139]. An adjusted p-value < 0.05 was used to detect significance in all cases. Further analysis of transcriptional signatures within cell types and across disease subtypes was performed using the R package UCell version 2.4[140]. UCell uses the Mann-Whitney U statistic to assign and rank signature scores from a set of genes based on their relative expression in cells.

Transcription factor motif and differentially accessible region analysis

DARs were determined with the *FindMarkers* function in Seurat using the logistic regression (LR) algorithm. Latent variables included sex and normalized counts in peaks for DARs. An adjusted p-value < 0.01 was used as the threshold to determine significance for DARs. TF motif accessibility was computed on the snATAC-seq data using the R package chromVAR[66], which enables the calculation of TF motif variability in sparse chromatin data. In brief, the motif position frequency matrix (pfm) was the human reference from the cisBP database[141] to produce a motif object which was then added to the Seurat object using the *AddMotif* function in Signac. TF motif activity was then calculated on the snATAC-seq object using the *RunChromVAR* function (GRCh38/hg38 reference genome assembly). Differential TF motifs were uncovered using the Wilcoxon Rank Sum Test algorithm in Seurat's *FindMarkers* command. The fold change value was changed to represent the average difference given to match the chromVAR output. Statistically significant TF motifs showed at least an average difference between conditions of 0.5 and an adjusted p-value<0.01. Overlap between differential results for DEG, DARs, and TF motifs were visualized using either the R packages ggvenn[142] or UpSetR[143].

Peak annotation and chromatin co-accessibility analysis

The annotation of peak and TF motif locations relative to the transcriptional start site TSS were performed with the ChIPseeker[56], and included regions ranging from promoters, 5′ UTR, 3′ UTR, exonic, intronic, and

distal intergenic. To identify co-accessible peaks in the snATAC-seq data, we computed co-accessibility peak scores using the R package Cicero[38]. In brief, Cicero takes the existing UMAP embeddings to cluster cells together, whereby the algorithm computes a regularized correlation matrix from binary chromatin accessibility values using a graphical LASSO approach. Here, the LASSO penalizes distant sites at a maximum of 500kb versus proximal sites. A *cis*-co-accessibility network (CCAN) is plotted using a given peak as a node and edges are co-accessible peaks connected from the node. An empirical CCAN score threshold of >25% of maximum score was set for each cell subtype.

APA Quantification and Profiling

Pseudo-bulk alignment files were generated for each cell type using the barcodes. We employed the MAAPER software to assign sequencing reads to known polyadenylation (poly A) sites, as defined in the PolyA_DB_v3 database[80,81]. Polyadenylation sites are considered only if there were 25 reads aligning to the sites. For identifying genes with significant changes in the length of their 3′-most exon, we used the REDu metric provided by MAAPER. REDu measures the relative expression levels between the two most differentially expressed isoforms in the 3′-most exon. A positive REDu value indicates transcript lengthening events, while a negative value points to shortening events. For pinpointing genes exhibiting intronic APA usage we used the REDi metric. REDi compares the relative expression levels of the top differentially expressed isoform in the 3′-most exon and the top differentially expressed isoform in an intron or internal exon. The RED score, comparing conditions 1 and 2, is computed using the formula:

$$RED = log2(\alpha 2d/\alpha 2p) - log2(\alpha 1d/\alpha 1p).$$

Where, the proportions of the proximal PAs in conditions 1 and 2 are denoted as α1p and α2p, and the proportions of the distal PAS are denoted as α1d and α2d[80,81].

Subsequently, we used the APAlog package to delve deeper into differential poly(A) site usage patterns. APAlog operates on the normalized counts of reads mapped to each poly(A) site to assess the extent and nature of differential usage. For a comprehensive comparison, APAlog was run in Pairwise Test mode, which enables the comparison of all possible pairs of poly(A) sites per transcript[104].

Gene Ontology Analysis

Gene Ontology (GO) analysis on snATAC-seq data was carried out using the rGREAT package[144]. All GO analyses for snRNA-seq data, including from DEGs and APA was performed using GSEA[145] with curated, no GO inferred electronic annotation gene sets downloaded from http://baderlab.org/GeneSets/ (June 2023 release). We used a minimum gene set size of 15 and the maximum gene set size of 200. Visualization of GO results focused on minimizing GO term redundancy using either individual enrichment or bar plots of select representative findings, or by summarizing results by plotting enrichment maps created with the Enrichment Map plugin for Cytoscape (v3.9.1) in Linux[146]. A q-value cutoff of 0.5 was used for plotting APA GO results.

Deep Learning Model Architecture

For the APA-Net architecture, we designed a state-of-the-art architecture aimed at dissecting the complex regulatory mechanism underlying APA in both C9-ALS and sALS cases. APA-Net employs a Convolutional Neural Network CNN architecture supplemented with a multi-head attention module, specifically optimized to identify cis-regulatory elements impacting PA site selection across varied cell types. The input region for the CNN and MAT modules encompasses 2 kb surrounding both the proximal and distal PAS. The CNN module contains a single convolutional layer, comprising 128 kernels, each with a size of 12 and a stride of 1. This is followed by a max pooling layer with a kernel size of 20 and stride of 20. The output from this stage feeds into a multi-head attention module, in which each position in the representation map functions as a distinct token . In this context, a 'token' refers to a discrete unit of information, which is essential for the attention mechanism to effectively process and interpret the complex patterns within the data. A residual connection links the CNN and multi-head attention modules. This residual connection, a key component in deep learning architectures, helps in mitigating the vanishing gradient problem by allowing the flow of information and gradients directly

across layers. The attention module's output, along with the RBP expression profile specific to the cell type, is forwarded to a multi-layer perceptron for final APA effect prediction through a regression task.

Deep Learning Model Interpretation

To extract learned RNA sequence motifs, we used the learned filter weights in the CNN module of APA-Net. Each sequence within the test dataset was scanned using the model's filters to identify regions with maximal filter activation, which are hypothesized to correspond to biologically relevant RNA sequence motifs, which were modeled as Position-Weight Matrices. We next aligned RBPs from the compendium of RNA-binding motifs[92] with these RNA sequence motifs. After RBP alignment, we identified filter (RBP motif) profile correlation interactions involved in APA regulation. This is done through analysis of filter activations (RBP motifs) across the test dataset. We hypothesize that highly correlated filters may represent RBPs involved in the same pathways.

RBP interaction dissimilarity across cell types

We used the Frobenius norm, a measure of the difference between two matrices (similar to Euclidean distance for vectors) defined as the square root of the sum of the absolute squares of their element-wise differences to measure the dissimilarity of the interaction profiles across the celltypes.
The Frobenius distance between two matrices X and Y is given by:

$$d(X, Y) = ||X - Y||_F = \sqrt{(tr((X - Y)^{T}(X - Y)))}$$

Where:
tr denotes the trace of a matrix (the sum of its diagonal elements).
X^T denotes the transpose of matrix X.
$||\cdot||_F$ denotes the Frobenius norm.

Data and code availability

Raw FASTQ snRNA-seq and snATAC-seq files are deposited in the National Institute of Health Sequencing Read Archive SRA (https://www.ncbi.nlm.nih.gov/sra) under the accession ID (TBA). Source code used in this study will be made available on GitHub upon publication. Processed snRNA-seq and snATAC-seq data will be available for interaction with a R Shiny app upon publication.

Acknowledgments

The authors are thankful to the ALS patients and families for their tissue donations in this study as well as the Douglas-Bell Canada Brain Bank (McGill University). This research was enabled in part by access to high performance computing clusters provided through the Digital Research Alliance of Canada (alliancecan.ca). This work was funded by the James Hunter Family Initiative in ALS Research, a Project Grant from the ALS Society of Canada, and an ERA-LEARN (E-Rare-3/Canadian Institute for Health Research) grant (063-REPETOMICS). P.M. McKeever was supported by the ALS Association Milton Safenowitz Postdoctoral Fellowship (2019-2021) and currently holds the Christopher Chiu Fellowship from ALS Double Play. J. Robertson is the James Hunter Family Chair in ALS Research. E. Rogaeva is supported in part by the Canadian Consortium on Neurodegeneration in Aging. This work was supported by NRNB (U.S. National Institutes of Health, National Center for Research Resources grant number P41 GM103504)

References

1. Renton, A. E. *et al.* A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron* **72**, 257–268 (2011).

2. DeJesus-Hernandez, M. *et al.* Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron* **72**, 245–256 (2011).

3. McGoldrick, P. & Robertson, J. Unraveling the impact of disrupted nucleocytoplasmic transport systems in C9orf72-associated ALS. *Front. Cell. Neurosci.* **17**, 1247297 (2023).

4. Bauer, C. S. *et al.* An interaction between synapsin and C9orf72 regulates excitatory synapses and is impaired in ALS/FTD. *Acta Neuropathol.* **144**, 437–464 (2022).

5. Frick, P. *et al.* Novel antibodies reveal presynaptic localization of C9orf72 protein and reduced protein levels in C9orf72 mutation carriers. *Acta Neuropathol. Commun.* **6**, 72 (2018).

6. Xiao, S., McKeever, P. M., Lau, A. & Robertson, J. Synaptic localization of C9orf72 regulates post-synaptic glutamate receptor 1 levels. *Acta Neuropathol. Commun.* **7**, 161 (2019).

7. Unable to find information for 15312587.

8. Tollervey, J. R. *et al.* Characterizing the RNA targets and position-dependent splicing regulation by TDP-43. *Nat. Neurosci.* **14**, 452–458 (2011).

9. Xiao, S. *et al.* RNA targets of TDP-43 identified by UV-CLIP are deregulated in ALS. *Mol. Cell. Neurosci.* **47**, 167–180 (2011).

10. Liu, E. Y. *et al.* Loss of Nuclear TDP-43 Is Associated with Decondensation of LINE Retrotransposons. *Cell Rep.* **27**, 1409-1421.e6 (2019).

11. Lagier-Tourenne, C. *et al.* Divergent roles of ALS-linked proteins FUS/TLS and TDP-43 intersect in processing long pre-mRNAs. *Nat. Neurosci.* **15**, 1488–1497 (2012).

12. Luisier, R. *et al.* Intron retention and nuclear loss of SFPQ are molecular hallmarks of ALS. *Nat. Commun.* **9**, 2010 (2018).

13. Taylor, R. *et al.* Prematurely terminated intron-retaining mRNAs invade axons in SFPQ null-driven neurodegeneration and are a hallmark of ALS. *Nat. Commun.* **13**, 6994 (2022).

14. Mackenzie, I. R. *et al.* TIA1 mutations in amyotrophic lateral sclerosis and frontotemporal dementia promote phase separation and alter stress granule dynamics. *Neuron* **95**, 808-816.e9 (2017).

15. Humphrey, J., Emmett, W., Fratta, P., Isaacs, A. M. & Plagnol, V. Quantitative analysis of cryptic splicing associated with TDP-43 depletion. *BMC Med. Genomics* **10**, 38 (2017).

16. Koike, Y. *et al.* TDP-43 and other hnRNPs regulate cryptic exon inclusion of a key ALS/FTD risk gene, UNC13A. *PLoS Biol.* **21**, e3002028 (2023).

17. Berson, A. *et al.* TDP-43 Promotes Neurodegeneration by Impairing Chromatin Remodeling. *Curr. Biol.* **27**, 3579-3590.e6 (2017).

18. Tibshirani, M. *et al.* Dysregulation of chromatin remodelling complexes in amyotrophic lateral sclerosis. *Hum. Mol. Genet.* **26**, 4142–4152 (2017).

19. Amlie-Wolf, A. *et al.* Transcriptomic Changes Due to Cytoplasmic TDP-43 Expression Reveal Dysregulation of Histone Transcripts and Nuclear Chromatin. *PLoS ONE* **10**, e0141836 (2015).

20. Ayala, Y. M. *et al.* TDP-43 regulates its mRNA levels through a negative feedback loop. *EMBO J.* **30**, 277–288 (2011).

21. Avendaño-Vázquez, S. E. *et al.* Autoregulation of TDP-43 mRNA levels involves interplay between transcription, splicing, and alternative polyA site selection. *Genes Dev.* **26**, 1679–1684 (2012).

22. Rot, G. *et al.* High-Resolution RNA Maps Suggest Common Principles of Splicing and Polyadenylation Regulation by TDP-43. *Cell Rep.* **19**, 1056–1067 (2017).

23. Modic, M. *et al.* Cross-Regulation between TDP-43 and Paraspeckles Promotes Pluripotency-Differentiation Transition. *Mol. Cell* **74**, 951-965.e13 (2019).

24. Melamed, Z. *et al.* Premature polyadenylation-mediated loss of stathmin-2 is a hallmark of TDP-43-dependent neurodegeneration. *Nat. Neurosci.* **22**, 180–190 (2019).

25. Imaizumi, K., Ideno, H., Sato, T., Morimoto, S. & Okano, H. Pathogenic Mutation of TDP-43 Impairs RNA Processing in a Cell Type-Specific Manner: Implications for the Pathogenesis of ALS/FTLD. *eNeuro* **9**, (2022).

26. Butti, Z. & Patten, S. A. RNA dysregulation in amyotrophic lateral sclerosis. *Front. Genet.* **9**, 712 (2018).

27. Prudencio, M. *et al.* Distinct brain transcriptome profiles in C9orf72-associated and sporadic ALS. *Nat. Neurosci.* **18**, 1175–1182 (2015).

28. Ziff, O. J. *et al.* Nucleocytoplasmic mRNA redistribution accompanies RNA binding protein mislocalization in ALS motor neurons and is restored by VCP ATPase inhibition. *Neuron* (2023) doi:10.1016/j.neuron.2023.06.019.

29. Andrews, T. S., Kiselev, V. Y., McCarthy, D. & Hemberg, M. Tutorial: guidelines for the computational

analysis of single-cell RNA sequencing data. *Nat. Protoc.* **16**, 1–9 (2020).

30. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
31. Buenrostro, J. D. *et al.* Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**, 486–490 (2015).
32. Stuart, T., Srivastava, A., Madad, S., Lareau, C. A. & Satija, R. Single-cell chromatin state analysis with Signac. *Nat. Methods* **18**, 1333–1341 (2021).
33. Granja, J. M. *et al.* ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat. Genet.* **53**, 403–411 (2021).
34. Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).
35. Traag, V. A., Waltman, L. & van Eck, N. J. From Louvain to Leiden: guaranteeing well-connected communities. *Sci. Rep.* **9**, 5233 (2019).
36. Clarke, Z. A. *et al.* Tutorial: guidelines for annotating single-cell transcriptomic maps using automated and manual methods. *Nat. Protoc.* **16**, 2749–2764 (2021).
37. Aevermann, B. *et al.* A machine learning method for the discovery of minimum marker gene combinations for cell type identification from single-cell RNA sequencing. *Genome Res.* **31**, 1767–1780 (2021).
38. Pliner, H. A. *et al.* Cicero Predicts cis-Regulatory DNA Interactions from Single-Cell Chromatin Accessibility Data. *Mol. Cell* **71**, 858-871.e8 (2018).
39. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
40. Li, J. *et al.* Divergent single cell transcriptome and epigenome alterations in ALS and FTD patients with C9orf72 mutation. *Nat. Commun.* **14**, 5714 (2023).
41. Pineda, S. S. *et al.* Single-cell profiling of the human primary motor cortex in ALS and FTLD. *BioRxiv* (2021) doi:10.1101/2021.07.07.451374.

42. Limone, F. *et al.* Single-nucleus sequencing reveals enriched expression of genetic risk factors sensitises Motor Neurons to degeneration in ALS. *BioRxiv* (2021) doi:10.1101/2021.07.12.452054.

43. Richards, L. M. *et al.* Gradient of Developmental and Injury Response transcriptional states defines functional vulnerabilities underpinning glioblastoma heterogeneity. *Nat. Cancer* **2**, 157–173 (2021).

44. Sar, M. & Stumpf, W. E. Androgen concentration in motor neurons of cranial nerves and spinal cord. *Science* **197**, 77–79 (1977).

45. Weiner, L. P. Possible role of androgen receptors in amyotrophic lateral sclerosis. A hypothesis. *Arch. Neurol.* **37**, 129–131 (1980).

46. Olah, M. *et al.* Single cell RNA sequencing of human microglia uncovers a subset associated with Alzheimer's disease. *Nat. Commun.* **11**, 6129 (2020).

47. Keren-Shaul, H. *et al.* A Unique Microglia Type Associated with Restricting Development of Alzheimer's Disease. *Cell* **169**, 1276-1290.e17 (2017).

48. Sun, N. *et al.* Human microglial state dynamics in Alzheimer's disease progression. *Cell* **186**, 4386-4403.e29 (2023).

49. Liu, X. & Quan, N. Microglia and CNS Interleukin-1: Beyond Immunological Concepts. *Front. Neurol.* **9**, 8 (2018).

50. Peckert-Maier, K. *et al.* CD83 expressed by macrophages is an important immune checkpoint molecule for the resolution of inflammation. *Front. Immunol.* **14**, (2023).

51. Aibar, S. *et al.* SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods* **14**, 1083–1086 (2017).

52. Yamanaka, K. & Komine, O. The multi-dimensional roles of astrocytes in ALS. *Neurosci. Res.* **126**, 31–38 (2018).

53. Humphrey, J. *et al.* Integrative transcriptomic analysis of the amyotrophic lateral sclerosis spinal cord implicates glial activation and suggests new risk genes. *Nat. Neurosci.* **26**, 150–162 (2023).

54. Zamanian, J. L. *et al.* Genomic analysis of reactive astrogliosis. *J. Neurosci.* **32**, 6391–6410 (2012).

55. Pekny, M. & Nilsson, M. Astrocyte activation and reactive gliosis. *Glia* **50**, 427–434 (2005).

56. Yu, G., Wang, L.-G. & He, Q.-Y. ChIPseeker: an R/Bioconductor package for ChIP peak annotation,

comparison and visualization. *Bioinformatics* **31**, 2382–2383 (2015).

57. Shibata, N. *et al.* Activation of signal transducer and activator of transcription-3 in the spinal cord of sporadic amyotrophic lateral sclerosis patients. *Neurodegener Dis* **6**, 118–126 (2009).

58. Hasan, R. *et al.* Transcriptomic analysis of frontotemporal lobar degeneration with TDP-43 pathology reveals cellular alterations across multiple brain regions. *Acta Neuropathol.* **143**, 383–401 (2022).

59. Pang, W. & Hu, F. C9ORF72 suppresses JAK-STAT mediated inflammation. *iScience* **26**, 106579 (2023).

60. Goutman, S. A. *et al.* Emerging insights into the complex genetics and pathophysiology of amyotrophic lateral sclerosis. *Lancet Neurol.* **21**, 465–479 (2022).

61. van Rheenen, W. *et al.* Common and rare variant association analyses in amyotrophic lateral sclerosis identify 15 risk loci with distinct genetic architectures and neuron-specific biology. *Nat. Genet.* **53**, 1636–1648 (2021).

62. Rosen, D. R. *et al.* Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis. *Nature* **362**, 59–62 (1993).

63. McGoldrick, P. *et al.* Unaffected mosaic C9orf72 case. *Neurology* **90**, e323–e331 (2018).

64. Shi, Y. *et al.* Haploinsufficiency leads to neurodegeneration in C9ORF72 ALS/FTD human induced motor neurons. *Nat. Med.* **24**, 313–325 (2018).

65. Liu, E. Y., Russ, J. & Lee, E. B. Neuronal Transcriptome from C9orf72 Repeat Expanded Human Tissue is Associated with Loss of C9orf72 Function. *Free Neuropathol* **1**, (2020).

66. Schep, A. N., Wu, B., Buenrostro, J. D. & Greenleaf, W. J. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat. Methods* **14**, 975–978 (2017).

67. Stevanovic, M. *et al.* SOX transcription factors as important regulators of neuronal and glial differentiation during nervous system development and adult neurogenesis. *Front. Mol. Neurosci.* **14**, 654031 (2021).

68. Lozzi, B., Huang, T.-W., Sardar, D., Huang, A. Y.-S. & Deneen, B. Regionally distinct astrocytes display unique transcription factor profiles in the adult brain. *Front. Neurosci.* **14**, 61 (2020).

69. Patel, M. *et al.* Gsx1 promotes locomotor functional recovery after spinal cord injury. *Mol. Ther.* **29**, 2469–2482 (2021).

70. Smith, A. M. *et al.* The transcription factor PU.1 is critical for viability and function of human brain microglia. *Glia* **61**, 929–942 (2013).

71. Kierdorf, K. *et al.* Microglia emerge from erythromyeloid precursors via Pu.1- and Irf8-dependent pathways. *Nat. Neurosci.* **16**, 273–280 (2013).

72. Chen, F. *et al.* The transcription factor NeuroD2 coordinates synaptic innervation and cell intrinsic properties to control excitability of cortical pyramidal neurons. *J Physiol (Lond)* **594**, 3729–3744 (2016).

73. Hou, P.-S., Miyoshi, G. & Hanashima, C. Sensory cortex wiring requires preselection of short- and long-range projection neurons through an Egr-Foxg1-COUP-TFI network. *Nat. Commun.* **10**, 3581 (2019).

74. Singh, A. *et al.* Tcf12 and NeuroD1 cooperatively drive neuronal migration during cortical development. *Development* **149**, (2022).

75. Corces, M. R. *et al.* The chromatin accessibility landscape of primary human cancers. *Science* **362**, (2018).

76. Olson, J. M. *et al.* NeuroD2 is necessary for development and survival of central nervous system neurons. *Dev. Biol.* **234**, 174–187 (2001).

77. Dastidar, S. G., Landrieu, P. M. Z. & D'Mello, S. R. FoxG1 promotes the survival of postmitotic neurons. *J. Neurosci.* **31**, 402–413 (2011).

78. Hidano, S. *et al.* STAT1 signaling in astrocytes is essential for control of infection in the central nervous system. *MBio* **7**, (2016).

79. Howe, C. L., Mayoral, S. & Rodriguez, M. Activated microglia stimulate transcriptional changes in primary oligodendrocytes via IL-1beta. *Neurobiol. Dis.* **23**, 731–739 (2006).

80. Li, W. V., Zheng, D., Wang, R. & Tian, B. MAAPER: model-based analysis of alternative polyadenylation using 3' end-linked reads. *Genome Biol.* **22**, 222 (2021).

81. Wang, R., Zheng, D., Yehia, G. & Tian, B. A compendium of conserved cleavage and polyadenylation

events in mammalian genes. *Genome Res.* **28**, 1427–1441 (2018).

82. O'Brien, J., Hayder, H., Zayed, Y. & Peng, C. Overview of microrna biogenesis, mechanisms of actions, and circulation. *Front Endocrinol (Lausanne)* **9**, 402 (2018).

83. Chen, C. Y. & Shyu, A. B. AU-rich elements: characterization and importance in mRNA degradation. *Trends Biochem. Sci.* **20**, 465–470 (1995).

84. Ma, X. R. *et al.* TDP-43 represses cryptic exon inclusion in the FTD-ALS gene UNC13A. *Nature* **603**, 124–130 (2022).

85. Brown, A.-L. *et al.* TDP-43 loss and ALS-risk SNPs drive mis-splicing and depletion of UNC13A. *Nature* **603**, 131–137 (2022).

86. Ranganathan, S. & Bowser, R. Alterations in G(1) to S phase cell-cycle regulators during amyotrophic lateral sclerosis. *Am. J. Pathol.* **162**, 823–835 (2003).

87. Galán, L., Gómez-Pinedo, U., Guerrero, A., García-Verdugo, J. M. & Matías-Guiu, J. Amyotrophic lateral sclerosis modifies progenitor neural proliferation in adult classic neurogenic brain niches. *BMC Neurol.* **17**, 173 (2017).

88. Chua, J. P., De Calbiac, H., Kabashi, E. & Barmada, S. J. Autophagy and ALS: mechanistic insights and therapeutic implications. *Autophagy* **18**, 254–282 (2022).

89. Smith, E. F., Shaw, P. J. & De Vos, K. J. The role of mitochondria in amyotrophic lateral sclerosis. *Neurosci. Lett.* **710**, 132933 (2019).

90. Koo, P. K. & Eddy, S. R. Representation learning of genomic sequence motifs with convolutional neural networks. *PLoS Comput. Biol.* **15**, e1007560 (2019).

91. Ghotra, R., Lee, N. K., Tripathy, R. & Koo, P. K. Designing Interpretable Convolution-Based Hybrid Networks for Genomics. *BioRxiv* (2021) doi:10.1101/2021.07.13.452181.

92. Ray, D. *et al.* A compendium of RNA-binding motifs for decoding gene regulation. *Nature* **499**, 172–177 (2013).

93. Masuda, A. *et al.* Position-specific binding of FUS to nascent RNA regulates mRNA length. *Genes Dev.* **29**, 1045–1057 (2015).

94. Shenouda, M., Zhang, A. B., Weichert, A. & Robertson, J. Mechanisms Associated with TDP-43

Neurotoxicity in ALS/FTLD. *Adv. Neurobiol.* **20**, 239–263 (2018).

95. Nasiri-Aghdam, M., Garcia-Garduño, T. C. & Jave-Suárez, L. F. CELF Family Proteins in Cancer: Highlights on the RNA-Binding Protein/Noncoding RNA Regulatory Axis. *Int. J. Mol. Sci.* **22**, (2021).

96. Shkurin, A., Pour, S. E. & Hughes, T. R. Known sequence features explain half of all human gene ends. *NAR Genom. Bioinform.* **5**, lqad031 (2023).

97. Feng, H. *et al.* Complexity and graded regulation of neuronal cell-type-specific alternative splicing revealed by single-cell RNA sequencing. *Proc Natl Acad Sci USA* **118**, (2021).

98. Fischl, H. *et al.* hnRNPC regulates cancer-specific alternative cleavage and polyadenylation profiles. *Nucleic Acids Res.* **47**, 7580–7591 (2019).

99. Schwich, O. D. *et al.* SRSF3 and SRSF7 modulate 3'UTR length through suppression or activation of proximal polyadenylation sites and regulation of CFIm levels. *Genome Biol.* **22**, 82 (2021).

100. Goodarzi, H. *et al.* Systematic discovery of structural elements governing stability of mammalian messenger RNAs. *Nature* **485**, 264–268 (2012).

101. Kowalski, M. H. *et al.* CPA-Perturb-seq: Multiplexed single-cell characterization of alternative polyadenylation regulators. *BioRxiv* (2023) doi:10.1101/2023.02.09.527751.

102. Anvar, S. Y. *et al.* Full-length mRNA sequencing uncovers a widespread coupling between transcription initiation and mRNA processing. *Genome Biol.* **19**, 46 (2018).

103. Simonelig, M. PABPN1 shuts down alternative poly(A) sites. *Cell Res.* **22**, 1419–1421 (2012).

104. Navickas, A. *et al.* An mRNA processing pathway suppresses metastasis by governing translational control from the nucleus. *BioRxiv* (2021) doi:10.1101/2021.10.04.463118.

105. Ishigaki, S. *et al.* Aberrant interaction between FUS and SFPQ in neurons in a wide range of FTLD spectrum diseases. *Brain* **143**, 2398–2405 (2020).

106. Malik, A. M. & Barmada, S. J. Matrin 3 in neuromuscular disease: physiology and pathophysiology. *JCI Insight* **6**, (2021).

107. Zheng, D. *et al.* Cellular stress alters 3'UTR landscape through alternative polyadenylation and isoform-specific degradation. *Nat. Commun.* **9**, 2268 (2018).

108. Ke, H. *et al.* Loss of TDP43 inhibits progression of triple-negative breast cancer in coordination with

SRSF3. *Proc Natl Acad Sci USA* **115**, E3426–E3435 (2018).

109. Couthouis, J. *et al.* A yeast functional screen predicts new candidate ALS disease genes. *Proc Natl Acad Sci USA* **108**, 20881–20890 (2011).

110. Yu, Z. *et al.* Neurodegeneration-associated TDP-43 interacts with fragile X mental retardation protein (FMRP)/Staufen (STAU1) and regulates SIRT1 expression in neuronal cells. *J. Biol. Chem.* **287**, 22560–22572 (2012).

111. Wang, I.-F., Wu, L.-S., Chang, H.-Y. & Shen, C.-K. J. TDP-43, the signature protein of FTLD-U, is a neuronal activity-responsive factor. *J. Neurochem.* **105**, 797–806 (2008).

112. Lomen-Hoerth, C. Characterization of amyotrophic lateral sclerosis and frontotemporal dementia. *Dement. Geriatr. Cogn. Disord.* **17**, 337–341 (2004).

113. Meier, S. L., Charleston, A. J. & Tippett, L. J. Cognitive and behavioural deficits associated with the orbitomedial prefrontal cortex in amyotrophic lateral sclerosis. *Brain* **133**, 3444–3457 (2010).

114. Lillo, P., Garcin, B., Hornberger, M., Bak, T. H. & Hodges, J. R. Neurobehavioral features in frontotemporal dementia with amyotrophic lateral sclerosis. *Arch. Neurol.* **67**, 826–830 (2010).

115. Lillo, P., Mioshi, E., Zoing, M. C., Kiernan, M. C. & Hodges, J. R. How common are behavioural changes in amyotrophic lateral sclerosis? *Amyotroph. Lateral Scler.* **12**, 45–51 (2011).

116. Tsujimoto, M. *et al.* Behavioral changes in early ALS correlate with voxel-based morphometry and diffusion tensor imaging. *J. Neurol. Sci.* **307**, 34–40 (2011).

117. Consonni, M., Cappa, S. F., Dalla Bella, E., Contarino, V. E. & Lauria, G. Cortical correlates of behavioural change in amyotrophic lateral sclerosis. *J. Neurol. Neurosurg. Psychiatr.* **90**, 380–386 (2019).

118. Crockford, C. *et al.* ALS-specific cognitive and behavior changes associated with advancing disease stage in ALS. *Neurology* **91**, e1370–e1380 (2018).

119. Bannwarth, S. *et al.* A mitochondrial origin for frontotemporal dementia and amyotrophic lateral sclerosis through CHCHD10 involvement. *Brain* **137**, 2329–2345 (2014).

120. Arefeen, A., Xiao, X. & Jiang, T. DeepPASTA: deep neural network based polyadenylation site analysis. *Bioinformatics* **35**, 4577–4585 (2019).

121. Li, Z. *et al.* DeeReCT-APA: Prediction of Alternative Polyadenylation Site Usage Through Deep Learning. *Genomics Proteomics Bioinformatics* (2021) doi:10.1016/j.gpb.2020.05.004.

122. Gittings, L. M. *et al.* Cryptic exon detection and transcriptomic changes revealed in single-nuclei RNA sequencing of C9ORF72 patients spanning the ALS-FTD spectrum. *Acta Neuropathol.* **146**, 433–450 (2023).

123. Brooks, B. R., Miller, R. G., Swash, M., Munsat, T. L. & World Federation of Neurology Research Group on Motor Neuron Diseases. El Escorial revisited: revised criteria for the diagnosis of amyotrophic lateral sclerosis. *Amyotroph. Lateral Scler. Other Motor Neuron Disord.* **1**, 293–299 (2000).

124. Xi, Z. *et al.* Jump from pre-mutation to pathologic expansion in C9orf72. *Am. J. Hum. Genet.* **96**, 962–970 (2015).

125. Lun, A. T. L., McCarthy, D. J. & Marioni, J. C. A step-by-step workflow for low-level analysis of single-cell RNA-seq data [version 1; peer review: 5 approved with reservations]. *F1000Res.* **5**, (2016).

126. Germain, P.-L., Lun, A., Macnair, W. & Robinson, M. D. Doublet identification in single-cell sequencing data using scDblFinder. *F1000Res.* **10**, 979 (2021).

127. Becht, E. *et al.* Dimensionality reduction for visualizing single-cell data using UMAP. *Nat. Biotechnol.* **37**, 38–44 (2018).

128. Morabito, S. *et al.* Single-nucleus chromatin accessibility and transcriptomic characterization of Alzheimer's disease. *Nat. Genet.* **53**, 1143–1155 (2021).

129. Thibodeau, A. *et al.* AMULET: a novel read count-based method for effective multiplet detection from single nucleus ATAC-seq data. *Genome Biol.* **22**, 252 (2021).

130. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).

131. Granja, J. M. *et al.* ArchR: An integrative and scalable software package for single-cell chromatin accessibility analysis. *BioRxiv* (2020) doi:10.1101/2020.04.28.066498.

132. Marsh, S., Salmon, M. & Hoffman, P. samuel-marsh/scCustomize: Version 1.1.3. *Zenodo* (2023) doi:10.5281/zenodo.8169188.

133. Bunis, D. G., Andrews, J., Fragiadakis, G. K., Burt, T. D. & Sirota, M. dittoSeq: universal user-friendly

single-cell and bulk RNA sequencing visualization toolkit. *Bioinformatics* **36**, 5535–5536 (2021).

134. Korsunsky, I., Nathan, A., Millard, N. & Raychaudhuri, S. Presto scales Wilcoxon and auROC analyses to millions of observations. *BioRxiv* (2019) doi:10.1101/653253.

135. Soneson, C. & Robinson, M. D. Bias, robustness and scalability in single-cell differential expression analysis. *Nat. Methods* **15**, 255–261 (2018).

136. Squair, J. W. *et al.* Confronting false discoveries in single-cell differential expression. *Nat. Commun.* **12**, 5692 (2021).

137. Zimmerman, K. D., Espeland, M. A. & Langefeld, C. D. A practical solution to pseudoreplication bias in single-cell studies. *Nat. Commun.* **12**, 738 (2021).

138. Amezquita, R. A. *et al.* Orchestrating single-cell analysis with Bioconductor. *Nat. Methods* **17**, 137–145 (2020).

139. Zhu, A., Ibrahim, J. G. & Love, M. I. Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. *Bioinformatics* **35**, 2084–2092 (2019).

140. Andreatta, M. & Carmona, S. J. UCell: Robust and scalable single-cell gene signature scoring. *Comput. Struct. Biotechnol. J.* **19**, 3796–3798 (2021).

141. Weirauch, M. T. *et al.* Determination and inference of eukaryotic transcription factor sequence specificity. *Cell* **158**, 1431–1443 (2014).

142. Yan, L. & Yan, M. L. Package 'ggvenn.' (2021).

143. Conway, J. R., Lex, A. & Gehlenborg, N. UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics* **33**, 2938–2940 (2017).

144. Gu, Z. & Hübschmann, D. rGREAT: an R/bioconductor package for functional enrichment on genomic regions. *Bioinformatics* **39**, (2023).

145. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15545–15550 (2005).

146. Reimand, J. *et al.* Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, Cytoscape and EnrichmentMap. *Nat. Protoc.* **14**, 482–517 (2019).

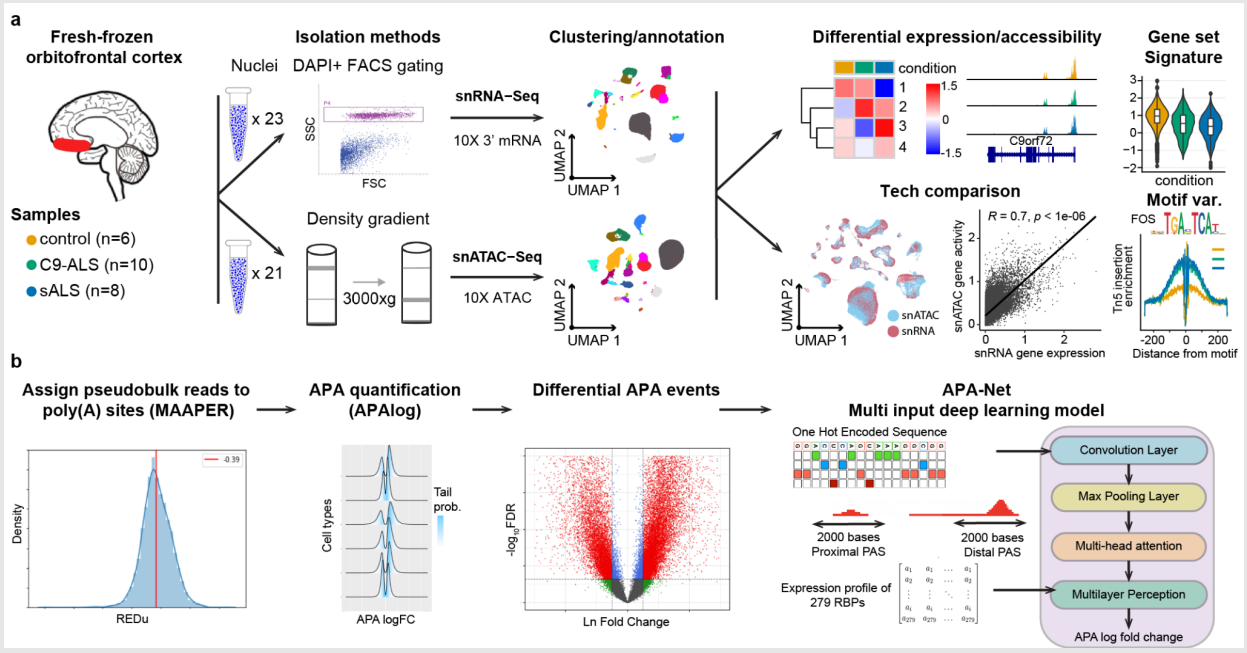


Figure 1. Epigenomic and transcriptomic atlas for the study of cell type changes in ALS brains. a, Workflow for the generation of a single nucleus atlas of the frontal cortex from C9-ALS (n=10 individuals), sALS (n=8), and control (n=6). The snRNA-seq data was derived from fluorescent activated cell sorted (FACS) nuclei labeled and gated with DAPI. The snATAC-seq data was obtained from nuclei purified by iodixanol density gradient. Analyses include clustering/annotation, differential expression/accessibility, gene set scoring, technology (tech) comparison and correlation, and motif variability (var.). **b**, Alternative polyadenylation (APA) analysis across ALS subtypes and cell types. The MAAPER software was used to assign pseudobulk snRNA-seq reads to poly(A) sites. The APAlog software facilitated quantification of APA events from assigned reads. APA grammar was decoded using a multi input deep learning model called APA-Net which consists of a convolutional neural network with a multi-head attention mechanism.

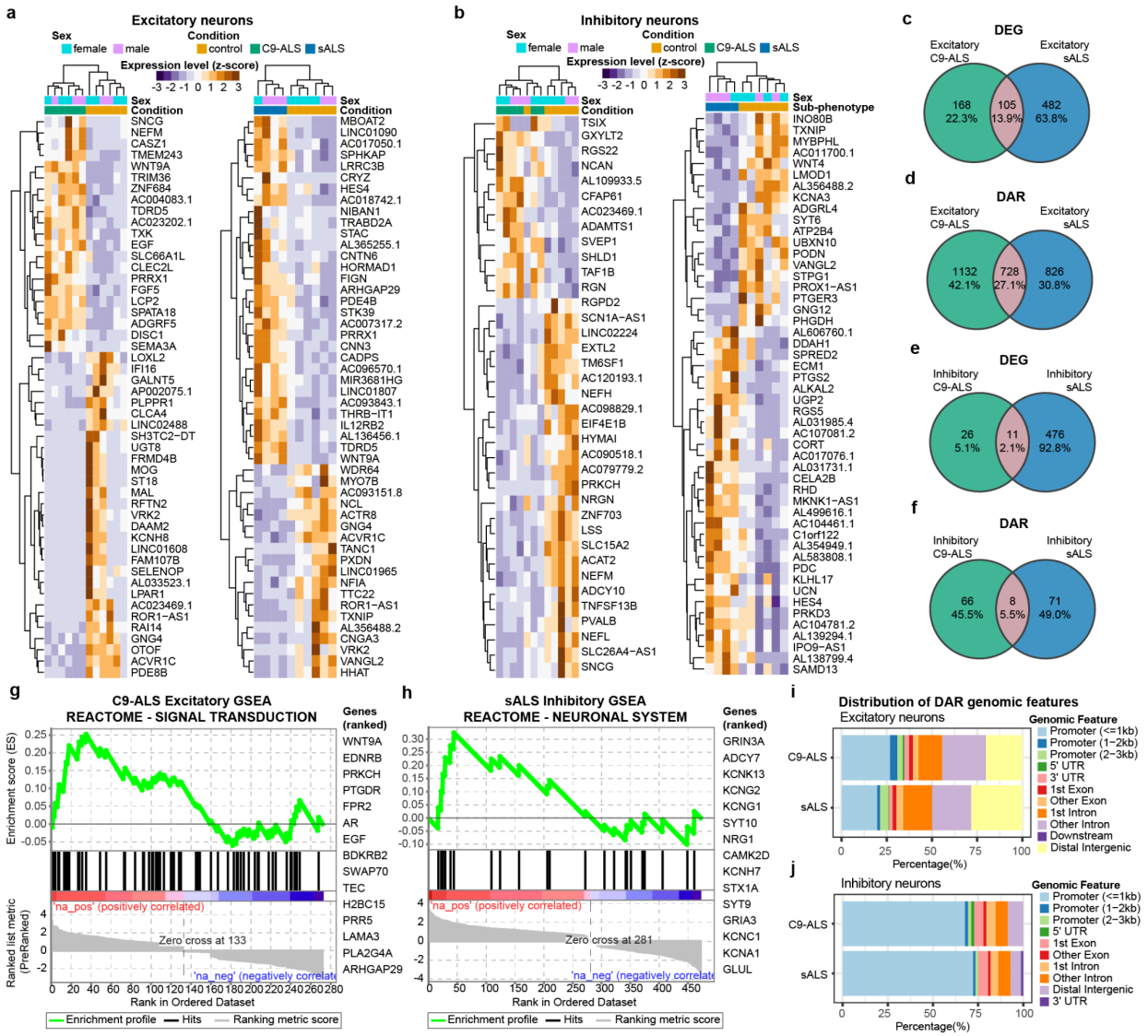
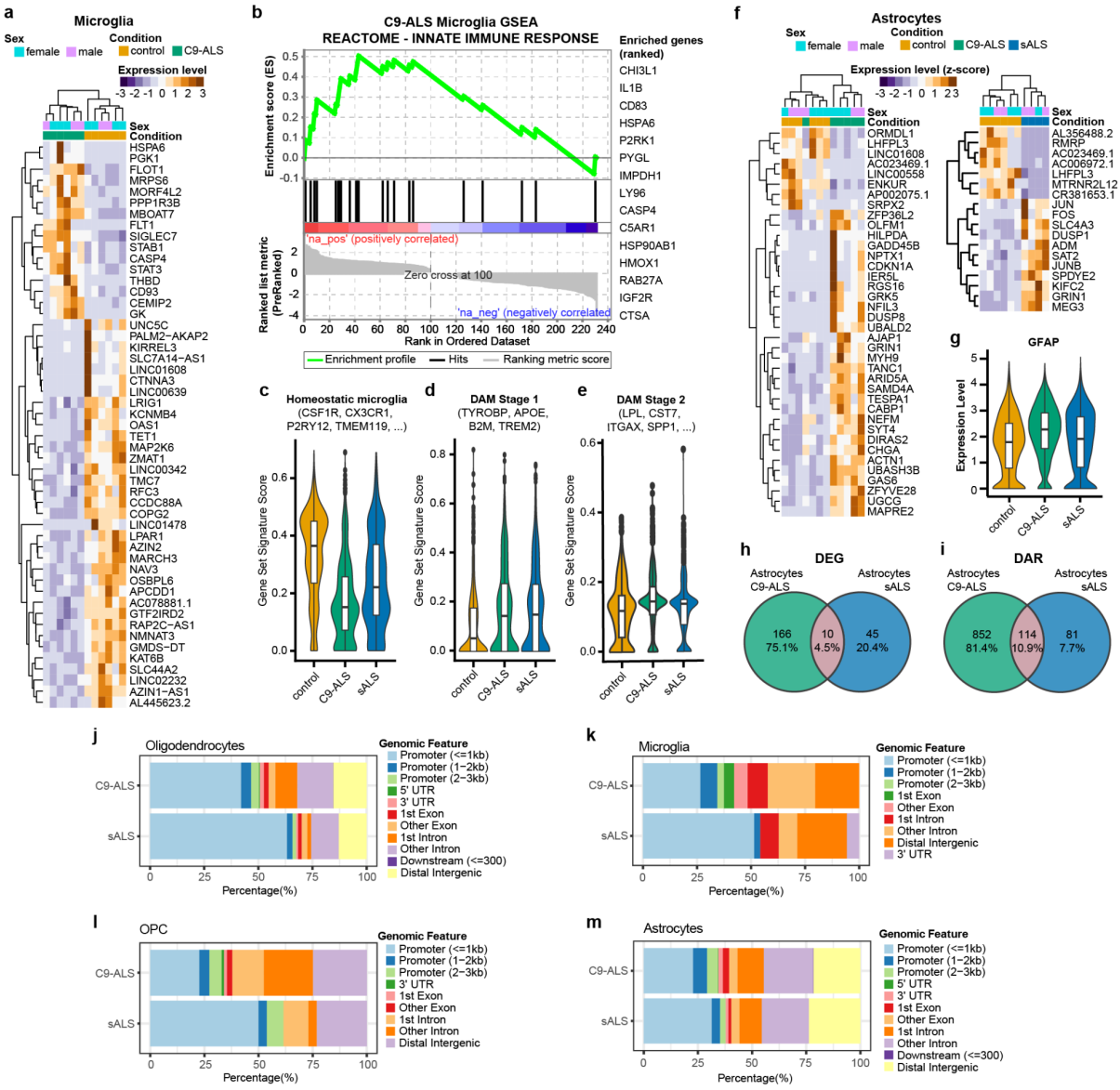
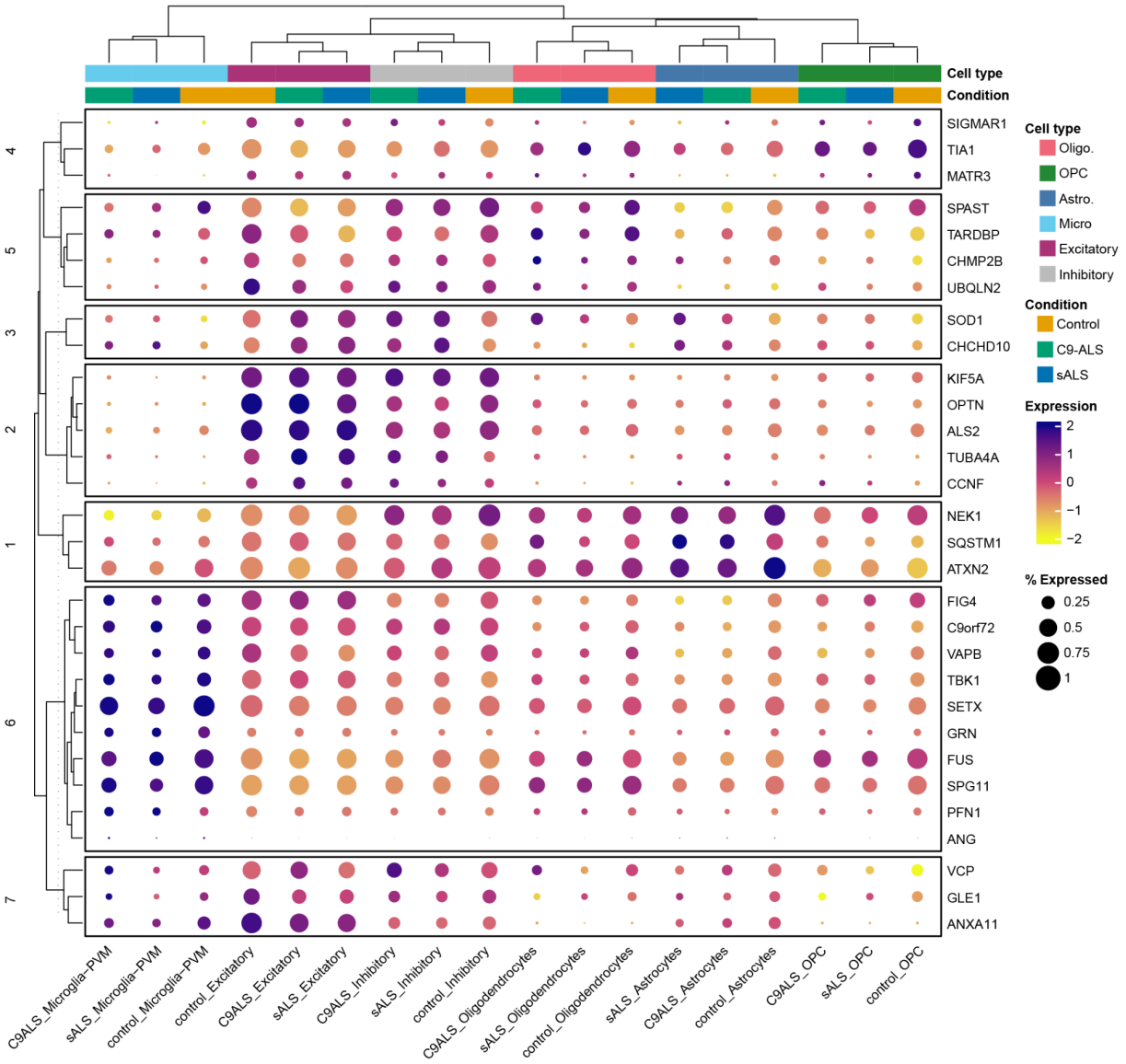


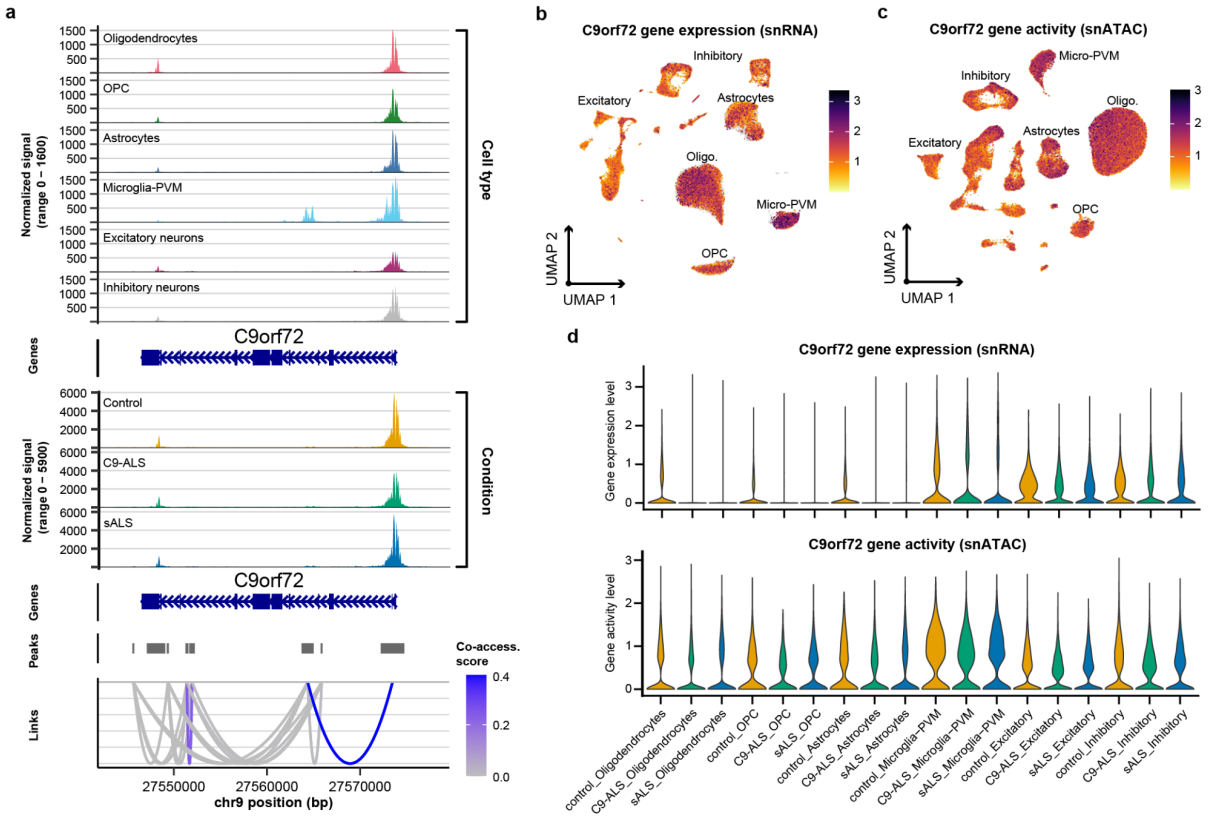
Figure 2. Neuronal transcriptomics and epigenomics show converging and diverging changes in C9-ALS and sALS versus controls. a, Heatmaps of top differentially expressed genes (DEG) uncovered using DESeq2 in excitatory neurons from C9-ALS (left) and sALS (right). **b**, Heatmaps of top DEG in inhibitory neurons from C9-ALS (left) and sALS (right). Venn diagrams indicating proportion of **c**, DEG overlap between C9-ALS and sALS in excitatory neurons; **d**, differentially accessible chromatin regions (DAR) in excitatory neurons between C9-ALS and sALS; **e**, DEG in inhibitory neurons comparing C9-ALS versus sALS; and **f**, DAR in inhibitory neuron from C9-ALS and sALS. **g**, Representative ranked gene set enrichment analysis (GSEA) plot from the Reactome database from DEG in C9-ALS. **h**, Ranked GSEA plot from inhibitory neurons in sALS. Genomic features distribution of DAR in **i**, excitatory neurons and **j**, inhibitory neurons from C9-ALS and sALS.



Extended Data Figure 1. Transcriptomics and epigenomics in glial cells from C9-ALS and sALS relative to controls. a, Heatmaps of top differentially expressed genes (DEG) uncovered using DESeq2 in microglia from C9-ALS. **b**, Representative ranked gene set enrichment analysis (GSEA) plot from the Reactome database from DEG in C9-ALS microglia. Violin and barplots of gene signature scores for **c**, homeostatic markers (*CX3CR1*, *CSF1R*, *P2RY12*, *HEXB*, *TMEM119*, *CSTC*, *SELPLG*, *TXNIP*, *CCR5*); **d**, disease associated microglia (DAM) stage 1 (*TYROBP*, *APOE*, *B2M*, *TREM2*); and **e**, DAM stage 2 (*LPL*, *CSTF*, *AXL*, *SPP1*, *CD9*, *CCL6*, *CSF1*). **f**, Heatmaps of top DEG in astrocytes from C9-ALS (left) and sALS (right). **g**, Violin and barplots of GFAP expression across disease subtypes. Venn diagrams indicating proportion of **h**, DEG overlap between C9-ALS and sALS in astrocytes; **i**, differentially accessible chromatin regions (DAR) in excitatory neurons between C9-ALS and sALS. Genomic features distribution of DARs in **j**, oligodendrocytes; **k**, oligodendrocyte precursor cells (OPCs); **l**, astrocytes; and **m**, microglia from C9-ALS and sALS.



Extended Data Figure 2. Expression of ALS-related genes across major cell types and disease subtypes. Clustered heatmap dot plot analysis comparing average ALS risk gene expression (y-axis) in major cell types and subtypes. Size of dot corresponds to percent of cells of a given cell type expressing the corresponding gene of interest. Cell type and condition identity indicated in the upper x-axis by colour coding and lower x-axis by label. Columns are clustered hierarchically, whereas rows are clustered and partitioned with k-means clustering.



Extended Data Figure 3. Cell type-specific enrichment of C9orf72 in frontal cortex. a, snATAC-seq gene coverage and link plots for major cell types (top coverage plot) and pseudobulk peaks (bottom coverage plot) for control and disease subtypes. **b**, UMAP plots demonstrating relative gene expression for C9orf72. **c**, UMAP plots demonstrating relative gene activity for C9orf72. **d**, Violin plots of C9orf72 snRNA-seq gene expression and snATAC-seq gene activity scores across major cell types and disease subtypes.

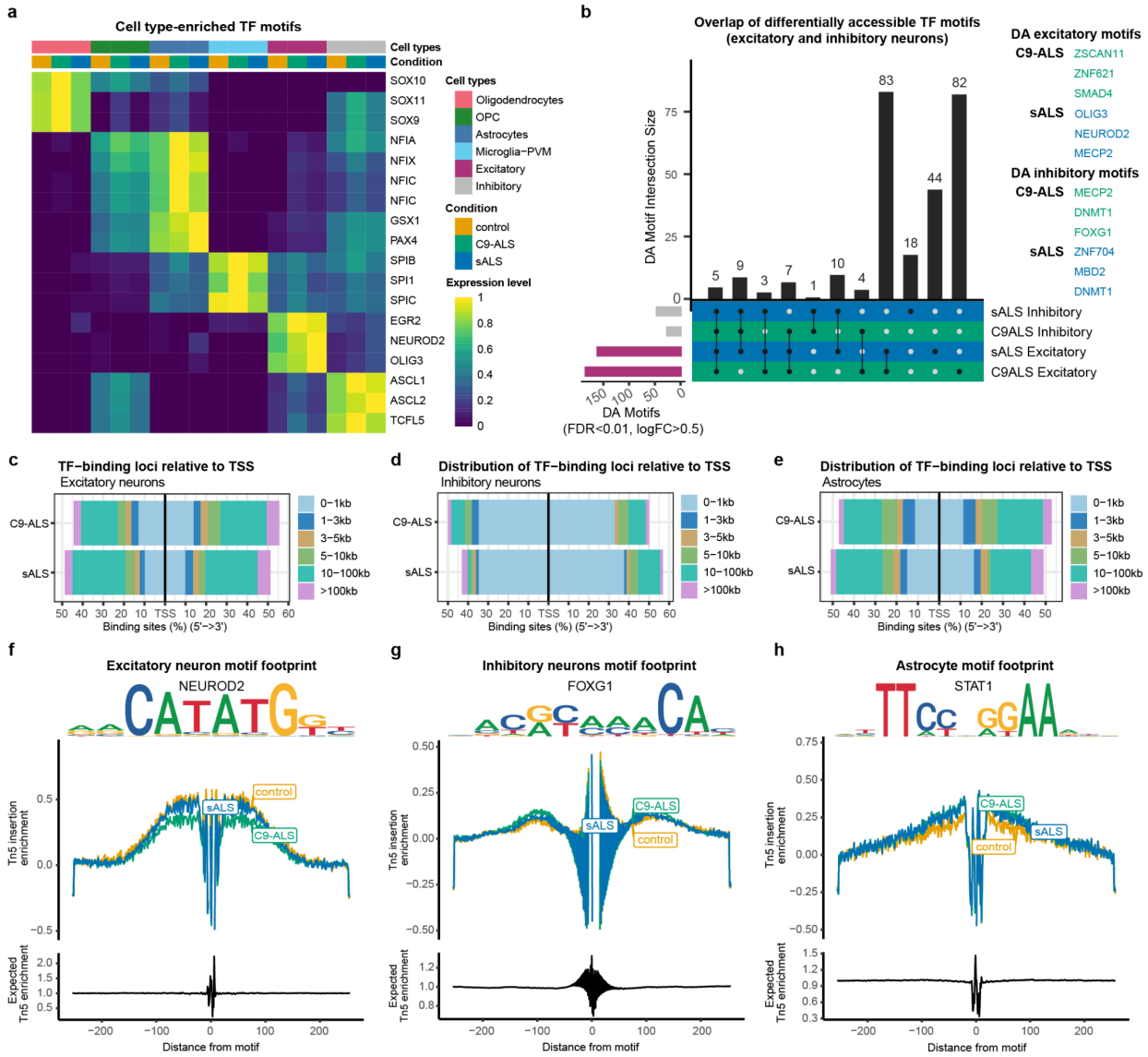
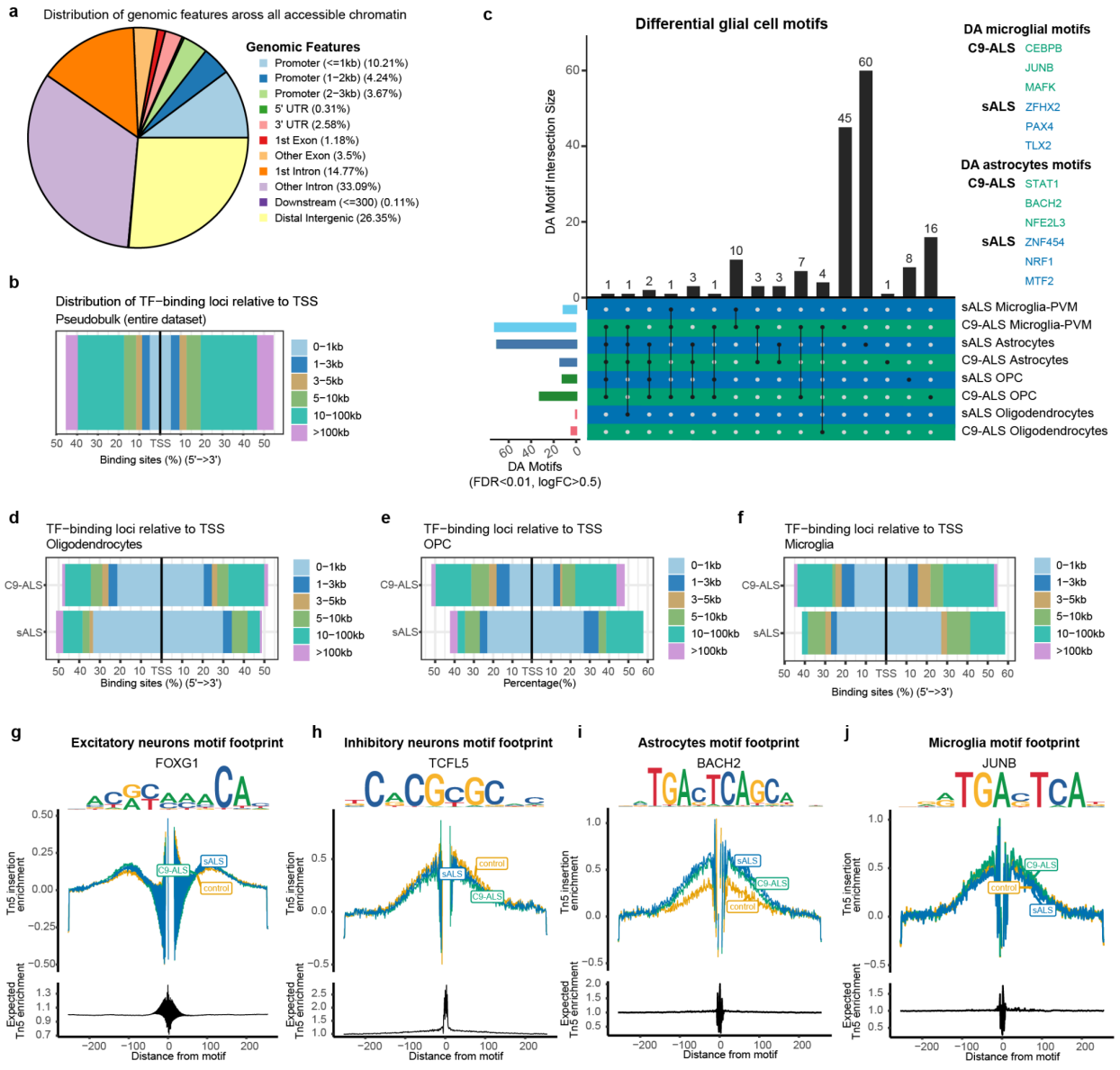


Figure 3. Transcription factor (TF) motif variability is altered across cell types in C9-ALS and sALS. a, Heatmap of average chromVAR deviation scores for the top 3 cell type-enriched motifs across disease subtypes. **b**, Upset plot showing the intersection of differentially accessible TF motifs in excitatory and inhibitory neurons from C9-ALS and sALS cohorts. Distribution of TF-binding loci relative to the transcriptional start site (TSS) in differentially accessible TF motifs in **c**, excitatory neurons, **d**, inhibitory neurons, and **e**, astrocytes. Transcription factor footprint analysis plots of enriched Tn5 insertion around **f**, NEUROD2 in excitatory neurons, **g**, FOXG1 in inhibitory neurons, and **h**, STAT1 in astrocytes.



Extended Data Figure 4. (Related to Fig. 3). Differentially altered transcription factor (TF) motifs in glial cells from C9-ALS and sALS. a, Pie chart representing the distribution of peaks in pseudobulk snATAC-seq dataset. **b**, TF-binding loci relative to the transcriptional start site (TSS) across the entire snATAC-seq dataset. **c**, Upset plot showing the intersection of differentially accessible TF motifs in glial cells from C9-ALS and sALS cohorts. Distribution of TF-binding loci relative to the transcriptional start site (TSS) in differentially accessible TF motifs in **d**, oligodendrocytes, **e**, oligodendrocyte precursor cells (OPC), and **f**, microglia. Transcription factor footprint analysis plots of enriched Tn5 insertion around **g**, FOXG1 in excitatory neurons, **h**, TCFL5 in inhibitory neurons, **i,** BACH2 in astrocytes, and **j,** JUNB in microglia.

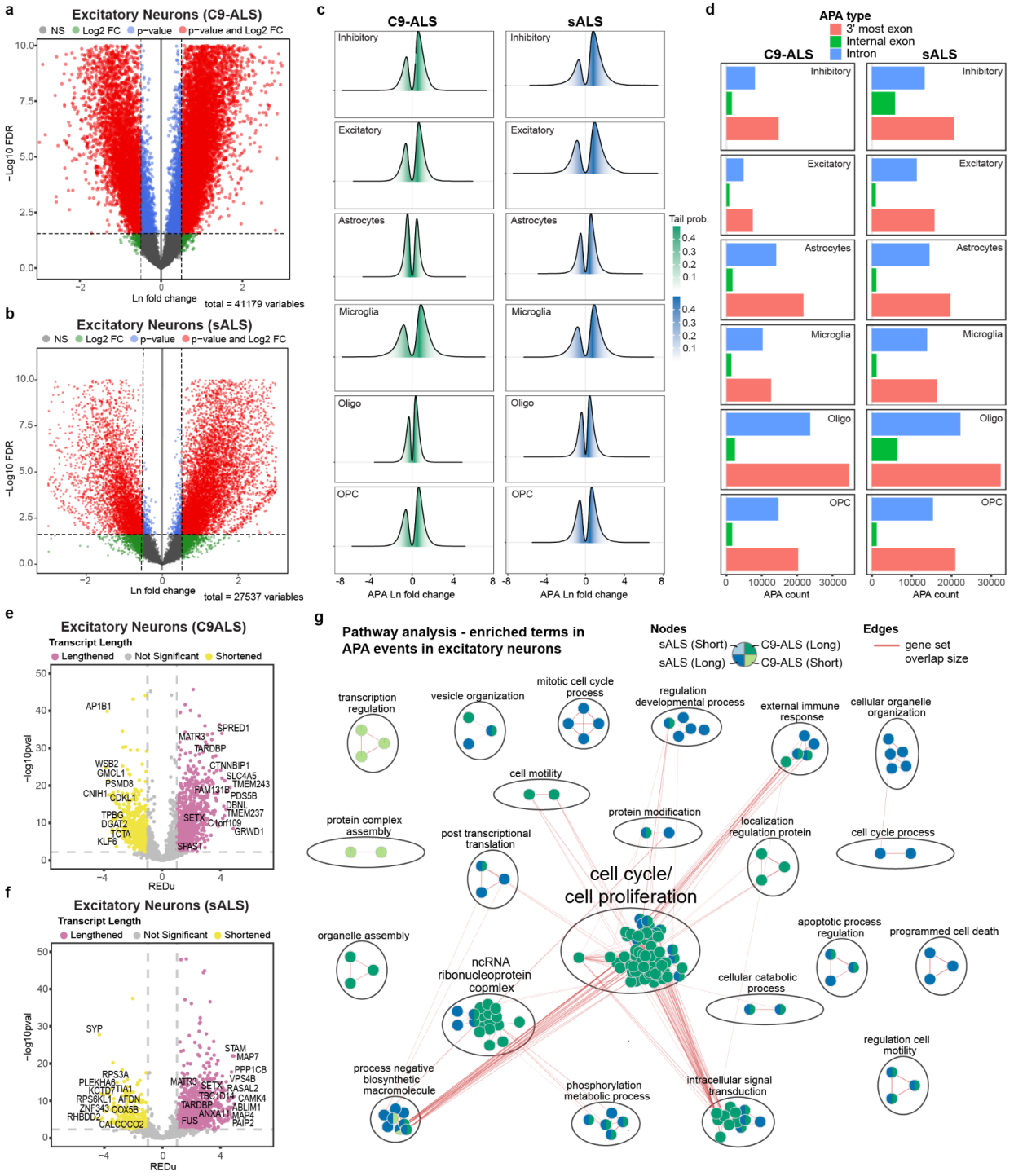
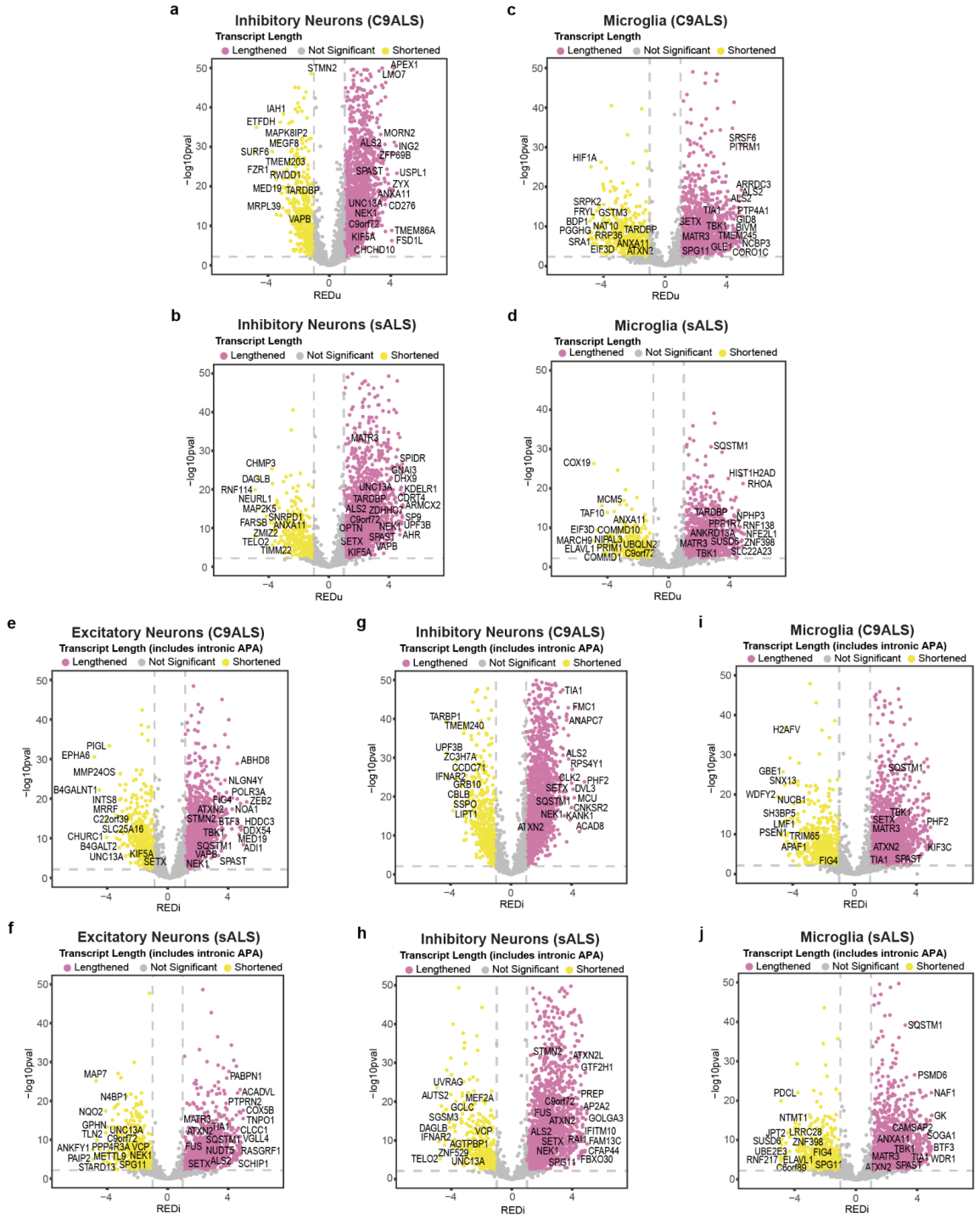


Figure 4. Dysregulation of alternative polyadenylation (APA) landscape in ALS compared to control. a, Volcano plot of APA events charting all feasible APA pairs between proximal and distal polyadenylation sites. The x-axis denotes the natural logarithm (Ln) fold change (LFC) of distal to proximal PAs usage. **a**, Represents these findings for excitatory neurons from C9-ALS versus control, while **b**, highlights the findings for excitatory neurons from sALS. **c,** Ridge plot elucidating the distribution of APA events in both C9-ALS and sALS relative to control samples, across the major cell types. P The "Tail probability" indicates the likelihood of an APA event in the given distribution. **d**, Annotation or the locational distribution of quantified APAs. **e**,

Volcano plot illustrating lengthened and shortened genes in excitatory neurons in C9-ALS relative to controls, as measured by the relative expression difference in the 3'UTR (REDu) metric (Methods). **f**, Volcano plot, analogous to **d**, but showing findings for excitatory neurons from sALS in comparison to control**. g**, Pathway analysis of the lengthened and shortened genes in excitatory neurons across ALS subtypes.



Extended Data Figure 5. (Related to Fig. 4). Dysregulation of Alternative Polyadenylation (APA) in ALS Across Cell Types. This figure presents volcano plots that depict lengthened and shortened genes in ALS subtypes across Excitatory neurons, Inhibitory neurons and Microglia, determined by the REDu and REDi metrics. REDu evaluates the relative expression difference between the top two isoforms in the 3′-most exon, indicating transcript lengthening (positive REDu) or shortening (negative REDu). REDi assesses the relative expression between the top differentially expressed isoform in the 3′-most exon versus an intron or internal exon. (Methods) The plots are organized as follows: **a**, REDu in Inhibitory Neurons in C9-ALS; **b**, REDu in Inhibitory Neurons in sALS; **c,** REDu in Microglia in C9-ALS; **d**, REDu in Microglia in sALS; **e**, REDi in Excitatory Neurons in C9-ALS; **f**, REDi in Excitatory Neurons in sALS; **g**, REDi in Inhibitory Neurons in C9-ALS; **h**, REDi in Inhibitory Neurons in sALS; **i**, REDu in Microglia in C9-ALS; **j**, REDu in Microglia in sALS.

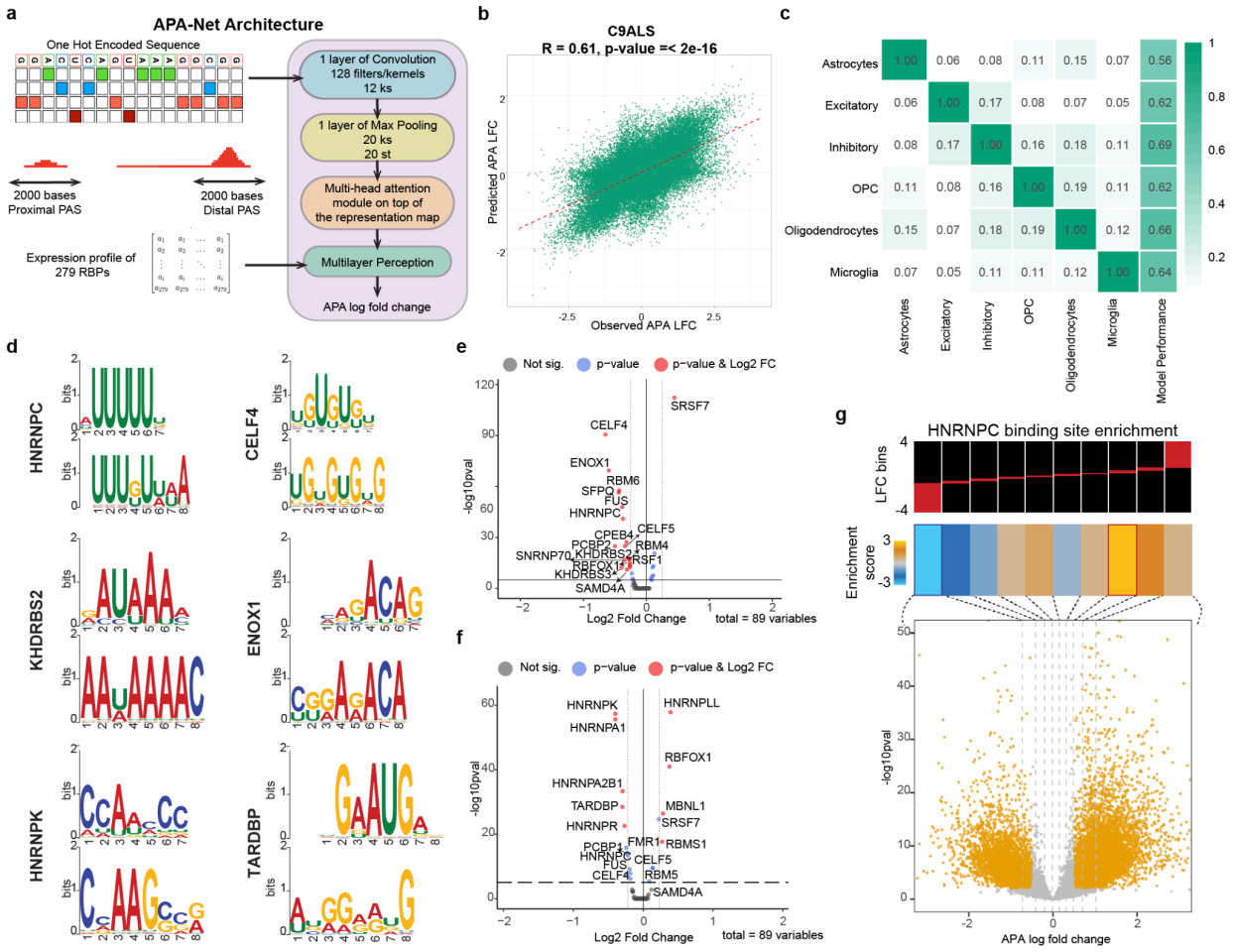
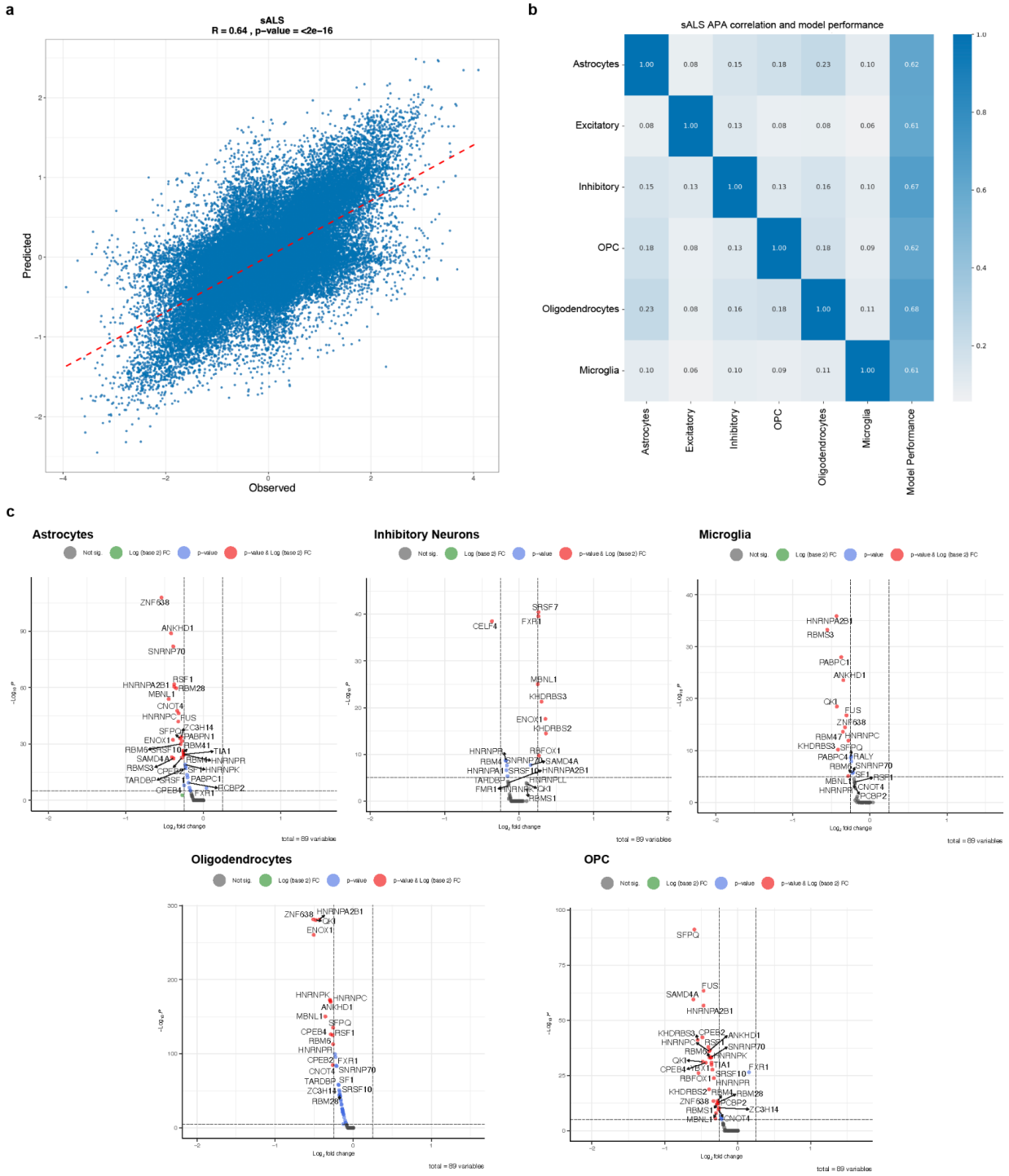
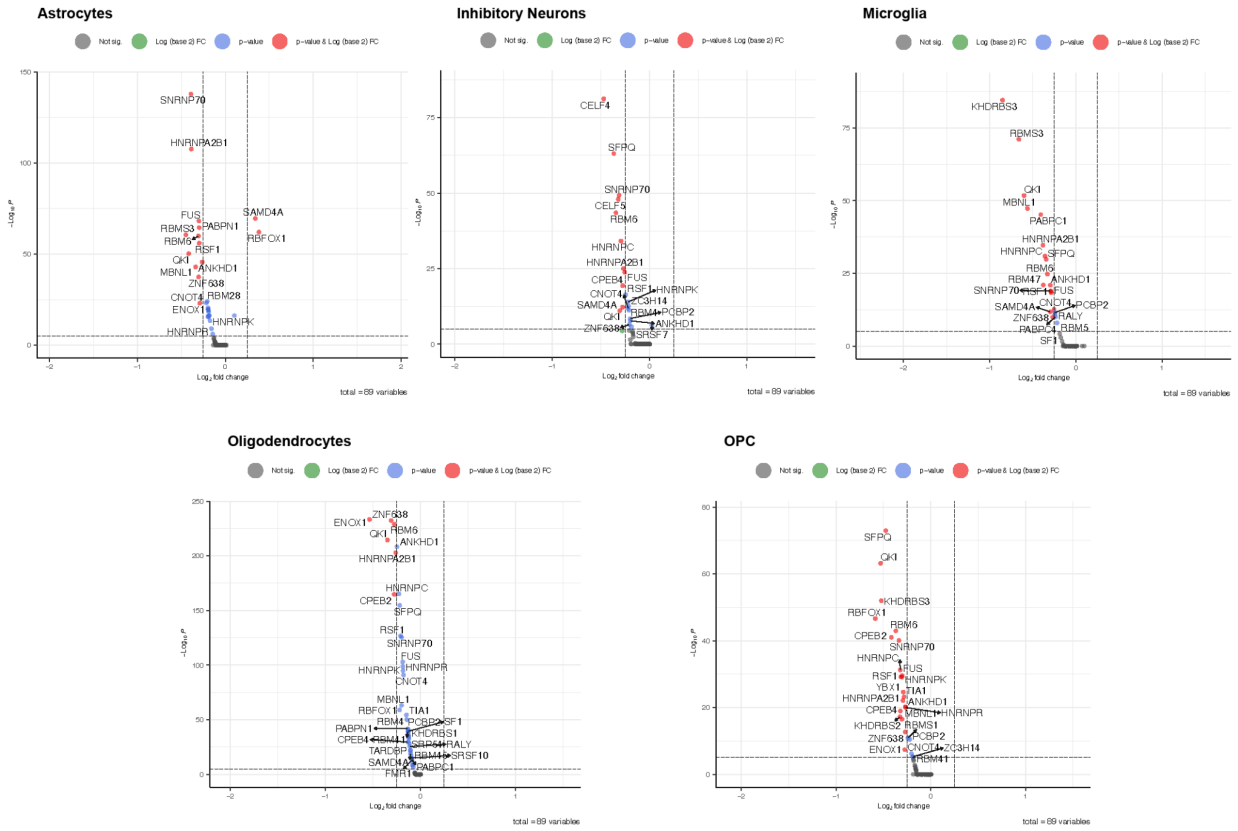


Figure 5. Development and analysis of a deep learning model to unravel the grammar of alternative polyadenylation (APA) regulation in C9-ALS and sALS cases. a, Schematic representation of the APA-Net architecture and inputs. The model uses two inputs: sequences surrounding the PA sites and the expression profiles of 279 RBPs. **b**, Scatter plot depicting the correlation between predicted APA log fold change (LFC) and observed APA LFC on a test dataset previously not exposed to the model for C9-ALS cases. **c**, Heatmap of APA correlations among different cell types. A low correlation suggests varied APA usages across cell types. The final column highlights the APA-Net's performance for each specific cell type, focusing on C9-ALS cases. **d**, APA-Net filters interpreted as motifs (Methods and Supplementary Data Figure 5), which were subsequently aligned to an RNA-binding protein (RBP) database to identify corresponding RBPs. **e,** Differential expression analysis performed on the identified RBPs. The volcano plot shows upregulated and downregulated RBPs in

excitatory neurons from C9-ALS and **f**, sALS. **g**, Enrichment analysis targeting HNRNPC binding sites for APA events. Top: APA sequences are divided into equally populated bins based on their LFC values. Red shows the proportion of APA events in each bin. Middle: Enrichment score indicating under-representation and over-representation. Bins with significant enrichment (hypergeometric test, corrected $P < 0.05$; red) or depletion (blue) of poly(U) motifs are denoted with a bolded border. Bottom: volcano plot showing the distribution of changes in APA LFC in excitatory neurons from C9-ALS compared to controls. Significant observations are highlighted in orange.



Extended Data Figure 6. (Related to Fig. 5) Deep learning model performance in sALS cell types versus control. a, Scatter plot depicting the correlation between Predicted APA log fold-change (LFC) (y-axis) and Observed APA LFC (x-axis) on a test dataset not previously exposed to the model for sALS cases. **b**, Heatmap illustrating APA correlations among different cell types. A low correlation suggests varied APAs across cell types. The final column highlights the model's performance for each major cell type, focusing on sALS cases. **c,** Results of differential expression analysis on the identified RBPs across the major cell types in sALS versus non-neurological controls. The volcano plot showcases upregulated and downregulated RBPs in each cell type.



Extended Data Figure 7. (Related to Fig. 5). Differential expression of RBPs in C9-ALS cell types versus control. Results of differential expression analysis on the identified RBPs across the major cell types in C9-ALS versus non-neurological controls. The volcano plot demonstrates the upregulated and downregulated RBPs in each cell type.

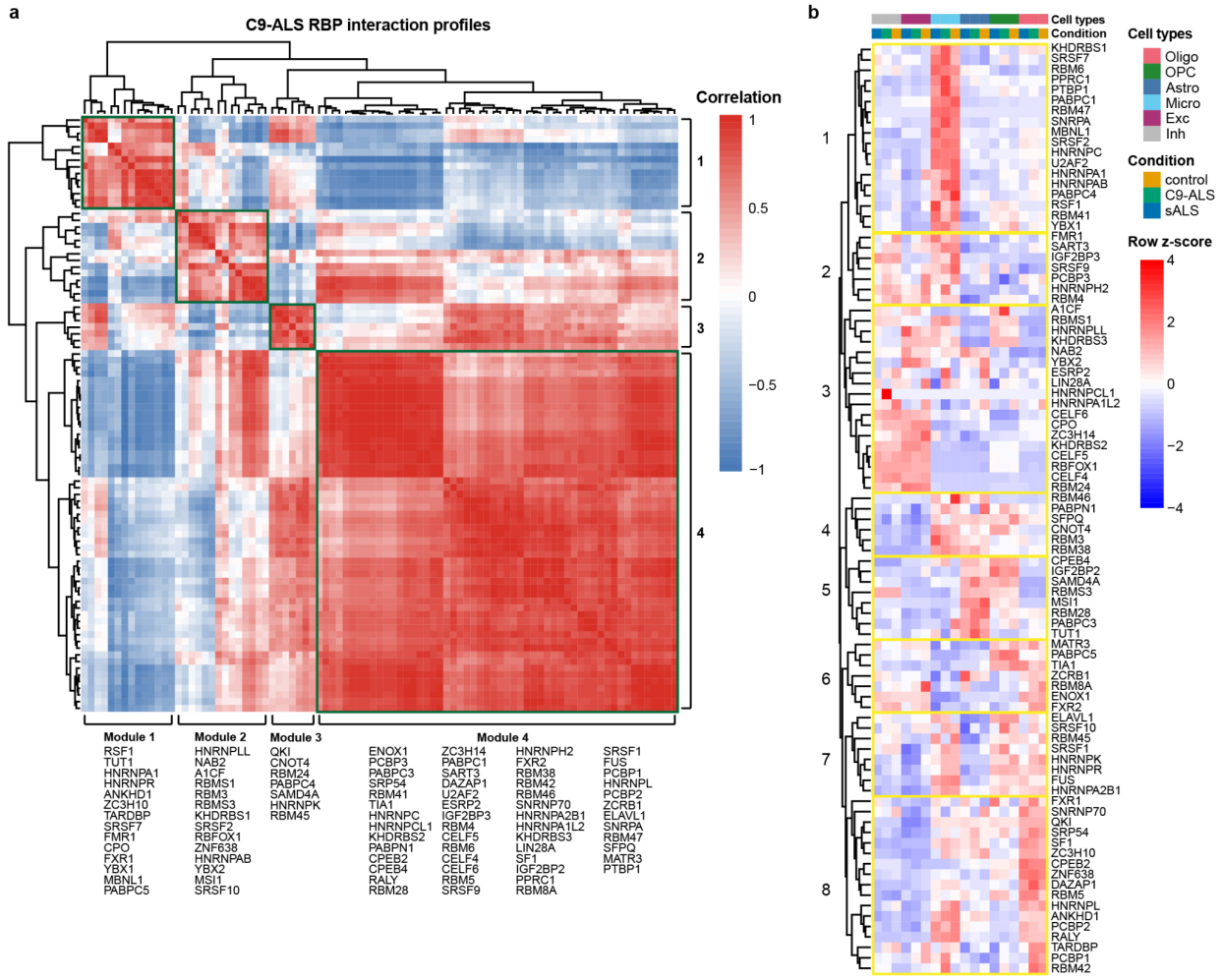
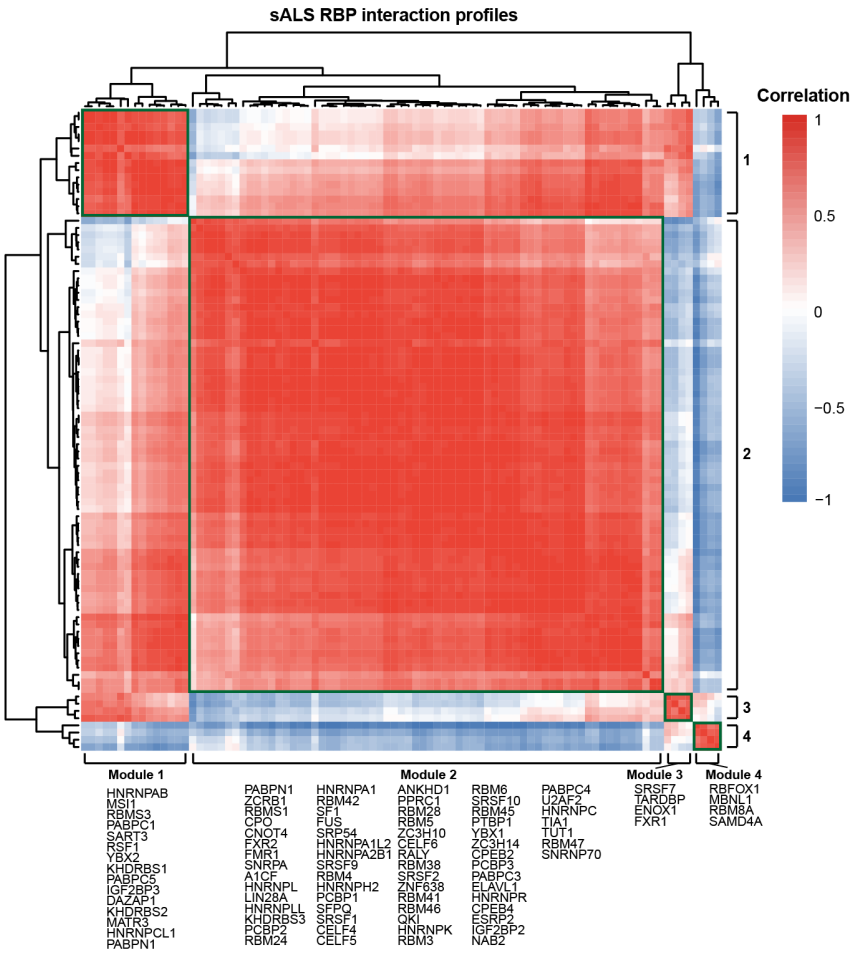
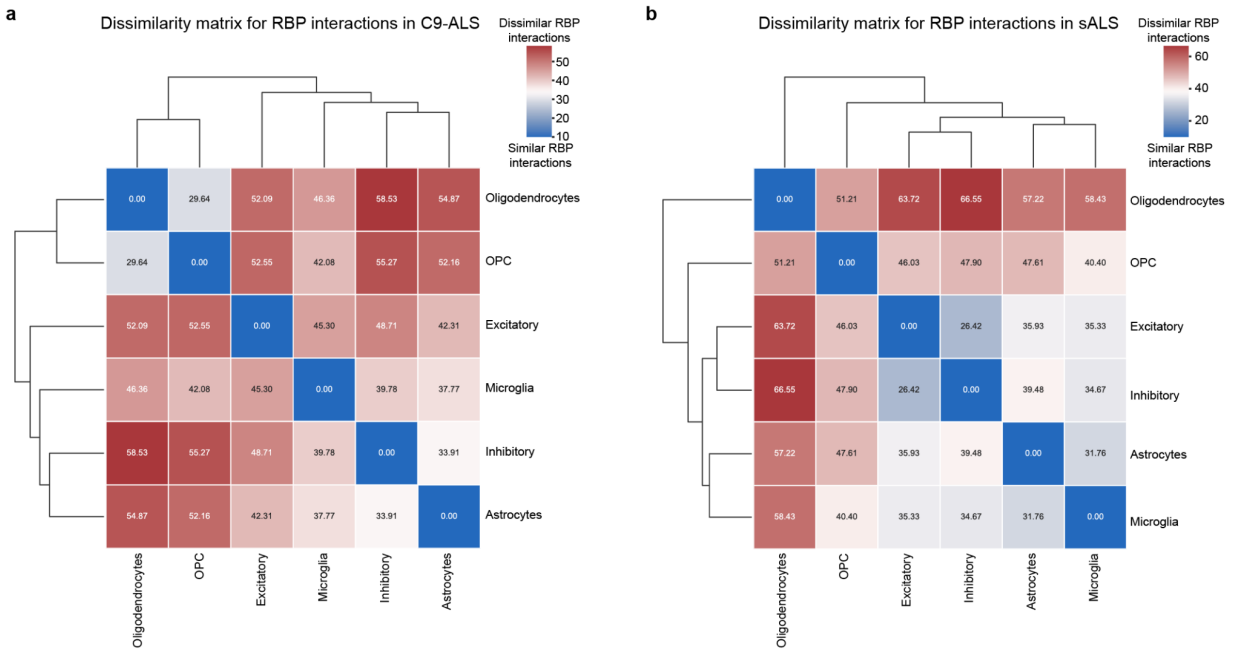


Figure 6. RNA-binding protein interactions reveal cell type-specific mechanisms and dysregulation of alternative polyadenylation in C9-ALS. a, Clustered RBP correlation profile heatmap computed using Pearson correlation in C9-ALS. Four RBP modules are defined based on this clustering (labeled on y- and x-axes). **b**, Clustered heatmap of RBP expression patterns across major cell types and disease subtypes. The dendrogram partitions the RBPs into cell type enrichment groups.

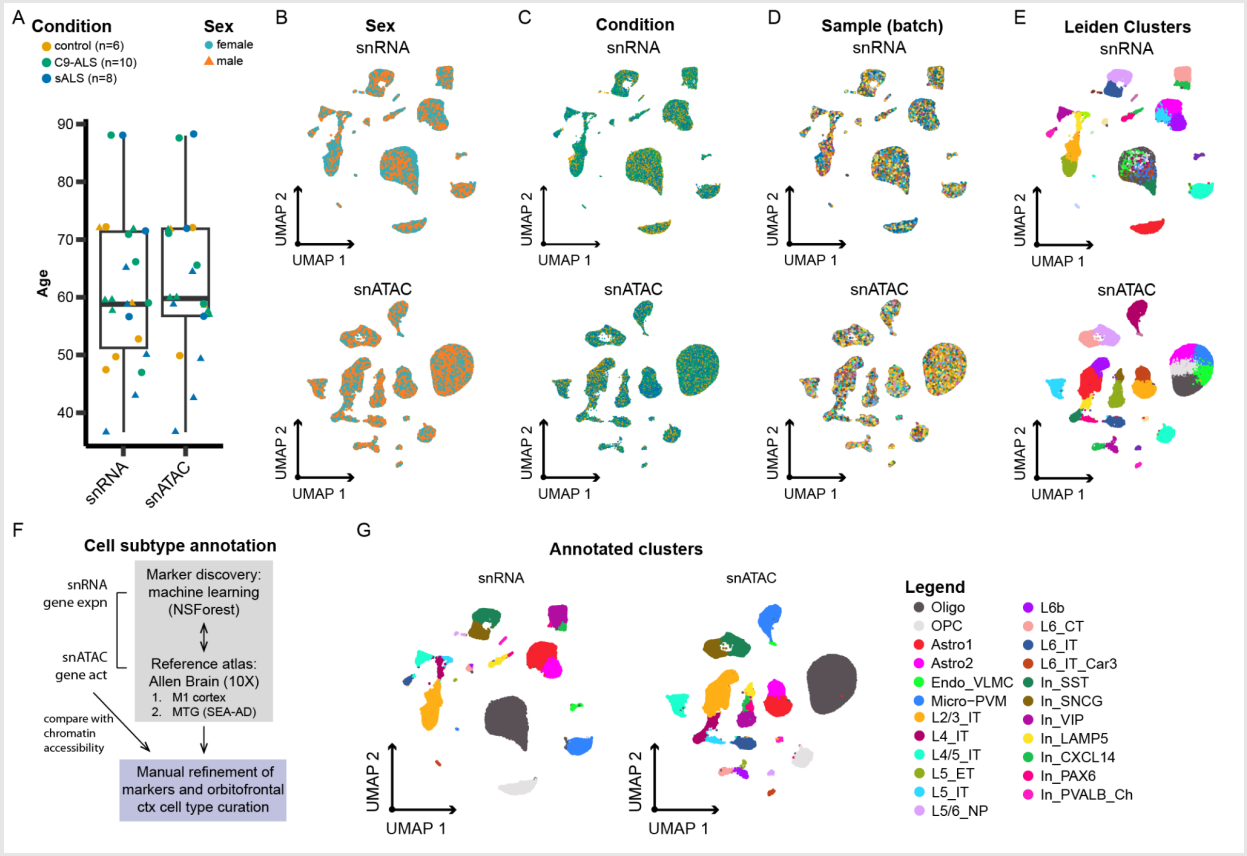


Extended Data Figure 8 (related to Fig. 6). RNA-binding protein interactions reveal cell type-specific mechanisms and dysregulation of alternative polyadenylation in sALS. Clustered RBP correlation profile heatmap computed using Pearson correlation in sALS. Four RBP modules are defined based on this clustering (labeled on y- and x-axes).

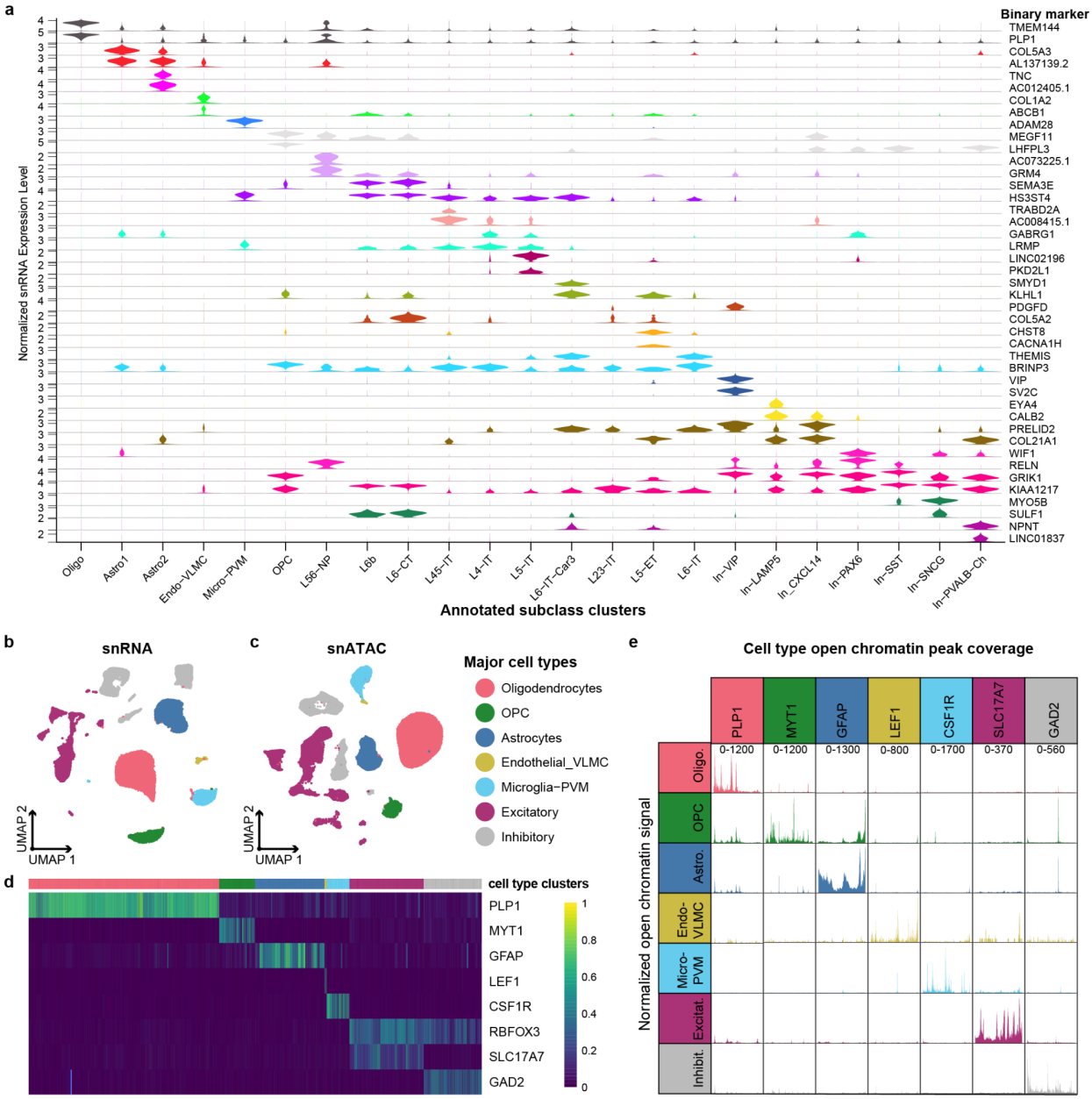


Extended Data Figure 9 (related to Fig. 6). RBP interaction dissimilarity across cell types in C9-ALS and sALS. Clustered heatmaps showing the dissimilarity of RBP correlation profile interactions measured by pairwise Frobenius norm of correlation profile matrices, as depicted e.g. in Figure 6a, among cell types in **a**, C9-ALS and **b**, sALS. (See Methods)

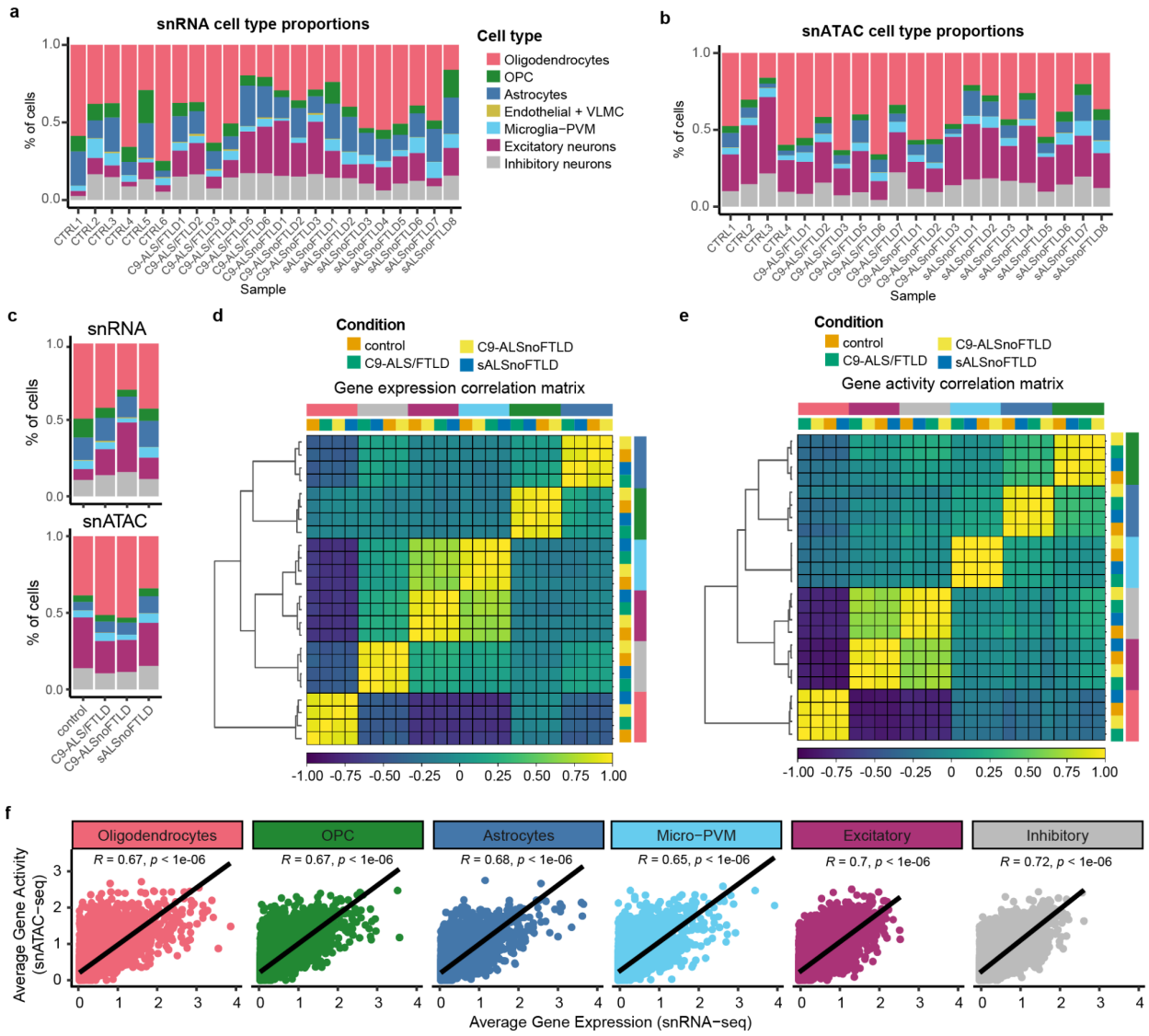
Supplementary Data Figures



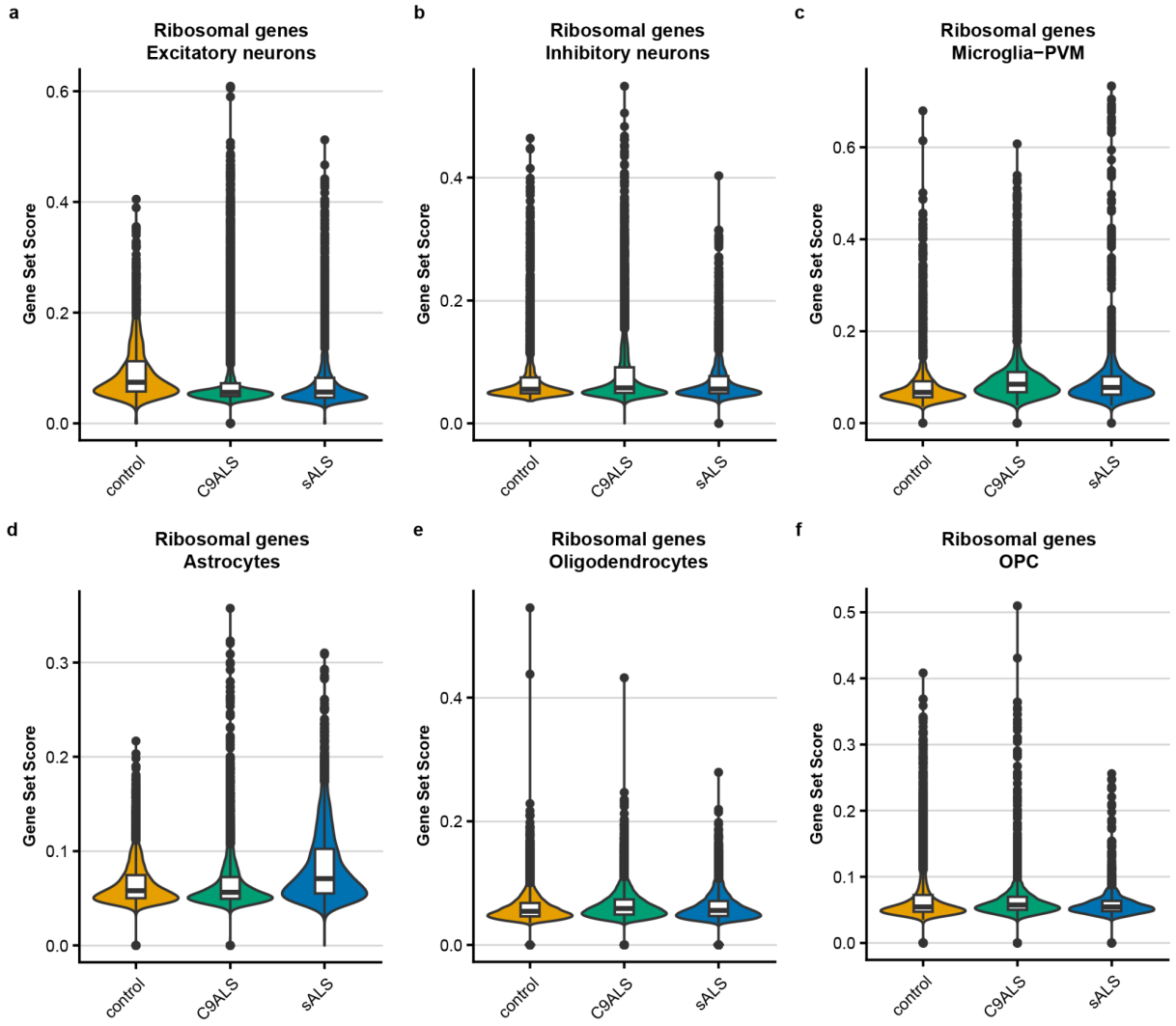
Supplementary Data Figure 1. Demographic information and cell type annotation in snRNA-seq and snATAC-seq. a, Boxplots of age and sex across all analyzed snRNA-seq (61.6 ± 13.3 years) and snATAC-seq (63.1 ± 13.0 years) samples. UMAP plots for **b**, sex; **c,** disease subtype; **d**, sample batches; and **e**, Leiden Clusters. **f**, flowchart of the cell subtype annotation approach. **g**, UMAP of annotated clusters from snRNA-seq (left) and snATAC-seq (right).



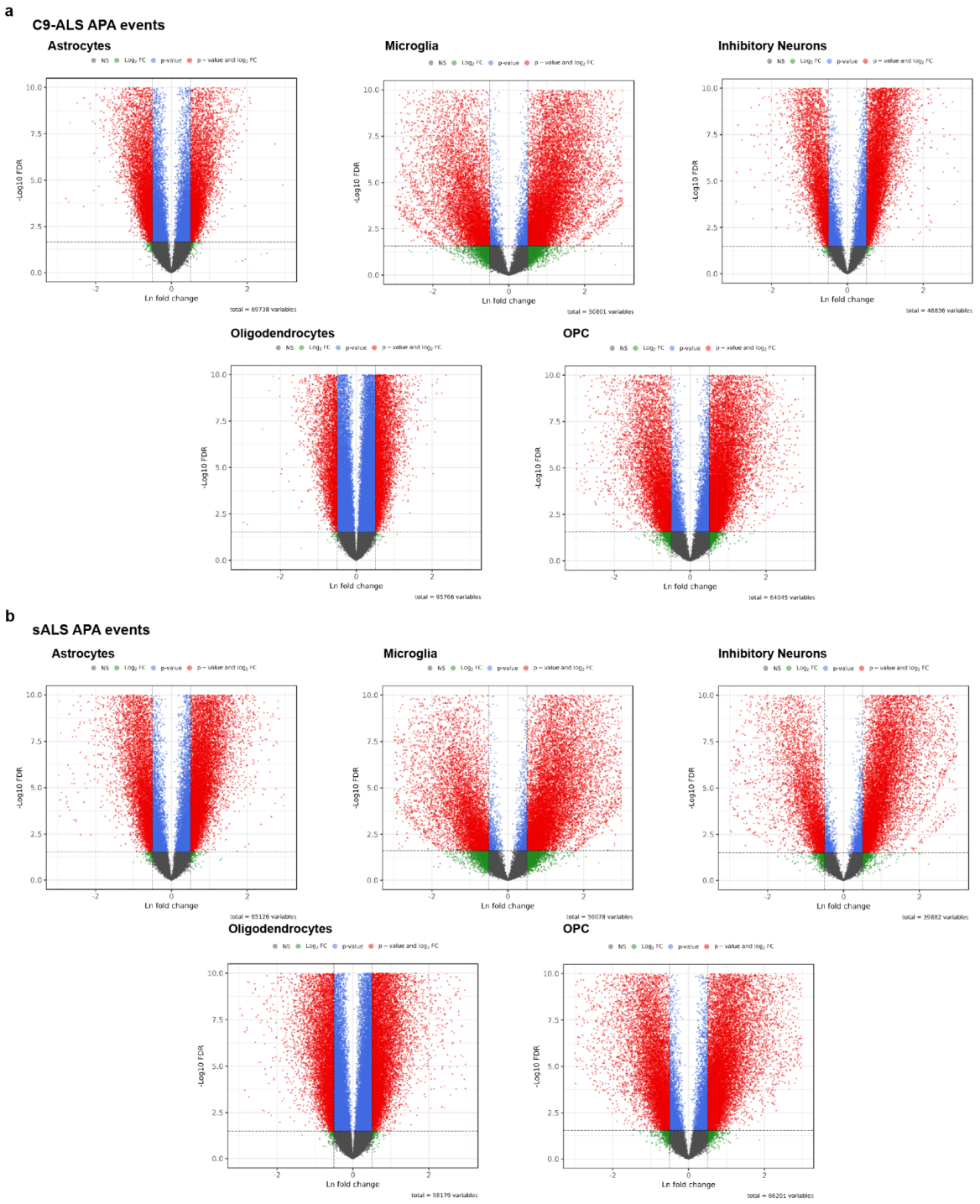
Supplementary Figure 2 (related to Figure 1). Markers of frontal cell subtypes in snRNA-seq and snATAC-seq. a, Stacked violin plot of binary markers identified by NS-Forest. **b**, UMAP of major cell type annotation in snRNA-seq. **c**, UMAP of major cell type annotation in snATAC-seq; **d**, snRNA-seq heatmap of canonical cell type markers. **e**, Chromatin accessibility peak coverage within the gene body of canonical cortical cell type markers in snATAC-seq. Oligo = oligodendrocytes; Astro = astrocytes; Endo-VLMC = endothelial and vascular lepotomeningeal cell; Mico-PVM = microglia and perivascular macrophages; OPC = oligodendrocytes precursor cells; excitatory neurons subtypes = Layer 2/3 intratelencephalic (L2/3 IT), L4 IT, L4/5 IT, L5 extra telencephalic-projecting (L5 ET), L5 IT, L5/6 near projecting (L5/6 NP) L6B, L6 corticothalamic-projecting (L6 CT), L6 IT, L6 IT Car3; enhibitory neuron subtypes: IN-SST (*SST*+), In-SNCG (*SNCG*+), In-VIP (*VIP*+), In-LAMP5 (*LAMP5*+), In-CXCL14 (*CXCL14*+), In-PAX6 (*PAX6*+), and In-PVALB-Ch (*PVALB*+ chandelier cells).



Supplementary Data Figure 3. (Related to Figure 1). Proportions and correlations between cell types across technologies and disease subtypes. Barplots of cell type proportions by sample in **a**, snRNA-seq and **b,** snATAC-seq. **c**, Cell type proportions by disease subtype in snRNA-seq (top) and snATAC-seq (bottom). **d**, Gene expression correlation matrix for snRNA-seq across major cell types and disease subtypes. **e**, Clustered correlation matrix heatmaps for gene activity, as estimated by per gene promoter peaks in snATAC-seq, for snRNA-seq across major cell types and disease subtypes. **f**, Scatter plots comparing the average gene expression level of a given gene for all frontal cortex cell types from snRNA-seq (x-axis) with the average gene activity score in snATAC-seq (y-axis). Regression line indicates correlation strength; Pearson correlation coefficient R considered significant at p<0.000001.



Supplementary Data Figure 4. (Related to Figure 2). Dysregulation of ribosomal subunits across cell types in C9-ALS and sALS. Violin and boxplots of ribosomal genes (RPL/S pattern) for **a,** excitatory neurons, **b,** inhibitory neurons, **c**, Microglia-PVM (perivascular macrophages), **d,** astrocytes, **e**, Oligodendrocytes, **f,** OPC. The y-axis represents gene set signature scores for ribosomal genes (Methods).



Supplementary Data Figure 5 (Related to Fig. 4). Volcano plots of APA events charting all feasible APA pairs between proximal and distal Polyadenylation sites. The x-axis denotes the natural logarithm (Ln) fold change of distal to proximal PAs usage. **a**, C9-ALS major cell types versus control. **b**, sALS major cell types versus control.

Supplementary Data Figure 6. (related to Fig. 5). Position Weight Matrix (PWM) representation of 128 filter activations for C9-ALS. Visualization of 128 sequence filters for C9-ALS. PWMs are generated by scanning the test sequences and aligning the subsequences with high activation levels.

Supplementary Data Figure 7. (related to Fig. 5). Position Weight Matrix (PWM) representation of 128 filter activations for sALS. Visualization of 128 sequence filters for sALS. PWMs are generated by scanning the test sequences and aligning the subsequences with high activation levels.