HHS Public Access

Author manuscript

Cell Rep. Author manuscript; available in PMC 2025 April 21.

Published in final edited form as:

Cell Rep. 2025 March 25; 44(3): 115402. doi:10.1016/j.celrep.2025.115402.

ALS molecular subtypes are a combination of cellular and pathological features learned by deep multiomics classifiers

Kathryn O'Neill[1], **Regina Shaw**[2,3], **Isobel Bolger**[2,3], **NYGC ALS Consortium**[6], **Oliver H. Tam**[2,3,*], **Hemali Phatnani**[4,5,*], **Molly Gale Hammell**[2,3,7,*]

[1]Cold Spring Harbor Laboratory School of Biological Sciences, Cold Spring Harbor, NY 11724, USA

[2]Institute for Systems Genetics, NYU Langone Health, New York, NY 10016, USA

[3]Department of Neuroscience & Neuroscience Institute, NYU Langone Health, New York, NY 10016, USA

[4]New York Genome Center, New York, NY 10013, USA

[5]Department of Neurology, Columbia University, New York, NY 10032, USA

[6]See Table S8 for a list of consortium members

[7]Lead contact

SUMMARY

Amyotrophic lateral sclerosis (ALS) is a complex syndrome with multiple genetic causes and wide variation in disease presentation. Despite this heterogeneity, large-scale genomics studies revealed that ALS postmortem samples can be grouped into a small number of subtypes, defined by transcriptomic signatures of mitochondrial dysfunction and oxidative stress (ALS-Ox), microglial activation and neuroinflammation (ALS-Glia), or TDP-43 pathology and associated transposable elements (ALS-TE). In this study, we present a deep ALS neural net classifier (DANCer) for ALS molecular subtypes. Applying DANCer to an expanded cohort from the NYGC ALS Consortium highlights two subtypes that strongly correlate with disease duration: ALS-TE in cortex and ALS-Glia in spinal cord. Finally, single-nucleus transcriptomes demonstrate that ALS subtypes are recapitulated in neurons and glia, with both ALS-wide and subtype-specific alterations in all cell

*Correspondence: oliver.tam@nyulangone.org (O.H.T.), hphatnani@nygenome.org (H.P.), molly.galehammell@nyulangone.org (M.G.H.).

AUTHOR CONTRIBUTIONS

K.O., O.H.T., H.P., and M.G.H. conceived the study. K.O. designed the neural net classifiers and analyzed data. O.H.T. designed CellRanger-TE and analyzed the majority of the bulk and single-nucleus data. R.S. performed single-nucleus isolation from postmortem cortex. I.B. contributed to figure design. H.P. and the NYGC ALS Consortium contributed bulk RNA-seq data for cortex and spinal cord samples. K.O., O.H.T., H.P., and M.G.H. contributed to the writing of the manuscript. All authors approved the manuscript.

DECLARATION OF INTERESTS

M.G.H. has the following additional affiliations: adjunct associate professor, Stony Brook University Graduate Program in Genetics; adjunct associate professor, Cold Spring Harbor Laboratory; and affiliate, New York Genome Center. M.G.H. serves on the scientific advisory board of a company called Transposon Therapeutics. M.G.H. holds the following patents: US9441223B2, "Transposable elements, TDP-43, and neurodegenerative disorders."

SUPPLEMENTAL INFORMATION

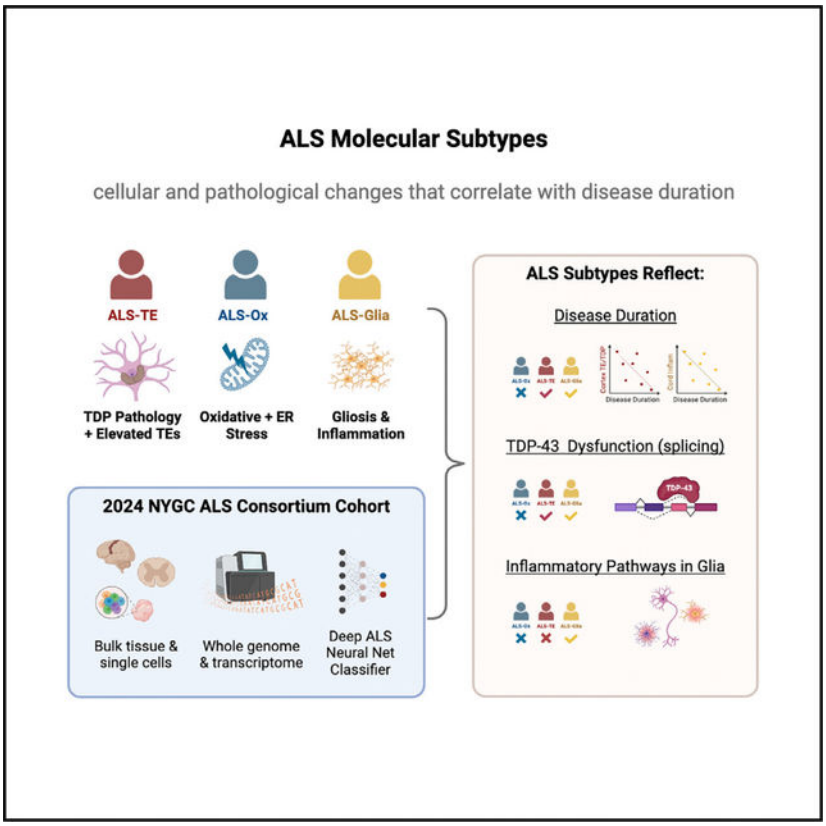
Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2025.115402.

types. In summary, ALS molecular subtypes represent a combination of cellular and pathological features that correlate with clinical features of ALS.

In brief

O’Neill et al. present a deep exploration of molecular subtypes in ALS postmortem tissues from the NYGC ALS Consortium. Matched bulk and single-cell profiles demonstrate that ALS subtypes are a combination of cellular alterations and pathological features that include neuronal stress, accumulation of TDP-43 dysfunction, and activation of inflammatory pathways.

Graphical abstract



INTRODUCTION

Amyotrophic lateral sclerosis (ALS) is a fatal neurodegenerative disease characterized by the progressive loss of motor neurons from the motor cortex and spinal cord.[1] Loss of motor neurons, which control voluntary motor function, typically leads to total paralysis and eventual death within 2–5 years of diagnosis.[2] Across patients living with ALS (pALSs), there is a high degree of variability in terms of the site of disease onset, age at diagnosis, rate of disease progression, and relative involvement of upper/lower motor neurons. Despite this heterogeneity, nearly all affected pALS tissues show pathological accumulations of phosphorylated TDP-43 protein.[3,4] TDP-43 pathology is present in most cases with known ALS causal mutations, typically called “familial ALS.”[5] At least 40 and as many as

90 genetic mutations have been associated with ALS.[6] These ALS-associated genes can generally be grouped by their inferred roles[5] in RNA processing, protein homeostasis, neuroinflammation, and cytoskeletal dynamics. TDP-43 pathology is additionally present in nearly all pALS cases without a family history of disease or known causal mutation, often called "sporadic ALS," with approximately 90% of all pALS cases estimated to have the sporadic form.[7]

Several other neurodegenerative diseases present with pathological inclusions of TDP-43,[8–10] most prominently in the TDP-43 subset of patients with frontotemporal dementia (FTD-TDP).[10] ALS and FTD-TDP share many common genetic factors and a high degree of co-morbidity, such that these two diseases are understood to exist on a spectrum.[11] Patients with inherited mutations in C9orf72 may present with pure ALS, pure FTD, or a combination of both.[12] In addition to shared pathology, other molecular alterations are commonly seen across the ALS/FTD spectrum. Previous work from our laboratories used genomics profiling of the NYGC ALS Consortium cohort to reveal that tissues from pALSs can be grouped into a small number of ALS molecular subtypes.[1] ALS molecular subtypes are defined by postmortem expression signatures of: mitochondrial dysfunction and oxidative stress (ALS-Ox), microglial activation and neuroinflammation (ALS-Glia), or dense TDP-43 pathology and associated transposable elements (ALS-TE). Work from groups on independent ALS cohorts has replicated these findings in motor cortex and detected subtypes in regions outside the central nervous system.[13] Work from other groups on the same cortical samples from the NYGC ALS Consortium cohort also replicated the original findings for ALS molecular subtypes.[14] Taken together, these studies have confirmed the presence of ALS molecular subtypes in frontal and motor cortex,[1,13,14] shown that these subtypes are detectable in peripheral biofluids,[13] and demonstrated an association between ALS subtypes and clinical correlates.[13,14]

Several open questions remain about the nature and implications of ALS subtypes. First, subtype studies have all used bulk RNA sequencing (RNA-seq) transcriptomics, abrogating our ability to determine cell-type-specific contributions. Second, these studies profiled a few hundred pALS samples in each cohort and may have been underpowered to detect rare subtypes. Finally, these studies have not yet integrated spinal cord tissues, despite the importance of spinal motor neuron degeneration in ALS. In this study, we have addressed these gaps in our understanding. Single-nucleus transcriptome profiles demonstrate that ALS subtypes cannot be explained purely by changes in cell composition. An updated set of NYGC ALS Consortium samples that triple the size of the original study shows no evidence of new subtypes. Finally, the availability of matched cortical and spinal cord tissues outlines the extent to which subtypes are present across motor tissues. To support these findings, we present a deep ALS neural net classifier (DANCer) that can operate in single-cell mode (scDANCer).

RESULTS

Expansion of the NYGC ALS Consortium cohort

The NYGC ALS Consortium study has greatly expanded since 2019, adding hundreds of new genomics profiles across cortical and spinal cord tissues. Since the original ALS

molecular-subtype study, the NYGC ALS Consortium has more than tripled the number of available samples to provide whole genomes and transcriptomes for 719 cortical samples from 398 pALSs and control individuals, with the majority of these samples derived from newly sequenced donors (Figure 1A). The cohort shows balance in terms of patient sex (57% male; Table S1), with a median age at diagnosis of 63 (Table S1) and a median disease duration of 34 months post- diagnosis (Table S1). Most of the samples in the NYGC ALS Consortium cohort were obtained from patients (n = 286) with a diagnosis of ALS; 8% were also diagnosed with the related disorder FTD (Table S1). This largely sporadic cohort showed few patients carrying known mutations of large effect (79% sporadic ALS; Figure 1B). Among the ALS-spectrum cases (ALS with or without FTD), 14% of the NYGC Consortium cohort carried repeat expansion mutations in the C9orf72 gene (Figure 1B), 2% carried mutations in SOD1, and 1% of ALS cases carried mutations in the FUS, ATXN2, or OPTN genes. All other mutations were present in a single case and involved mutations in: NEFH, VCP, PNPLA6, FIG4, UBQLN2, and PFN1. The distributions of pALS donors across site of onset, age of onset, and disease duration are similar to those seen in previous releases from the NYGC ALS Consortium[15–17] and to larger estimates of disease presentation in other cohorts, such as CREATE,[18] Project Mine,[19] and Answer ALS.[20]

DANCer accurately assigns ALS molecular subtypes

Identification of ALS molecular subtypes was enabled by a *de novo* clustering and pattern discovery method called non-negative matrix factorization (NMF).[21] NMF is a powerful technique to define new molecular pathways associated with samples in bulk transcriptomic data but is impractical to apply in the context of continuously expanding data collections. To address this limitation, we developed a classifier using an artificial neural network architecture to accurately predict ALS molecular subtypes. Here, we present our DANCer.

To reduce noise and redundancy within the transcriptome datasets, the transcriptome input to DANCer is first compressed using weighted gene correlation network analysis (WGCNA).[22] One WGCNA module/eigengene was excluded as it consisted entirely of sex-biased genes. The remaining 39 eigengenes were used as inputs for a multilayer feedforward perceptron (Figure 1C), which was trained to reproduce the labeled ALS molecular subtypes outlined in our primary study,[1] reaching 96.4% classification accuracy on the original labeled datasets (Figure 1D). Per-sample DANCer accuracy is represented by solid dots in an autoencoder latent space (Figure 1D), where the small number of incorrectly assigned samples are rendered translucent. Additional details about the training protocols, parameters used in the models, and generation of an autoencoder latent space can be found in the STAR Methods section and Figure S1. The code for DANCer is available on GitHub at https://github.com/mhammell-laboratory/DANCer.

DANCer classifications show that ALS molecular subtypes are robust across an expanded ALS cohort

We applied DANCer in prediction mode to the updated 2023 release from the NYGC ALS Consortium. The input data for DANCer consisted of whole-transcriptome RNA-seq datasets from 773 frontal and motor cortex samples of 431 subjects, as described above, and are visualized in a principal-component analysis (PCA) representation color coded by ALS

subtype (Figure 1E). Sample-level assignments were converted to a patient-level assignment for all donors, using a majority call for any discordant samples, as previously described.[1] The updated set of NYGC ALS Consortium patient-level subtypes were classified as 20% ALS-TE, 70% ALS-Ox, and 10% ALS-Glia (Figure 1F). This distribution is broadly similar to that seen previously, with small changes likely reflecting differences in patient populations and inclusion of pALS cases with co-morbid FTD. When looking just at patients with a pure FTD diagnosis, out of all those on the ALS/FTD spectrum, the distribution shifts more toward the ALS-TE and ALS-Glia subtypes with 83% of the FTD patients assigned as ALS-TE, 1% as ALS-Ox, and 13% as ALS-Glia (Table S1C).

We next explored whether any biological or technical co-variates influence subtype assignments, finding no significant difference between ALS-Glia or ALS-TE subtypes and controls for technical variables such as sample RNA integrity (RIN) values or tissue pH. The ALS-Ox subtype samples showed slightly higher RIN values when compared to controls (Wilcoxon $p < 6.5e{-}7$; Table S1D). We also explored biological co-variates, such as sex, site of onset, ALS/FTD diagnosis, age at diagnosis, and known ALS mutations. We found no significant association with mutations of large effect (Table S1E), although we were underpowered to assess associations with mutational status. The ALS-Ox subtype was significantly enriched for patients with a pure ALS diagnosis (odds ratio 19.2, Fisher's $p < 3.2e{-}6$) and depleted for patients with a pure FTD diagnosis (odds ratio 0.3, Fisher's $p < 3.0e{-}9$). In contrast, the ALS-TE subtype was strongly enriched among FTD patients (odds ratio 43.4, Fisher's $p < 3.1e{-}17$).

Finally, to determine whether new subtypes are present in this cohort, we ran the *de novo* pattern-finding algorithm that was originally used to define ALS subtypes, NMF.[21] NMF returned an optimal clustering of k = 3, with the three clusters showing a similar set of class assignments as previously described (see Figure S2). To highlight the similarity between the original NMF results and the DANCer subtype assignments, we provide PCA plots in Figure S2D of all samples in the NYGC ALS Consortium cohort, the original NMF-labeled samples, and the samples newly sequenced as part of this study. Taken together, these results demonstrate that ALS subtypes are robust in terms of defining molecular characteristics across large patient cohorts originating from many different geographical areas and clinical centers.

ALS spinal cord samples recapitulate the ALS-Ox and ALS-Glia subtypes

The NYGC ALS Consortium sample set also contains transcriptomes for 664 ALS and control spinal cord samples from 406 donors: 331 ALS and 75 non-ALS (Figure 2A). To determine whether the ALS subtypes would also be present in spinal cord samples, we ran NMF on the spinal cord transcriptomes. NMF returned an optimal clustering of k = 2, with just two dominant transcriptional patterns (Figures 2A and S3). The genes that distinguished these two clusters in spinal cord showed a clear pattern (Figure 2B): oxidative stress genes were elevated in the first group (NMF1) and inflammatory pathways in the second group (NMF2). Gene set enrichment analysis (GSEA)[23] returned several elevated pathways shared among all ALS spinal cord samples, including ribosomal proteins, genes associated with SOD1 mutant microglia,[24] and lysosomal components (Figure 2B).

Neuroactive ligands and receptors, calcium signaling, and classical ALS disease genes marked the NMF1 spinal group (Figure 2B, ALS-Ox specific, blue). In contrast, the second group (NMF2) showed stronger evidence for elevated inflammatory signatures (Figure 2B, ALS-Glia specific, gold). Furthermore, specific markers of pro-inflammatory astrocytes were significantly enriched in the NMF2/ALS-Glia spinal cord samples (Figure 2C), including those describing astrocytes induced via lipopolysaccharide (LPS) or oxidative stress.[25] To better reflect the genes expressed by these spinal cord subtypes, we re-named the first (NMF1/blue) group ALS-Ox-cord and the second (NMF2/gold) group ALS-Glia-cord (Figures 2A and 2B).

We next sought to understand whether the spinal cord subtypes mirror their cortical counterparts. We first compared the average log2 fold change for all genes in ALS-Glia spinal cord samples to the average log2 fold change for all genes in ALS-Glia cortex samples, relative to controls (Figure 2D), finding a significant correlation between the DE genes in pALS cortex and spinal cord ($r = 0.36$, $p < 1e{-}15$) despite the large differences in cellular composition between these two tissues. Similarly, we see a significant correlation for altered gene expression in ALS-Ox subtypes from the cortex and spinal cord (Figure 2F, $r = 0.43$, $p < 1e{-}15$). This suggests that the original ALS-Glia and ALS-Ox cortical subtype patterns are recapitulated in the spinal cord and conserved throughout the motor system. A full list of differentially expressed (DE) genes in each of the two spinal cord subtypes is given in Tables S2G and S2I.

Figure 2B uses a PCA (Figure 2E) to visualize the relative similarity of all ALS spinal cord and cortex signatures. PC1 cleanly separates inflammatory (ALS-Glia cord and cortex) signatures, while PC2 separates the remaining neuronal stress subtypes (ALS-Ox cord and cortex) from the ALS-TE subtype that largely represents TDP-43 dysfunction. Upset plots showing the lists of shared DE genes between ALS subtypes across cortex and spinal cord are shown in Figure S4 and listed in Table S3A. Figure S4 also displays similarity (upset plots and PCA) between each ALS subtype and several recent studies looking at ALS postmortem spinal cord samples[16] or induced motor neuron (iMN) ALS models,[26–28] with a list of shared DE genes given in Table S3. These studies have each nominated distinct dysregulated pathways in lower motor neurons: P53, ELAVL3, and Toll-like receptors. Each of these shares some overlap with the ALS-Ox and ALS-Glia subtypes present in spinal cord.

Integrative analysis across spinal cord and cortex reveals correlates of ALS molecular subtypes with TDP-43 dysfunction

We next asked whether each patient would show a single concordant pattern in both cord and cortical tissues (Figure 3A; Table S1J). There were 267 pALS donors with both motor tissues profiled. A majority of the patients with the ALS-Glia/inflammatory signature in the cortex also showed ALS-Glia/inflammatory signatures in the spinal cord (23 out of 27 = 85%). However, nearly all pALS cases showed evidence of inflammatory pathway signatures in spinal cord (189 out of 267 = 71%). This point emphasizes the stark differences in the relative frequency of inflammatory signatures in the cortex and cord. ALS-Glia

subtype patients formed the smallest group of all subtypes in the NYGC ALS Consortium cohort in cortex (10%) yet formed the majority of spinal cord cases (71%).

Our previous analysis of ALS cortex showed enrichment for TDP-43 target genes among the set of markers for the ALS-TE subtype and an enrichment for dense TDP-43 pathology in immunohistochemistry (IHC) stains of tissues from that group.[1] However, no equivalent of the ALS-TE cortical subtype was found in the spinal cord tissues. To better identify signatures of TDP-43 dysfunction, we took advantage of a recent study profiling ALS/FTD patients with and without functional TDP-43 protein in the nucleus where it normally resides.[29] Isolating all neuronal nuclei from the cortex, they counter-sorted for TDP-43-positive and -negative neurons and performed bulk RNA-seq of the two populations. Later studies incorporating these data have identified hallmark TDP-43 splicing targets such as STMN2,[31,32] UNC13A,[33,34] and others.[35] To compare these differential splicing changes to those seen in the NYGC ALS Consortium data, we used Leafcutter[36] to reveal TDP-43-mediated splicing defects across subtypes (Table S2). Figure 3D shows differential splicing changes plotted as delta percent spliced in (delta PSI) for each of the ALS subtypes vs. control in the cortex (y axis) as compared to the delta PSI values for those same genes in the TDP-43-depleted neurons from Liu et al.[29] ALS-TE subtype samples showed the strongest splicing correlation with TDP-43-negative neurons (Figure 3D, Pearson's r = 0.49, $p \approx 0$), including alterations in known TDP-43 target, UNC13B (Figure 3F). ALS-Ox subtype samples from cortex showed anti-correlated splicing changes with TDP-43-depleted neurons (indicating a closer splicing profile to control/TDP-43-positive neurons). These data suggest that both ALS-Glia and ALS-TE patient cortex tissues show TDP-43 dysfunction-mediated splicing alterations, with ALS-TE samples showing the greatest evidence for TDP-43 dysfunction. These results mirror the results from our previous ALS subtype study,[1] where ALS-TE subtype tissues had the highest levels of TDP-43 pathology.

To extend this analysis to the spinal cord samples, we performed a similar analysis using a set of previously published laser micro-dissected spinal motor neurons from ALS and control spinal cord samples.[30] Using this spinal motor neuron reference set, we saw strong correlations between splicing alterations seen in the NYGC ALS Consortium cohort ALS-Glia spinal cord samples (Figure 3E) and the splicing alterations seen in the Krach et al.[30] motor neurons (Pearson's r = 0.71, $p \approx 0$), including known TDP-43 target genes UNC13B[34,37] and STMN2[31,32] (Figure 3G).

Integrative analysis across spinal cord and cortex reveals correlates of ALS molecular subtypes with disease duration

We next asked whether any ALS subtypes were associated with disease duration. Disease duration is defined as the number of months between the date of ALS diagnosis and the date of death or tracheostomy, which varied between 6 and 264 months in our cohort. For all patients with cortical samples, we plotted the eigengene scores for each ALS subtype against the disease duration in months (Figures 4B, S5A, and S5B). Only the ALS-TE samples showed a significant correlation between ALS subtype eigenvalues and disease duration in cortex ($r = -0.25$, $p < 0.05$). This is somewhat equivalent to saying that increased TDP-43 dysfunction in cortex is associated with shortened survival times, since the ALS-TE

subtype showed the strongest evidence for TDP-43-associated dysfunction (Figure 3D). No association was seen between expression of the ALS-Ox or ALS-Glia eigengenes and disease duration in cortex (see Figures S5A and S5B).

We repeated this exploration of clinical correlates for the spinal cord. ALS-Glia signatures in spinal cord are significantly correlated with shortened disease duration. This was especially true for patients assigned to the ALS-Glia group (Figure 3C, gold, r = −0.35, $p < 2e{-}10$) but was also true for patients assigned to the ALS-Ox group (Figure S5C, blue, r = −0.29, $p <$ 6e–06). This suggests that some patients in the ALS-Ox group in the cord show moderate expression of the inflammatory signature genes (e.g., have some positive eigengene values), and, to the extent that they do, this correlates with shortened survival times.

scDANCER shows that cell-type composition changes among ALS subtypes in cortex

To explore the possibility that tissue composition changes contribute to ALS subtypes, we generated single-nucleus RNA-seq (snRNA-seq) samples from several patient motor cortex samples from our cohort.[1] We integrated these with a larger set of ALS motor cortex snRNA-seq samples obtained from an additional cohort.[38] This integrated dataset represents 48 ALS snRNA-seq transcriptomes (from 26 cases) as well as 28 control motor cortex samples (from 14 non-ALS donors) (Figure 4A). An integrated uniform manifold approximation and projection (UMAP) of all 509,966 cells (Figure 4B) shows good representation of expected cortical cell types, which cluster by cell type and not by ALS/control, sex, or data source (Figure S6). A heatmap of cell-type-specific marker genes (Figure 4C) shows faithful expression of cell-lineage markers in each corresponding cell type, with a layer 5 extrathalamic (L5ET) cortical neuron signature marked by FEZF2 and LRATD2.[39] L5ET excitatory neurons include cortical motor neurons (Betz cells).

For samples sequenced in this study, ALS subtypes are known from bulk RNA-seq profiles. For the additional set of published ALS snRNA-seq samples,[38] no corresponding data were available to assign subtypes. To overcome this, we retrained our classifier to take in pseudobulk data from the snRNA-seq composite transcriptomes and accurately assign subtypes (Figure 4D). Specifically, we took the labeled bulk RNA-seq data that had previously been used to train DANCer and downsampled the data to only select genes detectable in snRNA-seq. scDANCer reached 91.8% accuracy on test/train data after 200 training epochs, using the same 20-fold cross-validation strategy to estimate classifier accuracy as was used for DANCer (Figure S7; Table S4A). scDANCer accurately called the correct ALS subtype for all samples where matched bulk and snRNA-seq data were available (Table S4B). For the samples without matched bulk data,[38] we note that 95% of the replicate samples from the same donors gave the same subtype (Table S4C). All pseudobulk transcriptomes are shown on a PCA plot (Figure 4E) that includes bulk NYGC ALS Consortium samples (Figure 4E), matched snRNA-seq data from the ALS Consortium (large diamonds), and the additional snRNA-seq samples[38] (large circles).

We next explored cell composition changes across ALS subtypes. As shown in Figure 4F, we see slight differences in the number of cells sampled across ALS subtypes. ALS-Ox samples show a cell-type composition that was largely reflective of that seen in control motor cortex with mild but significant enrichments for excitatory ($p < 0.01$) and inhibitory (p

< 0.001) neurons and slightly fewer microglia ($p < 0.0002$) captured. Similarly, the ALS-TE samples show a slight enrichment for excitatory ($p < 0.01$) and inhibitory ($p < 0.01$) neurons and relative depletion of oligodendrocyte precursor cells ($p < 0.006$) and microglia ($p < 0.002$). ALS-Glia samples show a more pronounced depletion of both excitatory ($p < 0.006$) and inhibitory ($p < 0.01$) neurons and a relative increase in microglia ($p < 0.01$). Looking more specifically at just the L5ET neuron populations (Figure 4G), only the ALS-Glia/inflammatory subtype samples show a significant loss of L5ET neurons, with 70% fewer L5ET/motor neurons recovered ($p < 6e–06$).

ALS L5ET neuron profiles show both pan-ALS and ALS subtype-specific alterations

We next set out to determine whether differences would emerge between L5ET neurons from tissues with different ALS subtypes. Compared to control L5ET neurons, we found hundreds of DE genes in each ALS subtype (Figure 5A; Table S5). We estimated relative levels of TDP-43 dysfunction in these L5ET neurons using GSEA based upon the set of genes known to be downregulated in sorted TDP-43-negative cortical neurons,[29] as described above. ALS-TE subtype L5ET neurons showed the greatest evidence for TDP-43 dysfunction out of all L5ET neurons (normalized enrichment score [NES] = −2.1, false discovery rate [FDR] < 0.003). L5ET neurons from the ALS-Ox and ALS-Glia subtypes showed no significant alterations in the genes that mark TDP-43-negative neurons.[29]

While the results in Figures 5A and 5B highlight differences between the ALS subtypes in terms of gene expression, there were also shared alterations or pan-ALS L5ET expression changes. Looking across all L5ET neurons in the ALS samples, we found 17 significantly upregulated and 30 significantly downregulated genes across all subtypes (Figures 5B; Table S5H). These shared alterations (Figures 5C and 5D) largely reflect neuronal stress genes that have previously been associated with ALS (NEFL, NEFM), pro-apoptotic genes (PNMA2, BIN1), and genes generally involved in proteostasis (DUSP1, SCRN2). UMAP displays of gene expression in ALS and control L5ET neurons for a subset of these pan-ALS altered genes are shown in Figure 5E below the L5ET marker gene FEZF2.

In addition to these shared ALS-wide expression alterations, L5ET neurons also displayed subtype-specific expression patterns reflective of gene pathways previously identified in the bulk tissue profiles (Table S5). In the ALS-TE subtype, we found broad upregulation of transposable element (TE) transcripts from young and active TEs (Figure 5F) as well as dysregulation of known targets of TDP-43 (Figure 5F): the long non-coding RNAs NEAT1[40,41] and MALAT1[40,41] and the TDP-43 kinase CSNK1E[30] (Figure 5F). In addition to these selected genes that overlap with those previously identified in bulk profiles, we show the top 10 most subtype-selective DE up- and downregulated genes in ALS-TE L5ET neurons (Figure 5I). In the ALS-Ox subtype (Figure 5G), we found upregulation of genes associated with autophagy and lysosomes (VCP, OPTN) and chaperones (HSBP1L1, DNAJB4, DNAJC8). In addition, we show the top 10 most subtype-selective DE up- and downregulated genes in ALS-Ox L5ET neurons (Figure 5J). Motor/L5ET neurons from ALS-Glia/inflammatory subtype (Figure 5H) show upregulation of genes associated with inflammatory pathways (IFI6 and IFIT1) and synaptic function (SYT2, SV2A, SV2B).

Figure 5K shows the top 10 most subtype-selective DE up- and downregulated genes in ALS-Glia L5ET neurons.

ALS subtype-specific differences in pro-inflammatory pathway activation in astrocytes

Given that astrocytes play central roles in mediating neuroinflammatory processes in the central nervous system,[42] we wondered whether these cells would also show ALS subtype-specific differences. A UMAP of all motor cortex astrocytes is given in Figure 6A, where Leiden clustering[43] identified 14 astrocyte clusters, nine of which showed no evidence of doublets or cell-type assignment errors (Table S6). Looking at the genes that marked each astrocyte cluster (Table S6A) and comparing to previously published markers of astrocyte states,[25,44,45] we noticed that our nine clusters could be grouped into four broad expression programs, which are labeled on the UMAP representation (Figure 6A): those expressing homeostatic astrocyte markers (WIF1; Figure 6B, top left), an inflammatory cluster marked by high CD44 expression (Figure 6B, top right), genes that mark disease-associated astrocytes (DAAs)[44] and show high expression of TMSB4X (Figure 6B, bottom left) and those that express high levels of TEs and P2RY14 (Figure 6B, bottom right). Splitting the UMAP representations by ALS subtype revealed differences in the relative frequency of astrocyte clusters in each ALS group (Figure 6D). ALS-Glia subtype samples contained more astrocytes of the interferon-driven inflammatory cluster (Figure 6D), which was significantly different than the distribution of astrocytes in control samples (Figure 6E). Similarly, the ALS-Ox subtype samples were enriched for astrocytes expressing DAA markers (Figures 6D and 6E). Finally, the ALS-TE subtype astrocytes show both elevated TE expression and fewer homeostatic-cluster astrocytes than controls (Figures 6D and 6E).

This cluster-based analysis of DAA and inflammatory astrocyte markers across subtypes was also significant when added to a standard GSEA[23] using a modified MSigDB database[23] (see STAR Methods) that included published marker sets representing the transcriptional response of astrocytes to pro-inflammatory signals: reactive astrocytes induced by lipopolysaccharide (RA-LPS[25]), oxidative stress (RA-MCAO[25]), and DAAs[44]. ALS-Glia subtype astrocytes showed significant upregulation of genes in pathways associated with interferon signaling (Figures 6C; Tables S6B and S6C) as well as significant enrichment for RA-LPS and RA-MCAO markers. ALS-Ox subtype astrocytes showed elevation of genes associated with the DAA signature (Tables S6D and S6E), which was not enriched in other subtypes.

Shared and distinct DE genes in each subtype are shown in violin plots and volcano plots in Figures 6F–6H and listed in Tables S6C–S6I. Briefly, ALS-Glia astrocytes show significant elevation of pro-inflammatory marker genes (Figure 6F top: AQP1 and CRYAB). ALS-Ox astrocytes show elevation of known ALS-associated genes[46] (Figure 6G, top: CHCHD10 and SOD1). ALS-TE astrocytes show significant elevation of young TEs such as L1HS and AluSx4 (Figure 6H, violins) as well as the P2Y receptor P2RY14, which has previously been associated with response to pro-inflammatory signaling in astrocytes.[47] All ALS astrocytes showed downregulation of the heat shock proteins HSPH1 and HSPD1, regardless of subtype (Figures 6F–6H, bottom panels).

ALS subtype-specific differences in pro-inflammatory pathway activation in microglia

Microglia are tightly integrated with astrocytes and neurons in neuroinflammatory cascades, adopting a multitude of states in response to neuronal stress and inflammatory stimuli.[48] As such, we were curious whether ALS subtypes would reflect differences in microglial cell states. We first extracted and re-clustered all captured microglia, with a UMAP of their expression in Figure 7A. Leiden clustering returned 11 clusters, seven of which were present without evidence of doublets or cell-type assignment errors (Table S7A). Looking at the genes that marked each of these microglia clusters (Table S7) and comparing to previously published studies of microglial state markers,[24,49] we noticed that the microglia clusters could be grouped into four broad expression programs, which are labeled on the UMAP in Figure 7A: genes associated with elevated metabolic activity and marked by GAPDH (Figure 7B, top left), those expressing homeostatic markers of microglia function and marked by P2RY12 (Figure 7B, top right), those expressing interferon signaling pathways and marked by high APOE expression (Figure 7B, bottom right), and genes that have been previously associated with phagocytic activity and marked by high expression of SLC11A1 (Figure 7B, bottom left). Splitting the UMAP representations by ALS subtype shows striking differences in the representation of microglial clusters among ALS subtypes. ALS-Glia subtypes were dominated by microglia from the inflammatory cluster (Figures 7D and 7E). Microglia from ALS-Ox subtype samples were predominantly expressing signatures of elevated metabolic activity (Figures 7D and 7E). ALS-TE subtype samples showed strong enrichment for microglia from the homeostatic cluster (Figure 7D).

GSEA revealed significant upregulation of innate immune-response genes in ALS-Glia subtype microglia, including interferon pathways (Figures 7D; Table S7C). To the standard GSEA lists, we added published markers representing the transcriptional response of microglia to pro-inflammatory stimuli, including microglia induced by lipid droplets (Microglia-LDAMs),[50] pro-inflammatory microglia induced by SOD1 mutations (SOD1-DAMs),[24] and disease-associated microglia induced by constitutive TREM2 activity (TREM2-DAMs).[49] Of the experimentally defined pro-inflammatory microglia pathways, only the SOD1-DAM set[24] was enriched.

Additional sets of shared and distinct DE genes in each subtype are shown in violin plots and volcano plots in Figures 7F–7H and listed in Table S7. Briefly, ALS-Glia subtype microglia (Figure 7F) show significant elevation of pro-inflammatory marker genes (ACSL1, CTSB) and innate immune pathway markers (IL7R, STAT4). ALS-Ox microglia (Figure 7G) show elevation of mitochondrial and synaptic function (SNAP25, NDUFA4). ALS-TE microglia (Figure 7H) show significant elevation of young TEs (L1HS, SVA-D). Heat shock proteins were among the most downregulated genes across all ALS subtypes, as noted by HSPA1B (Figures 7F–7H). A full table of shared DE genes is shown in Table S7H.

DISCUSSION

In this study, we present an analysis of the largest ALS postmortem molecular profiling study available, with whole transcriptomes available for hundreds of pALSs from multiple affected motor system tissues. The size of this study allowed us to conclude that ALS molecular subtypes are highly robust across multiple clinical centers in large ALS cohorts.

We used *de novo* pattern-finding methods to determine that no new molecular subtypes were present in this larger patient cohort, which quadrupled the size of the patient samples available in our initial study.[1]

Upon integrating the cortex and spinal cord data available in the updated NYGC ALS Consortium cohort, we noted that ALS subtype patterns in the spinal cord largely mirror two of those previously seen in the motor cortex. Patients with an ALS-Glia/inflammatory subtype in the motor cortex are more likely to show inflammatory patterns in their spinal cord samples, as well, suggesting a small number of patients with systemic neuroinflammation. The TDP-43-driven ALS-TE subtype does not appear in the bulk spinal cord tissue samples, which may reflect differences in cellular composition of the two tissues. That said, single-nucleus transcriptomes revealed that ALS subtypes, originally discovered in bulk transcriptome data, reflect cell-state alterations from multiple cell types and are not simply reflections of neuronal cell loss.

Strong correlations with disease duration were seen for ALS-TE subtype expression in the cortex and ALS-Glia subtype expression in spinal cord. Taken together, these results suggest that disease duration in ALS reflects, in part, the degree of TDP-43 dysfunction in the cortex and inflammatory processes in the spinal cord. In summary, the ALS-Ox subtype represents neuronal stress without TDP-43 dysfunction or inflammatory pathway activation; the ALS-TE subtype reflects accumulation of TDP-43 dysfunction without interferon-related inflammatory pathways; and, finally, the ALS-Glia subtype represents the presence of both TDP-43 dysfunction and pro-inflammatory expression.

Limitations of the study

While this study represents the largest postmortem genomics profiling study available for ALS motor cortex and spinal cord samples, several questions remain unanswered. Postmortem studies are limited to describing processes at play at the end stage of disease and cannot be used to determine the relative timing of molecular alterations. Postmortem studies use cells present at the end stages of disease, which may be biased toward those that are most resilient to disease-associated processes. Finally, we were underpowered to determine the extent to which genetic mutations drive predisposition to ALS molecular subtypes.

STAR★METHODS

RESOURCE AVAILABILITY

Lead contact—Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Molly Gale Hammell (molly.galehammell@nyulangone.org).

Materials availability—This study did not generate any new reagents; all reagents are commercially available.

Data and code availability

- RNA-seq and snRNA-seq data have been deposited at SRA and are publicly available as of the date of publication. Accession numbers are listed in the key resources table. Published snRNA-seq data[38] are available from BioProject PRJNA1073234.
- Code generated for this study (DANCer and CellRanger-TE) has been deposited in GitHub and is publicly available as of the date of publication. The URL and DOI are listed in the key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANTS DETAILS

ALS postmortem samples—The NYGC ALS Consortium samples in this study were acquired through various IRB protocols from member sites and transferred to NYGC in accordance with all applicable foreign, domestic, federal, state, and local laws and regulations for processing, sequencing, and analyses. Generation of snRNA-seq libraries from postmortem material at CSHL was approved by the CSHL internal review board (IRB-17–015). Generation of genomics libraries from postmortem material in the NYGC ALS Consortium is covered by IRB# BRANY-15–08-292–385. All material has been anonymized, was obtained postmortem, and thus is considered "not human subjects". All available de-identified clinical and pathological records were collected and used together with C9orf72 genotypes to summarize patient demographics and disease features (see Table S1). Care was taken to ensure sex/gender distribution of patient and control individuals from whom samples were obtained were close to equal.

METHOD DETAILS

Generation of bulk RNA-seq data—RNA was extracted from flash-frozen patient samples homogenized in Trizol (15596026, ThermoFisher Scientific, Waltham, MA, USA) -Chloroform and purified using the QIAGEN RNeasy Mini kit (74104, QIAGEN, Germantown, MD, USA). RNA was assessed using the Bioanalyzer (G2939BA, Agilent, Santa Clara, CA, USA). RNA-seq libraries were prepared from 500 ng of total RNA using the KAPA Stranded RNA-seq kit with RiboErase (07962304001, Kapa Biosystems, Wilmington, MA, USA) for rRNA depletion and Illumina compatible indexes (NEXTflex RNA-seq Barcodes, NOVA-512915, PerkinElmer, Waltham, MA, USA). Pooled libraries (average insert size: 375bp) were sequenced on an Illumina HiSeq 2500 or NextSeq V1 using a paired end 125 nucleotide setting, to yield 40–50 million reads per library.

Extraction of nuclei from fresh frozen tissue—Single nuclei were extracted from archived human postmortem motor cortex tissue using a method adapted with modifications from Maitra et al.[57] A Dounce homogenizer was used to homogenize fresh frozen cortex tissue on ice. Five mL of chilled Nucleus wash buffer[57] was added to the homogenate, to quench the lysis. A 30-μm MACs SmartStrainer (Miltenyi Biotech cat. # 130–098-458) was used to remove cell debris. The lysed, filtered homogenate was then promptly centrifuged in a desktop centrifuge equipped with a swinging bucket rotor at 500 G for 5 min at 4°C.

After centrifugation, the sample was removed and supernatant decanted without disrupting the nuclei pellet, and placed on ice. The pelleted nuclei were gently resuspended on ice. A fresh 15-mL centrifuge tube was equipped with a 30 μM MACs SmartStrainer and the filtration and resuspension and wash processes were repeated as above. One mL of 50% (weight/volume) working solution of iodixanol (Optiprep) was added to the resuspended nuclei and mixed gently, to obtain 2 mL of nuclei in 25% (wt/vol) iodixanol solution. For gradient centrifugation, 0.5 mL of the nucleus suspension was gently pipetted on top of an iodixanol cushion (0.5 mL of 29% (wt/vol) iodixanol) by letting it slowly run down the wall of the tube, to prevent mixing. Without disturbing the layers, the LoBind tubes containing the nuclei on top of the iodixanol cushion were centrifuged at 10,000 G for 30 min at 4°C. A second gradient centrifugation step was carried out at 10,000 G for 10 min at 4°C. Supernatant was then removed, leaving the nuclei in up to 200 μL buffer volume. Pelleted nuclei were then gently resuspended and filtered through a 40 μm FlowMi Tip Strainer (SP Scienceware cat. #H13680–0040) into a 1.5 mL LoBind tube. Extracted nuclei were immediately placed on ice.

Generation of snRNA-seq data—All samples were sorted using the Sony SH800 sorter into 0.04% BSA in PBS and counted on a Countess II FL automated cell counter using 1:1 AOPI stain (Nexcelom CS2–0106). Nuclei were pelleted for 5 min at 500 G in a refrigerated swinging bucket centrifuge and resuspended to a target concentration of 1,000 nuclei/μL prior to loading the 10× Chromium chips. Single-cell gene expression libraries were prepared using the Single Cell 3′ Gene Expression kit v3.1 (10× Genomics, #1000268) according to manufacturer's instructions. Libraries were sequenced on an Illumina Nextseq2000 using 100-cycle kits to a mean depth of ~35,000 reads per cell.

QUANTIFICATION AND STATISTICAL ANALYSIS

Analysis of bulk RNA-seq data—Reads from samples with RIN ≥ 5.5 were aligned to the hg38 human genome using STAR v2.7.6,[55] allowing for a 4% mismatch rate and up to 100 alignments per read to ensure capture of young transposon sequences. Abundance of gene and transposon sequences was calculated with TEtranscripts v2.2.3.[56] For differential expression analysis, we employed DESeq2,[51] using the DESeq normalization strategy and negative binomial modeling. B-H corrected FDR p value threshold of $p < 0.05$ was used to determine significance. For heatmap visualization, the reads were normalized using a variance stabilizing transformation in DESeq2. Alternative splicing was performed using Leafcutter[36] following the procedure as outlined in their documentation (http://davidaknowles.github.io/leafcutter/articles/Usage.html).

Analysis of snRNA-seq data (CellRanger-TE)—In addition to the snRNA-seq data described above, raw data from single nucleus sequencing of ALS and pathologically normal (PN) samples[38] (GEO: GSE174332) were downloaded from GEO. All libraries were analyzed using 10x Genomics Cell Ranger version 5.0.1,[58] using a GRCh38 reference genome database built with GENCODE v35 gene annotation[59] and RepeatMasker TE annotation.[60] Libraries with mapping rate <80%, intergenic (non-gene and non-TE) reads >5% and total read count <10 million were excluded from subsequent analyses. Gene expression estimates of each cell were correlated using Spearman correlation with identified

cell types from the human M1 10x dataset,[61] with the cell labelled according to the M1 cell type with the highest correlation.

The processed count matrices from Cell Ranger were integrated in Seurat,[54] with PCA and UMAP performed on the combined dataset. Differential analysis between ALS and control cell types was performed using a modified DESeq2 protocol, where size factor was calculated using the scran R package,[53] a gamma-Poisson GLM (glmGamPoi R package[52]) was used to fit the data, and a likelihood ratio test was used for the statistical comparison. Genes and TEs with an FDR < 0.05 were considered differentially expressed.

Gene set enrichment analysis—Gene set enrichment was performed using GSEA[23] and a manually curated set of gene pathways. From MSigDB,[23] this included the Hallmark and KEGG database. From the curated literature, we included differentially regulated genes from TDP-43 depleted neurons,[29] KRAB zinc fingers,[62] and several lists of genes dysregulated in astrocytes[25,44] and microglia[24,49,50] from diseased models or other pro-inflammatory stimuli. Gene sets with an FDR < 0.05 were considered significantly enriched/depleted.

Statistical data analysis—All enrichments calculations for discrete variables (e.g., subtype and clinical variables, subtype and cell types or cell clusters) used Fisher's Exact test, with an alpha = 0.05 and a Benjamini-Hochberg adjustment for Type I error. All linear correlations used Pearson's R with an alpha = 0.05 and a Benjamini-Hochberg adjustment for Type I error. Statistics for differential expression of bulk and single-cell data are described in each section and all involve an alpha = 0.05 and a Benjamini-Hochberg adjustment for Type I error. Statistics for differential splice junction usage of bulk data are described above, using an alpha = 0.05 and a Benjamini-Hochberg adjustment for type I error. Statistics for GSEA of bulk and single-cell data use an alpha = 0.05 and a Benjamini-Hochberg adjustment for type I error.

Training and validation of neural networks—WGCNA[22] was used to create eigengenes for the ALS bulk cortex data with a power parameter of 9, resulting in 40 modules. One of these modules was determined to be entirely comprised of sex-specific genes and was removed for the purpose of classifier training. Eigenvalues for each ALS sample were calculated for the remaining 39 modules and used as input to DANCer. DANCer is a feedforward multilayer perceptron composed of 39 input nodes, a fully connected hidden layer of 6 nodes, and a softmax output layer of 3 nodes that report the probability of classification to each subtype. The classifier was built using the Keras library[63] as a wrapper for TensorFlow version 2.[64] The labeled ALS subtype samples (n = 176) from Tam et al.[1] were split 80/20 for training/testing, respectively. Training proceeded for 200 epochs with a loss function of categorical cross entropy with callbacks for early stopping to prevent overfitting.

scDANCer is a single-cell (nucleus) adapted version of DANCer that follows the same architecture and training regime. Single nucleus transcriptomes were compressed into pseudobulk representations by summing unique molecular identifier (UMI) transcript counts for all detected genes in all cells of a given sample. scDANCer was trained using the

same original bulk RNA-seq data used for DANCer, but the list of input genes was restricted to the set of detectable genes in our pseudobulked snRNA-seq transcriptomes. The WGCNA network and eigenmodules were recomputed using those pseudobulked snRNA-seq transcriptomes, again using a power parameter of 9, which resulted in 17 eigenmodules. The scDANCer architecture involved 17 input nodes, 5 fully connected inner nodes, and a softmax output layer of 3 nodes. The labeled ALS subtype samples ($n = 176$) from Tam et al.[1] were split 80/20 for training/testing, respectively. Training proceeded for 200 epochs with a loss function of categorical cross entropy with callbacks for early stopping to prevent overfitting.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

We wish to thank the Target ALS Human Postmortem Tissue Core for providing postmortem brain and spinal cord samples, the CSHL Single Cell Core Facility and CSHL Sequencing Facility (supported by NIH Cancer Center support grant 5P30CA045508) for single-cell data generation and sequencing support, and the Harms laboratory for performing repeat-primed PCR to identify C9ORF72 expansions in the Target ALS samples. Schematic images were created with BioRender.com. M.G.H. was supported by grants from the Chan Zuckerberg Initiative (SVCF2018–191863), the Ride For Life Foundation, the Rita Allen Foundation (of which M.G.H. is a scholar), and the NIH/NINDS (RF1NS118570 to M.G.H. and H.P.). H.P. was supported by grants from the NIH/NINDS (RF1NS118570 to M.G.H. and H.P., R01NS118183, and R01NS116350). K.O. acknowledges funding from the National Science Foundation Graduate Research Fellowship and an NIH training grant (2T32GM065094–16). All NYGC ALS Consortium activities are supported by the ALS Association (15-LGCA-234) and the Tow Foundation.

REFERENCES

1. Tam OH, Rozhkov NV, Shaw R, Kim D, Hubbard I, Fennessey S, Propp N, NYGC ALS Consortium; Fagegaltier D, and Harris BT (2019). Postmortem Cortex Samples Identify Distinct Molecular Subtypes of ALS: Retrotransposon Activation, Oxidative Stress, and Activated Glia. Cell Rep. 29, 1164–1177.e5. 10.1016/j.celrep.2019.09.066. [PubMed: 31665631]
2. Feldman EL, Goutman SA, Petri S, Mazzini L, Savelieff MG, Shaw PJ, and Sobue G (2022). Amyotrophic lateral sclerosis. Lancet 400, 1363–1380. 10.1016/S0140-6736(22)01272-7. [PubMed: 36116464]
3. Neumann M, Sampathu DM, Kwong LK, Truax AC, Micsenyi MC, Chou TT, Bruce J, Schuck T, Grossman M, Clark CM, et al. (2006). Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. Science 314, 130–133. 10.1126/science.1134108. [PubMed: 17023659]
4. Arai T, Hasegawa M, Akiyama H, Ikeda K, Nonaka T, Mori H, Mann D, Tsuchiya K, Yoshida M, Hashizume Y, and Oda T (2006). TDP-43 is a component of ubiquitin-positive tau-negative inclusions in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. Biochem. Biophys. Res. Commun. 351, 602–611. 10.1016/j.bbrc.2006.10.093. [PubMed: 17084815]
5. Ghasemi M, and Brown RH Jr. (2018). Genetics of Amyotrophic Lateral Sclerosis. Cold Spring Harb. Perspect. Med. 8, a024125. 10.1101/cshperspect.a024125. [PubMed: 28270533]
6. Van Daele SH, Moisse M, van Vugt JJFA, Zwamborn RAJ, van der Spek R, van Rheenen W, Van Eijk K, Kenna K, Corcia P, Vourc'h P, et al. (2023). Genetic variability in sporadic amyotrophic lateral sclerosis. Brain 146, 3760–3769. 10.1093/brain/awad120. [PubMed: 37043475]
7. Ravits J, Appel S, Baloh RH, Barohn R, Brooks BR, Elman L, Floeter MK, Henderson C, Lomen-Hoerth C, Macklis JD, et al. (2013). Deciphering amyotrophic lateral sclerosis: what phenotype, neuropathology and genetics are telling us about pathogenesis. Amyotroph. Lateral Scler. Frontotemporal Degener. 14, 5–18. 10.3109/21678421.2013.778548. [PubMed: 23678876]

8. Nelson PT, Dickson DW, Trojanowski JQ, Jack CR, Boyle PA, Arfanakis K, Rademakers R, Alafuzoff I, Attems J, Brayne C, et al. (2019). Limbic-predominant age-related TDP-43 encephalopathy (LATE): consensus working group report. Brain 142, 1503–1527. 10.1093/brain/awz099. [PubMed: 31039256]
9. Meneses A, Koga S, O'Leary J, Dickson DW, Bu G, and Zhao N (2021). TDP-43 Pathology in Alzheimer's Disease. Mol. Neurodegener. 16, 84. 10.1186/s13024-021-00503-x. [PubMed: 34930382]
10. Kwong LK, Neumann M, Sampathu DM, Lee VMY, and Trojanowski JQ (2007). TDP-43 proteinopathy: the neuropathology underlying major forms of sporadic and familial frontotemporal lobar degeneration and motor neuron disease. Acta Neuropathol. 114, 63–70. 10.1007/s00401-007-0226-5. [PubMed: 17492294]
11. Couratier P, Corcia P, Lautrette G, Nicol M, and Marin B (2017). ALS and frontotemporal dementia belong to a common disease spectrum. Rev. Neurol. 173, 273–279. 10.1016/j.neurol.2017.04.001. [PubMed: 28449882]
12. Gendron TF, and Petrucelli L (2018). Disease Mechanisms of C9ORF72 Repeat Expansions. Cold Spring Harb. Perspect. Med. 8, a024224. 10.1101/cshperspect.a024224. [PubMed: 28130314]
13. Marriott H, Kabiljo R, Hunt GP, Khleifat AA, Jones A, Troakes C, Project MinE ALS Sequencing Consortium; TargetALS Sequencing Consortium; Pfaff AL, and Quinn JP (2023). Unsupervised machine learning identifies distinct ALS molecular subtypes in post-mortem motor cortex and blood expression data. Acta Neuropathol. Commun. 11, 208. 10.1186/s40478-023-01686-8. [PubMed: 38129934]
14. Eshima J, O'Connor SA, Marschall E, NYGC ALS Consortium; Bowser R, Plaisier CL, and Smith BS (2023). Molecular subtypes of ALS are associated with differences in patient prognosis. Nat. Commun. 14, 95. 10.1038/s41467-022-35494-w. [PubMed: 36609402]
15. Krupp S, Hubbard I, Tam O, Hammell GM, and Dubnau J (2023). TDP-43 pathology in Drosophila induces glial-cell type specific toxicity that can be ameliorated by knock-down of SF2/SRSF1. PLoS Genet. 19, e1010973. 10.1371/journal.pgen.1010973. [PubMed: 37747929]
16. Humphrey J, Venkatesh S, Hasan R, Herb JT, de Paiva Lopes K, Küçükali F, Byrska-Bishop M, Evani US, Narzisi G, Fagegaltier D, et al. (2023). Integrative transcriptomic analysis of the amyotrophic lateral sclerosis spinal cord implicates glial activation and suggests new risk genes. Nat. Neurosci. 26, 150–162. 10.1038/s41593-022-01205-3. [PubMed: 36482247]
17. Conlon EG, Fagegaltier D, Agius P, Davis-Porada J, Gregory J, Hubbard I, Kang K, Kim D, and New York Genome Center ALS Consortium; and Phatnani H. (2018). Unexpected similarities between C9ORF72 and sporadic forms of ALS/FTD suggest a common disease mechanism. Elife 7, e37754. 10.7554/eLife.37754. [PubMed: 30003873]
18. Placek K, Benatar M, Wuu J, Rampersaud E, Hennessy L, Van Deerlin VM, Grossman M, Irwin DJ, Elman L, McCluskey L, et al. (2021). Machine learning suggests polygenic risk for cognitive dysfunction in amyotrophic lateral sclerosis. EMBO Mol. Med. 13, e12595. 10.15252/emmm.202012595. [PubMed: 33270986]
19. Project MinE ALS Sequencing Consortium (2018). Project MinE: study design and pilot analyses of a large-scale whole-genome sequencing study in amyotrophic lateral sclerosis. Eur. J. Hum. Genet. 26, 1537–1546. 10.1038/s41431-018-0177-4. [PubMed: 29955173]
20. Baxi EG, Thompson T, Li J, Kaye JA, Lim RG, Wu J, Ramamoorthy D, Lima L, Vaibhav V, Matlock A, et al. (2022). Answer ALS, a large-scale resource for sporadic and familial ALS combining clinical and multi-omics data from induced pluripotent cell lines. Nat. Neurosci. 25, 226–237. 10.1038/s41593-021-01006-0. [PubMed: 35115730]
21. Gaujoux R, and Seoighe C (2010). A flexible R package for nonnegative matrix factorization. BMC Bioinf. 11, 367. 10.1186/1471-2105-11-367.
22. Langfelder P, and Horvath S (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinf. 9, 559. 10.1186/1471-2105-9-559.
23. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, and Mesirov JP (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA 102, 15545–15550. 10.1073/pnas.0506580102. [PubMed: 16199517]

24. Chiu IM, Morimoto ETA, Goodarzi H, Liao JT, O'Keeffe S, Phatnani HP, Muratet M, Carroll MC, Levy S, Tavazoie S, et al. (2013). A neurodegeneration-specific gene-expression signature of acutely isolated microglia from an amyotrophic lateral sclerosis mouse model. Cell Rep. 4, 385–401. 10.1016/j.celrep.2013.06.018. [PubMed: 23850290]
25. Zamanian JL, Xu L, Foo LC, Nouri N, Zhou L, Giffard RG, and Barres BA (2012). Genomic analysis of reactive astrogliosis. J. Neurosci. 32, 6391–6410. 10.1523/JNEUROSCI.6221-11.2012. [PubMed: 22553043]
26. Catanese A, Rajkumar S, Sommer D, Masrori P, Hersmus N, Van Damme P, Witzel S, Ludolph A, Ho R, Boeckers TM, and Mulaw M (2023). Multiomics and machine-learning identify novel transcriptional and mutational signatures in amyotrophic lateral sclerosis. Brain 146, 3770–3782. 10.1093/brain/awad075. [PubMed: 36883643]
27. Ho R, Workman MJ, Mathkar P, Wu K, Kim KJ, O'Rourke JG, Kellogg M, Montel V, Banuelos MG, Arogundade OA, et al. (2021). Cross-Comparison of Human iPSC Motor Neuron Models of Familial and Sporadic ALS Reveals Early and Convergent Transcriptomic Disease Signatures. Cell Syst. 12, 159–175.e9. 10.1016/j.cels.2020.10.010. [PubMed: 33382996]
28. Ziff OJ, Neeves J, Mitchell J, Tyzack G, Martinez-Ruiz C, Luisier R, Chakrabarti AM, McGranahan N, Litchfield K, Boulton SJ, et al. (2023). Integrated transcriptome landscape of ALS identifies genome instability linked to TDP-43 pathology. Nat. Commun. 14, 2176. 10.1038/s41467-023-37630-6. [PubMed: 37080969]
29. Liu EY, Russ J, Cali CP, Phan JM, Amlie-Wolf A, and Lee EB (2019). Loss of Nuclear TDP-43 Is Associated with Decondensation of LINE Retrotransposons. Cell Rep. 27, 1409–1421.e6. 10.1016/j.celrep.2019.04.003. [PubMed: 31042469]
30. Krach F, Batra R, Wheeler EC, Vu AQ, Wang R, Hutt K, Rabin SJ, Baughn MW, Libby RT, Diaz-Garcia S, et al. (2018). Transcriptome-pathology correlation identifies interplay between TDP-43 and the expression of its kinase CK1E in sporadic ALS. Acta Neuropathol. 136, 405–423. 10.1007/s00401-018-1870-7. [PubMed: 29881994]
31. Klim JR, Williams LA, Limone F, Guerra San Juan I, Davis-Dusenbery BN, Mordes DA, Burberry A, Steinbaugh MJ, Gamage KK, Kirchner R, et al. (2019). ALS-implicated protein TDP-43 sustains levels of STMN2, a mediator of motor neuron growth and repair. Nat. Neurosci. 22, 167–179. 10.1038/s41593-018-0300-4. [PubMed: 30643292]
32. Melamed Z, López-Erauskin J, Baughn MW, Zhang O, Drenner K, Sun Y, Freyermuth F, McMahon MA, Beccari MS, Artates JW, et al. (2019). Premature polyadenylation-mediated loss of stathmin-2 is a hallmark of TDP-43-dependent neurodegeneration. Nat. Neurosci. 22, 180–190. 10.1038/s41593-018-0293-z. [PubMed: 30643298]
33. Ma XR, Prudencio M, Koike Y, Vatsavayai SC, Kim G, Harbinski F, Briner A, Rodriguez CM, Guo C, Akiyama T, et al. (2022). TDP-43 represses cryptic exon inclusion in the FTD-ALS gene UNC13A. Nature 603, 124–130. 10.1038/s41586-022-04424-7. [PubMed: 35197626]
34. Brown AL, Wilkins OG, Keuss MJ, Hill SE, Zanovello M, Lee WC, Bampton A, Lee FCY, Masino L, Qi YA, et al. (2022). TDP-43 loss and ALS-risk SNPs drive mis-splicing and depletion of UNC13A. Nature 603, 131–137. 10.1038/s41586-022-04436-3. [PubMed: 35197628]
35. Mehta PR, Brown AL, Ward ME, and Fratta P (2023). The era of cryptic exons: implications for ALS-FTD. Mol. Neurodegener. 18, 16. 10.1186/s13024-023-00608-5. [PubMed: 36922834]
36. Li YI, Knowles DA, Humphrey J, Barbeira AN, Dickinson SP, Im HK, and Pritchard JK (2018). Annotation-free quantification of RNA splicing using LeafCutter. Nat. Genet. 50, 151–158. 10.1038/s41588-017-0004-9. [PubMed: 29229983]
37. Imaizumi K, Ideno H, Sato T, Morimoto S, and Okano H (2022). Pathogenic Mutation of TDP-43 Impairs RNA Processing in a Cell Type-Specific Manner: Implications for the Pathogenesis of ALS/FTLD. eNeuro 9, ENEUR.O0061–22.2022. 10.1523/ENEURO.0061-22.2022.
38. Pineda SS, Lee H, Ulloa-Navas MJ, Linville RM, Garcia FJ, Galani K, Engelberg-Cook E, Castanedes MC, Fitzwalter BE, Pregent LJ, et al. (2024). Single-cell dissection of the human motor and prefrontal cortices in ALS and FTLD. Cell 187, 1971–1989.e16. 10.1016/j.cell.2024.02.031. [PubMed: 38521060]
39. BRAIN Initiative Cell Census Network BICCN (2021). A multimodal cell census and atlas of the mammalian primary motor cortex. Nature 598, 86–102. 10.1038/s41586-021-03950-0. [PubMed: 34616075]

40. Tollervey JR, Curk T, Rogelj B, Briese M, Cereda M, Kayikci M, König J, Hortobágyi T, Nishimura AL, Zupunski V, et al. (2011). Characterizing the RNA targets and position-dependent splicing regulation by TDP-43. Nat. Neurosci. 14, 452–458. 10.1038/nn.2778. [PubMed: 21358640]
41. Polymenidou M, Lagier-Tourenne C, Hutt KR, Huelga SC, Moran J, Liang TY, Ling SC, Sun E, Wancewicz E, Mazur C, et al. (2011). Long pre-mRNA depletion and RNA missplicing contribute to neuronal vulnerability from loss of TDP-43. Nat. Neurosci. 14, 459–468. 10.1038/nn.2779. [PubMed: 21358643]
42. Hasel P, and Liddelow SA (2021). Astrocytes. Curr. Biol. 31, R326–R327. 10.1016/j.cub.2021.01.056. [PubMed: 33848482]
43. Traag VA, Waltman L, and van Eck NJ (2019). From Louvain to Leiden: guaranteeing well-connected communities. Sci. Rep. 9, 5233. 10.1038/s41598-019-41695-z. [PubMed: 30914743]
44. Habib N, McCabe C, Medina S, Varshavsky M, Kitsberg D, Dvir-Szternfeld R, Green G, Dionne D, Nguyen L, Marshall JL, et al. (2020). Disease-associated astrocytes in Alzheimer's disease and aging. Nat. Neurosci. 23, 701–706. 10.1038/s41593-020-0624-8. [PubMed: 32341542]
45. Hasel P, Rose IVL, Sadick JS, Kim RD, and Liddelow SA (2021). Neuroinflammatory astrocyte subtypes in the mouse brain. Nat. Neurosci. 24, 1475–1487. 10.1038/s41593-021-00905-6. [PubMed: 34413515]
46. Akcimen F, Lopez ER, Landers JE, Nath A, Chio A, Chia R, and Traynor BJ (2023). Amyotrophic lateral sclerosis: translating genetic discoveries into therapies. Nat. Rev. Genet. 24, 642–658. 10.1038/s41576-023-00592-y. [PubMed: 37024676]
47. Hamby ME, Coppola G, Ao Y, Geschwind DH, Khakh BS, and Sofroniew MV (2012). Inflammatory mediators alter the astrocyte transcriptome and calcium signaling elicited by multiple G-protein-coupled receptors. J. Neurosci. 32, 14489–14510. 10.1523/JNEUROSCI.1256-12.2012. [PubMed: 23077035]
48. Hickman S, Izzy S, Sen P, Morsett L, and El Khoury J (2018). Microglia in neurodegeneration. Nat. Neurosci. 21, 1359–1369. 10.1038/s41593-018-0242-x. [PubMed: 30258234]
49. Keren-Shaul H, Spinrad A, Weiner A, Matcovitch-Natan O, DvirSzternfeld R, Ulland TK, David E, Baruch K, Lara-Astaiso D, Toth B, et al. (2017). A Unique Microglia Type Associated with Restricting Development of Alzheimer's Disease. Cell 169, 1276–1290.e17. 10.1016/j.cell.2017.05.018. [PubMed: 28602351]
50. Marschallinger J, Iram T, Zardeneta M, Lee SE, Lehallier B, Haney MS, Pluvinage JV, Mathur V, Hahn O, Morgens DW, et al. (2020). Lipid-droplet-accumulating microglia represent a dysfunctional and proinflammatory state in the aging brain. Nat. Neurosci. 23, 194–208. 10.1038/s41593-019-0566-1. [PubMed: 31959936]
51. Love MI, Huber W, and Anders S (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550. 10.1186/s13059-014-0550-8. [PubMed: 25516281]
52. Ahlmann-Eltze C, and Huber W (2021). glmGamPoi: fitting Gamma-Poisson generalized linear models on single cell count data. Bioinformatics 36, 5701–5702. 10.1093/bioinformatics/btaa1009. [PubMed: 33295604]
53. Lun ATL, McCarthy DJ, and Marioni JC (2016). A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. F1000Res. 5, 2122. 10.12688/f1000research.9501.2. [PubMed: 27909575]
54. Butler A, Hoffman P, Smibert P, Papalexi E, and Satija R (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat. Biotechnol. 36, 411–420. 10.1038/nbt.4096. [PubMed: 29608179]
55. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, and Gingeras TR (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21. 10.1093/bioinformatics/bts635. [PubMed: 23104886]
56. Jin Y, Tam OH, Paniagua E, and Hammell M (2015). TEtranscripts: a package for including transposable elements in differential expression analysis of RNA-seq datasets. Bioinformatics 31, 3593–3599. 10.1093/bioinformatics/btv422. [PubMed: 26206304]

57. Maitra M, Nagy C, Chawla A, Wang YC, Nascimento C, Suderman M, Théroux JF, Mechawar N, Ragoussis J, and Turecki G (2021). Extraction of nuclei from archived postmortem tissues for single-nucleus sequencing applications. Nat. Protoc. 16, 2788–2801. 10.1038/s41596-021-00514-4. [PubMed: 33972803]
58. Zheng GXY, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, Ziraldo SB, Wheeler TD, McDermott GP, Zhu J, et al. (2017). Massively parallel digital transcriptional profiling of single cells. Nat. Commun. 8, 14049. 10.1038/ncomms14049. [PubMed: 28091601]
59. Frankish A, Carbonell-Sala S, Diekhans M, Jungreis I, Loveland JE, Mudge JM, Sisu C, Wright JC, Arnan C, Barnes I, et al. (2023). GENCODE: reference annotation for the human and mouse genomes in 2023. Nucleic Acids Res. 51, D942–D949. 10.1093/nar/gkac1071. [PubMed: 36420896]
60. Smit A, Hubley R & Green P (2013–2015). RepeatMasker Open-4.0. http://www.repeatmasker.org.
61. Bakken TE, Jorstad NL, Hu Q, Lake BB, Tian W, Kalmbach BE, Crow M, Hodge RD, Krienen FM, Sorensen SA, et al. (2021). Comparative cellular analysis of motor cortex in human, marmoset and mouse. Nature 598, 111–119. 10.1038/s41586-021-03465-8. [PubMed: 34616062]
62. Imbeault M, Helleboid PY, and Trono D (2017). KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks. Nature 543, 550–554. 10.1038/nature21683. [PubMed: 28273063]
63. Chollet F, (2015). Keras. https://github.com/fchollet/keras.
64. Martín Abadi AA, Paul B, Brevdo E, Zhifeng Chen CC, Corrado GS, Davis A, Jeffrey Dean MD, Sanjay G, Goodfellow I, Andrew Harp GI, et al. (2015). TensorFlow: Large-scale machine learning on heterogeneous systems. tensorflow.org.

Highlights

- ALS tissues are dominated by signatures of inflammation in cord and stress in cortex
- Disease duration correlates with TDP-43 dysfunction in cortex and inflammation in cord
- Subtypes are recapitulated in all cells, with subtype-specific changes in neurons and glia
- A deep ALS neural net classifier (DANCer) identifies subtype in bulk and single-nucleus data

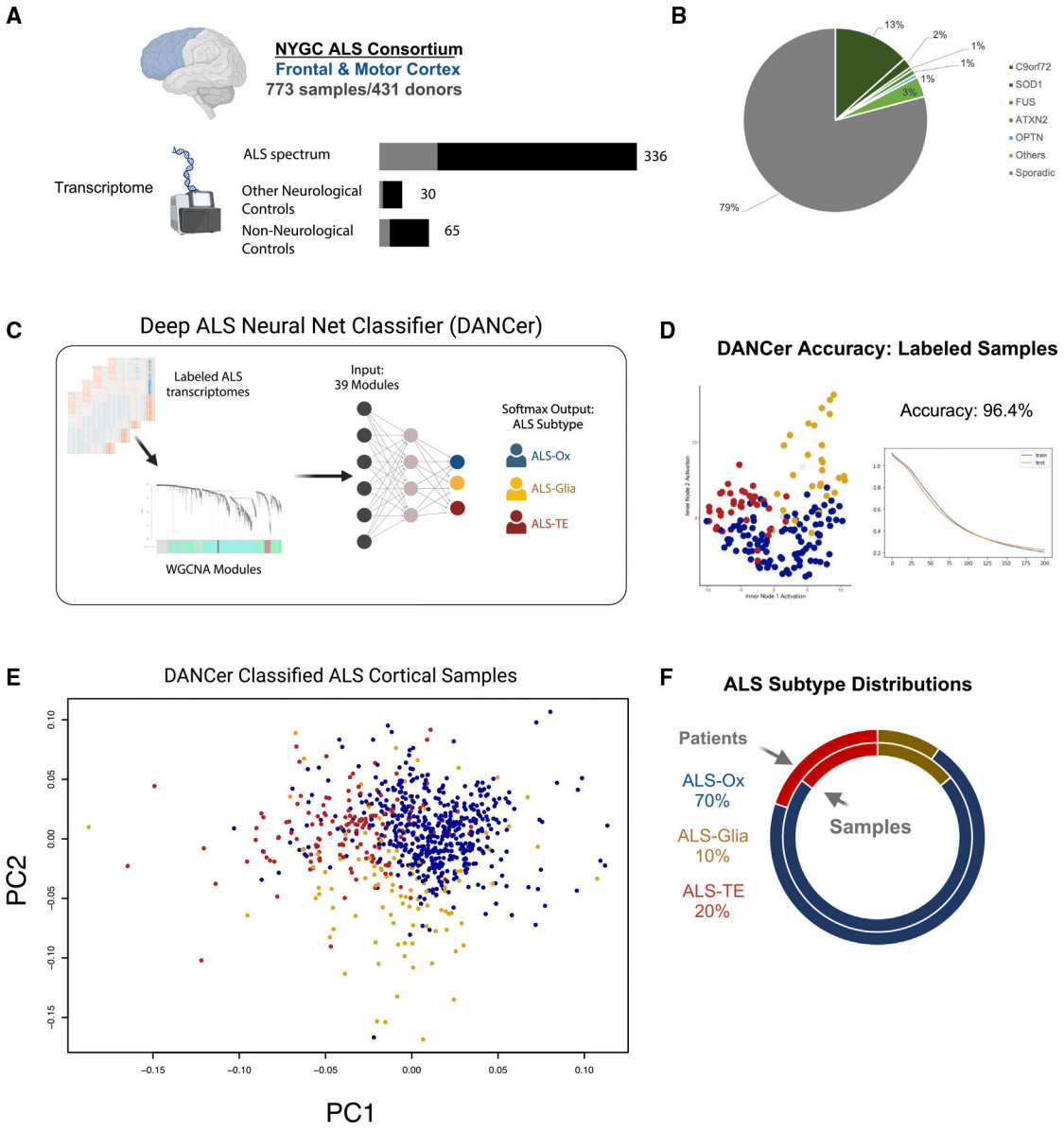


Figure 1. DANCer accurately assigns ALS molecular subtypes to an expanded NYGC ALS Consortium cohort of ALS cortical samples

(A) The 2023 NYGC ALS Consortium genomics profiling cohort consists of whole-genome and -transcriptome libraries from 773 samples of frontal and motor cortex from 431 donors.

(B) For cases on the ALS spectrum, 79% of all tissues were from sporadic ALS donors, while 13% carried known ALS-associated mutations.

(C) DANCer takes whole-transcriptome data as input and assigns ALS subtypes using a multilayer, fully connected neural network.

(D) DANCer accuracy on labeled samples (96.4%) is graphically displayed for ALS-Ox (blue), ALS-Glia (gold), and ALS-TE (red) samples.

(E) All DANCer-classified ALS spectrum samples are displayed in PCA space.

(F) The distribution across ALS subtypes is displayed at the patient level (outer circle) and sample level (inner circle).

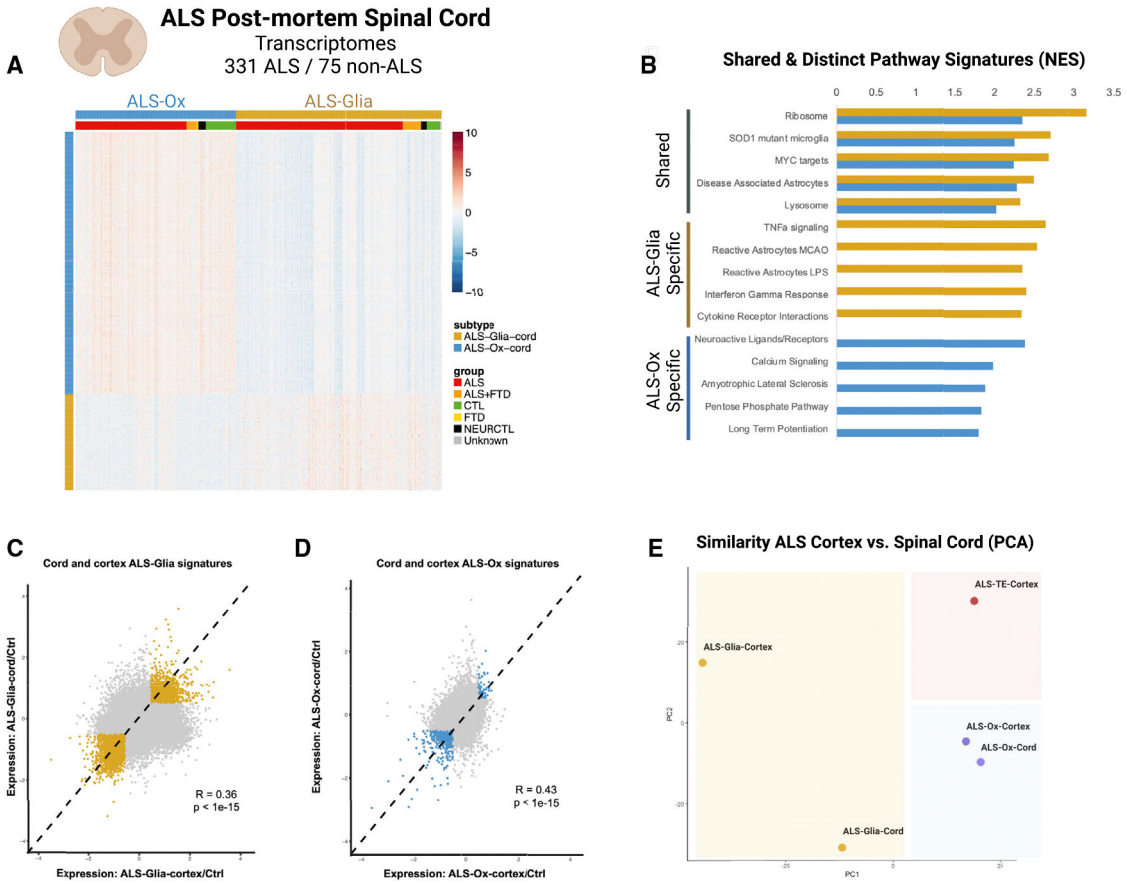


Figure 2. ALS spinal cord transcriptional patterns are dominated by inflammatory and oxidative stress pathways, showing strong overlap with ALS-Glia and ALS-Ox cortical subtypes

(A) The 2023 NYGC ALS Consortium cohort additionally contains 331 ALS and 75 non-ALS spinal cord samples. Non-negative matrix factorization (NMF) returned two transcriptome subtypes, with marker genes that indicated either oxidative stress (blue) or inflammatory (gold) gene expression patterns.

(B) Gene set enrichment analysis (GSEA) of these spinal cord subtypes returned pro-inflammatory pathways for the ALS-Glia-cord samples (gold) and classical ALS genes for ALS-Ox-cord (blue). Both subtypes showed significant enrichment for markers of SOD1 mutation-induced disease-associated microglia[24] (SOD1-DAMs).

(C) To stress the similarity of the inflammatory signatures in spinal cord and cortex samples, we plotted the log2 fold changes (log2FCs) for all genes in ALS-Glia cortical samples vs. controls (y axis) against all measured fold changes for ALS-Glia-cord samples vs. controls (x axis). Spearman's R coefficient and associated *p* value is shown

(D) A similar scatterplot of log2FC is presented for all DE genes in ALS-Ox cortical and spinal cord samples.

(E) PCA plots of all cord and cortex signatures in the ALS Consortium separate inflammatory (ALS-Glia) cord and cortex samples from non-inflammatory (ALS-TE and ALS-Ox) samples.

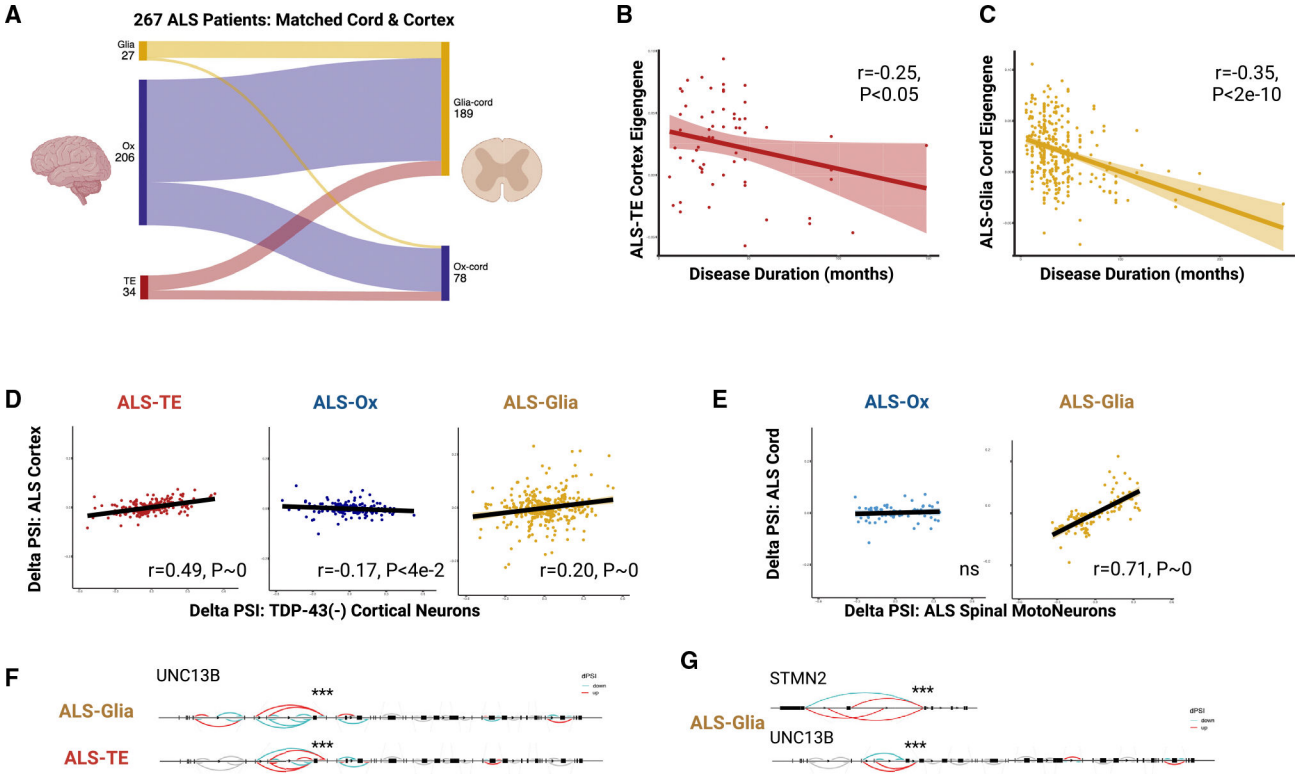


Figure 3. Integrative analysis across cortical and spinal cord tissues shows correlates of ALS molecular subtypes with clinical features

(A) For 267 of the ALS cases in the NYGC ALS Consortium cohort, matched spinal cord and cortex samples from the same donors enable a comparison of ALS subtypes in each region. ALS-Ox dominates calls for the cortex (70%) but were far less frequent in spinal cord (78 out of 267 = 29%). ALS-Glia subtypes dominate in the spinal cord (189 out of 267 = 71%).

(B) The ALS-TE subtype signature shows a strong negative correlation with disease duration, indicating shorter disease duration for these ALS cases.

(C) In spinal cord, inflammatory (ALS-Glia) gene expression signatures were negatively correlated with disease duration.

(D) The ALS-TE subtype showed the strongest signature for TDP-43 dysfunction in cortex, as measured by comparing splicing differences (percentage spliced in [PSI]) for all ALS-TE (red), ALS-Ox (blue), and ALS-Glia (gold) transcriptomes to the splicing differences (PSI) for sorted TDP-43-negative cortical neurons.[29] Spearman's R coefficients and *p* values are shown for each subtype.

(E) The ALS-Glia-cord subtype also showed a strong signature for TDP-43 dysfunction, as measured by comparing PSI values for ALS-Ox-cord (blue) and ALS-Glia-cord (gold) transcriptomes to PSI values for laser micro-dissected ALS motor neurons.[30] Spearman's R coefficients and *p* values are shown for each subtype.

(F) Altered splicing patterns for the known TDP-43 target gene UNC13B in ALS-TE and ALS-Glia cortex subtypes.

(G) Altered splicing patterns for the known TDP-43 target genes STMN2 and UNC13B in the ALS-Glia-cord subtype.

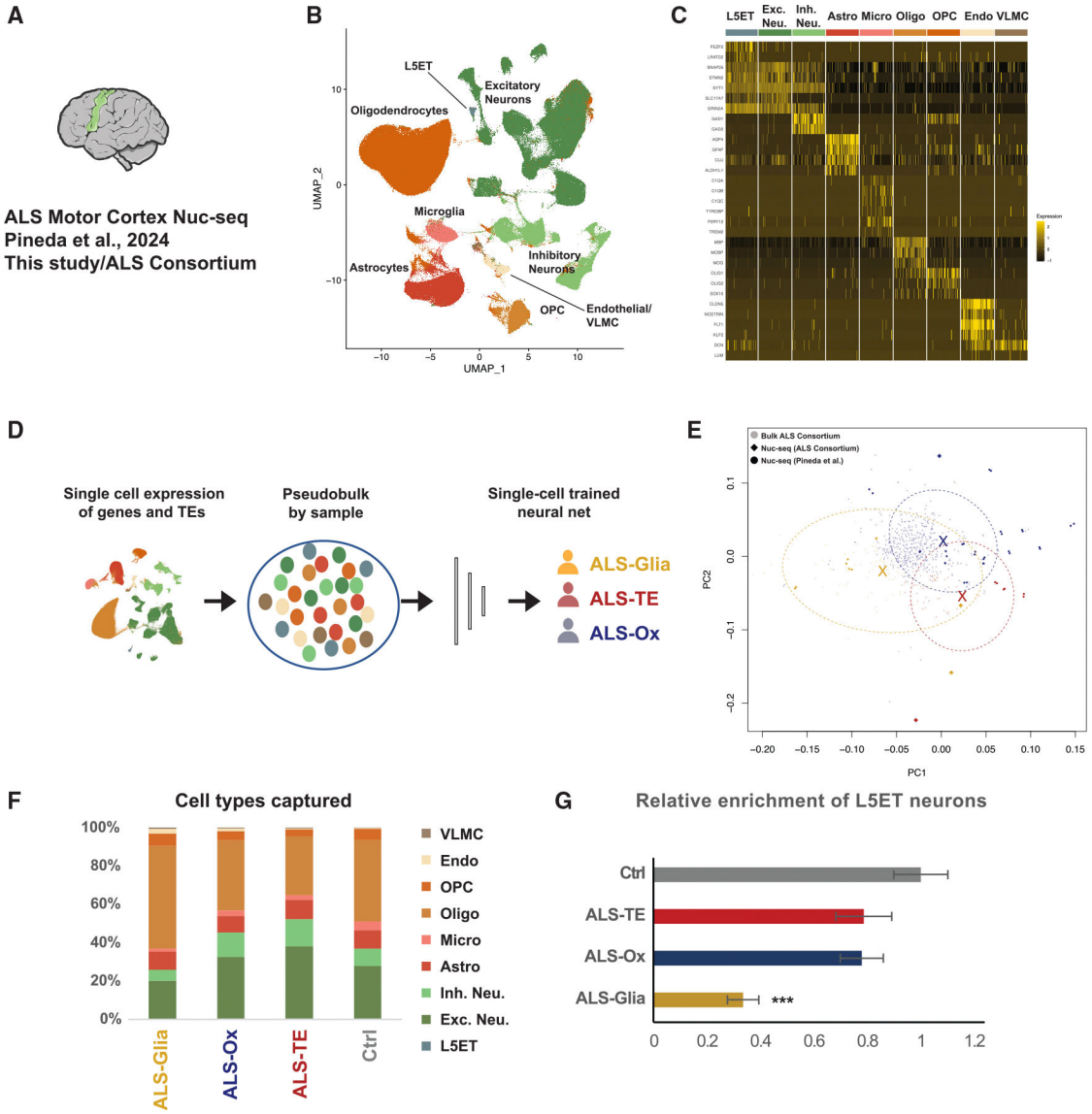


Figure 4. scDANCer shows cell-type composition differences between ALS subtypes

(A) Single-nucleus RNA-seq (snRNA-seq) transcriptomes were generated for six motor cortex samples from the NYGC ALS Consortium, which were integrated with an additional 58 ALS and control snRNA-seq motor cortex from Pineda et al.[38]

(B) A UMAP of the full integrated snRNA-seq dataset shows representation of all major cell types in the motor cortex, where L5ET neurons include motor neurons.

(C) A heatmap of cell-type-specific expression patterns for each of the identified cell types in the motor cortex.

(D) A schematic of the adapted snRNA-seq version of our ALS classifier, scDANCer.

(E) Both bulk RNA-seq and pseudobulked snRNA-seq transcriptomes occupy the same region of a PCA plot, which separates samples by ALS subtype and not by data platform (bulk/snRNA-seq) or dataset origin (this study/Pineda et al.[38]). Each sample is colored by ALS subtype (ALS-TE, red; ALS-Ox, blue; ALS-Glia, gold). Ellipses encompass 90% of all data from each subtype.

(F) Across ALS subtypes, ALS-Ox samples have the same cell-type composition as pathologically normal (PN) samples, while ALS-Glia and ALS-TE subtype samples show deviations.

(G) We see significant depletion in the relative fraction of L5ET/motor neurons present in ALS-Glia subtype samples. Data are represented as mean ± SEM. ***$p < 6e–6$ by Bonferroni corrected one-tailed t test.

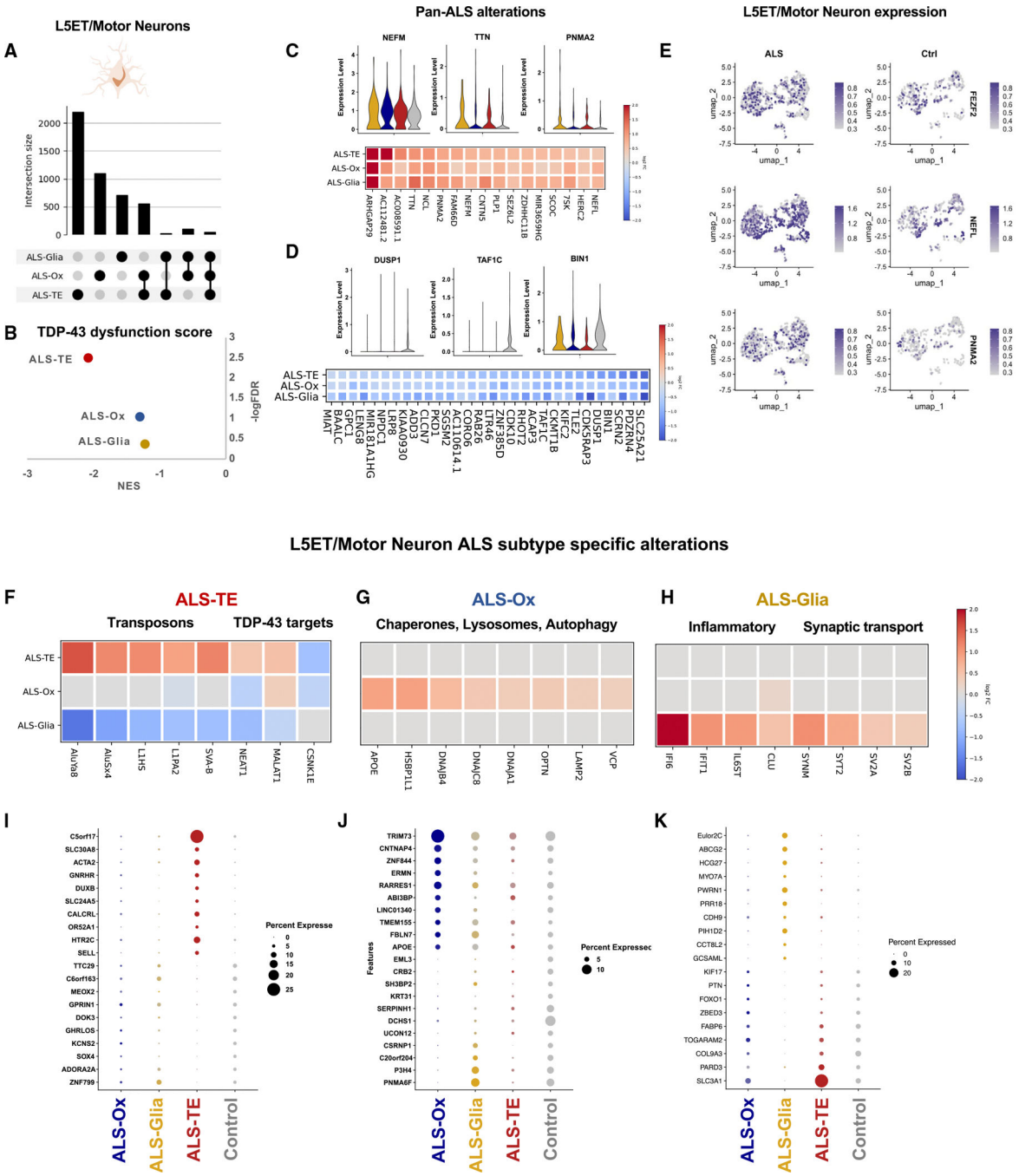


Figure 5. ALS L5ET single-neuron expression profiles show both ALS-wide and ALS-subtype-specific alterations

(A) Upset plots of shared and subtype-specific differentially expressed (DE) genes in L5ET neurons for each ALS subtype versus control samples.

(B) Normalized enrichment scores (NESs) comparing DE genes in each subtype and significant alterations in known TDP-43 target genes (identified from fluorescence-activated cell sorting [FACS]-sorted TDP-43-negative neurons[29]). Only L5ET neurons from ALS-TE subtype show significant alterations in TDP-43 targets.

(C) Shared upregulated genes across all ALS subtype L5ET neurons are shown as a heatmap of fold-change values, with violin plots for selected genes.
(D) Shared downregulated genes across all ALS subtype L5ET neurons are shown as a heatmap of fold-change values, with violin plots for selected genes.
(E) A heatmap of gene expression is overlayed for all L5ET cells in the UMAP for the motor neuron lineage gene FEZF2 as well as selected pan-ALS upregulated genes.
(F–H) A heatmap of log2FC values is shown for selected ALS subtype-specific alterations in L5ET neurons that were also subtype specific in bulk ALS subtype samples.
(I–K) Dot plots of the top 10 most subtype-specific up- and downregulated genes, where dot size corresponds to the percentage of cells expressing that gene for ALS-TE (I), ALS-Ox (J), and ALS-Glia (K) samples.

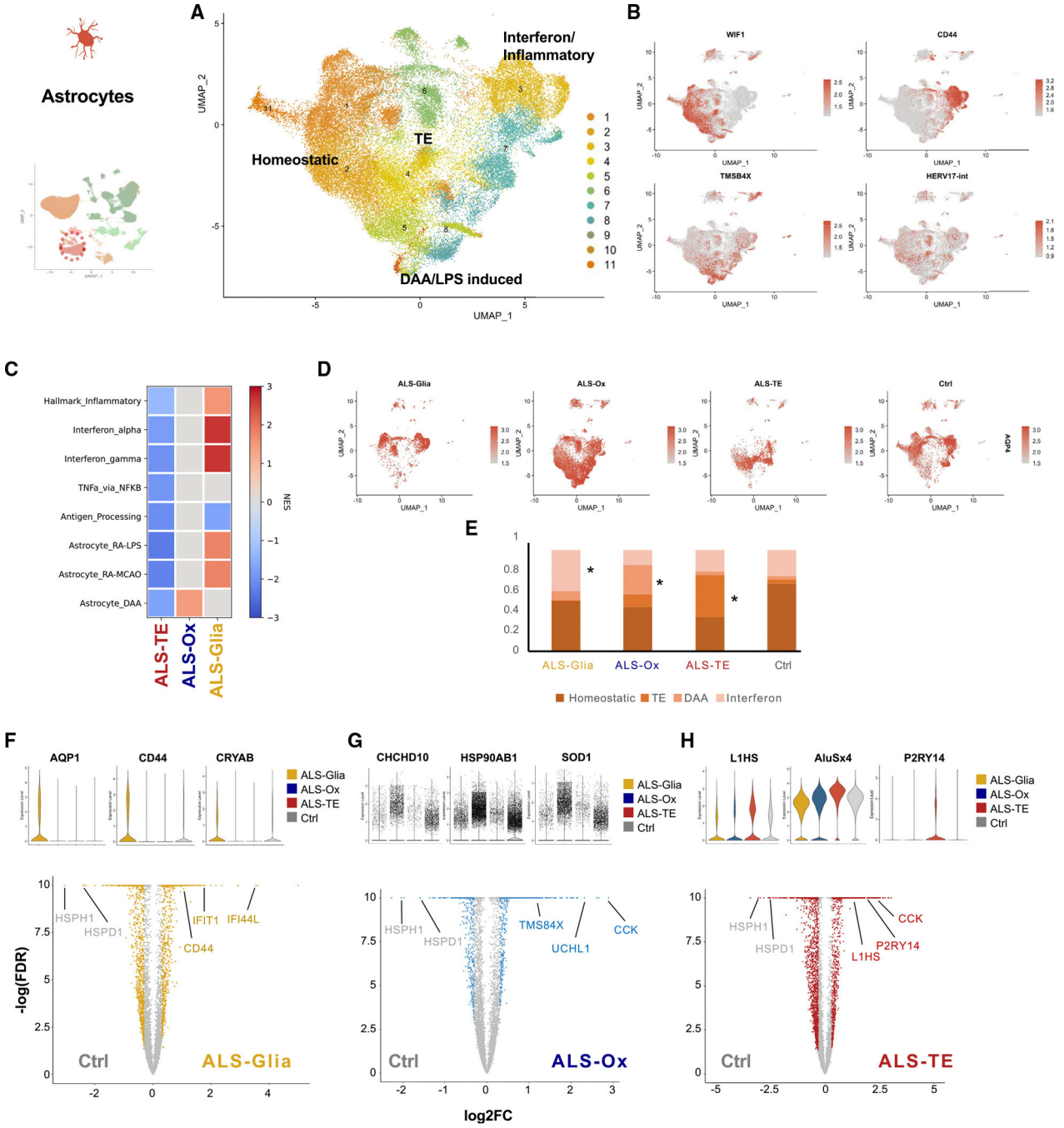


Figure 6. ALS subtypes show differences in pro-inflammatory pathway activation in astrocytes

(A) Astrocytes from all ALS and PN samples were re-clustered using Leiden, which returned four main categories marked by genes associated with interferon signaling, disease-associated astrocytes (DAAs), homeostatic astrocyte markers, and transposable elements (TEs).

(B) For each of these broad categories, a heatmap of marker-gene expression is overlayed on the UMAPs: homeostatic marker WIF1; pro-inflammatory marker CD44; DAA marker gene TMSB4X; and a young TE, HERV17.

(C) GSEA NESs for several published astrocyte pro-inflammatory gene sets across ALS-subtype astrocytes.
(D) Astrocytes from each ALS subtype occupy different regions of the UMAP space, although all express the pan-astrocyte marker AQP4.
(E) Differential occupancy of the UMAP space also corresponds to differential representation in the astrocyte state clusters. *$p < 0.05$ by two-tailed Fisher’s exact test.
(F) Comparing gene expression for all ALS-Glia astrocytes to controls identified several ALS-Glia-specific DE genes as well as downregulated genes shared by all ALS astrocytes (HSPH1, HSPD1).
(G) Volcano and violin plots for ALS-Ox subtype DE genes.
(H) Volcano and violin plots for ALS-TE DE genes in astrocytes.

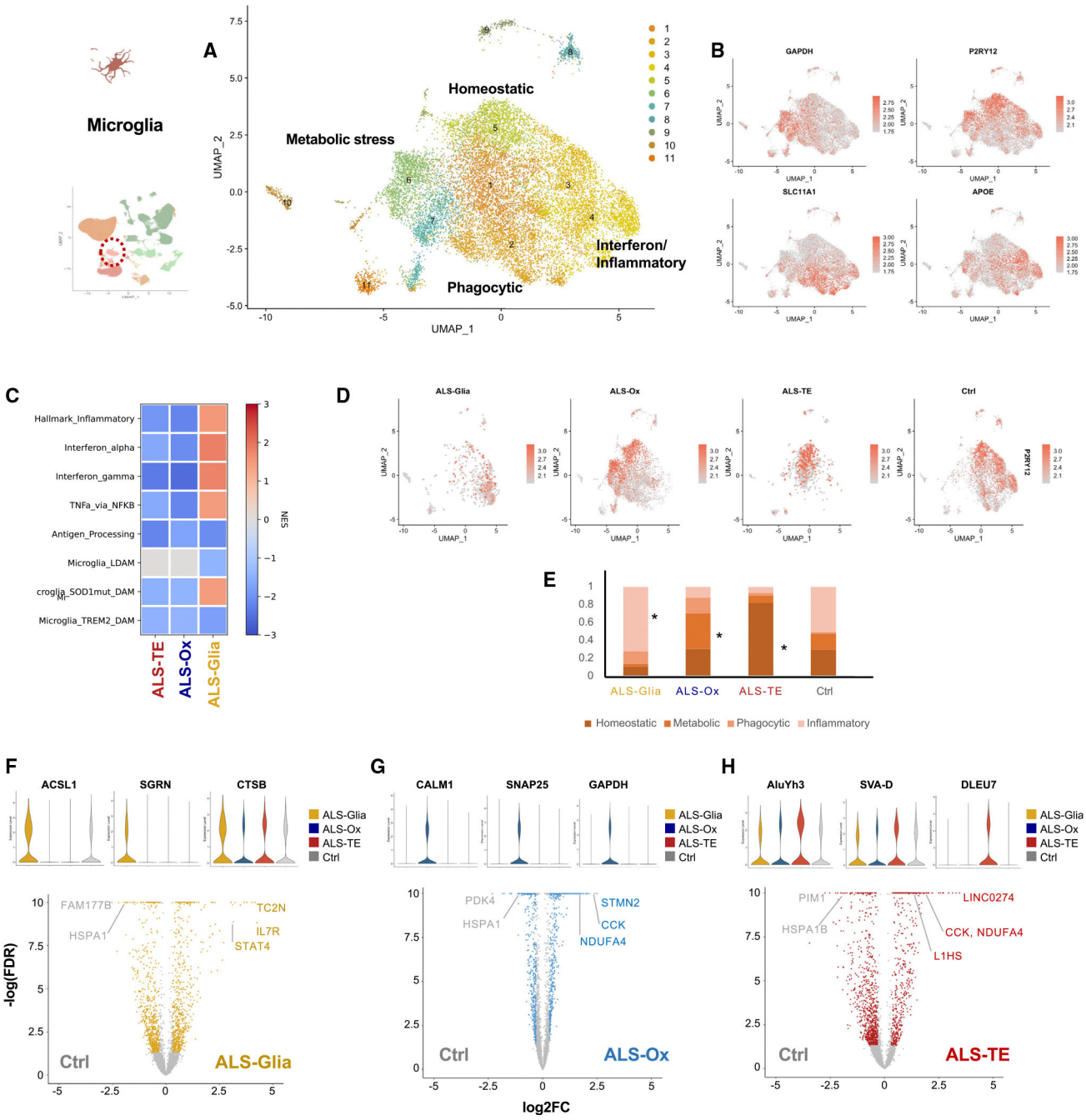


Figure 7. ALS subtypes show differences in pro-inflammatory pathway activation in microglia

(A) Leiden clusters for all ALS microglia group into four main categories marked by genes associated with interferon-driven pathways, phagocytic expression programs, metabolic stress, and homeostatic markers.

(B) For each of these broad categories, a heatmap shows markers for each cluster, including the metabolic marker GAPDH, the homeostatic marker P2RY12, the apolipoprotein APOE, and the phagocytic marker SLC11A1.

(C) GSEA NESs for several published microglia pro-inflammatory markers.

(D) Microglia from each ALS subtype occupy different regions of the UMAP space, highlighted using a pan-microglia marker, P2RY12.
(E) Differential occupancy of the UMAP space also corresponds to differential representation of microglial states across ALS subtypes. *$p < 0.05$ by two-tailed Fisher’s exact test.
(F) Comparing gene expression for all ALS-Glia microglia to controls identified several ALS-Glia-specific DE genes (ACSL1, SRGN, CTSB) as well as a downregulated heat shock gene shared by all ALS microglia (HSPA1B).
(G) Violin and volcano plots display ALS-Ox-specific DE genes.
(H) Violin and volcano plots highlight DE genes in ALS-TE microglia.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Postmortem frontal/motor cortex samples from ALS patients, neurological and non-neurological controls	NYGC ALS Consortium	https://www.nygenome.org/science-technology/collaborative-research-programs/neurodegenerative-disease-research/als-consortium/
Postmortem spinal cord samples from ALS patients, neurological and non-neurological controls	NYGC ALS Consortium	https://www.nygenome.org/science-technology/collaborative-research-programs/neurodegenerative-disease-research/als-consortium/
Chemicals, peptides, and recombinant proteins		
Trizol	ThermoFisher Scientific	Cat# 15596026
Optiprep iodixanol	Sigma Aldritch	Cat# D1556
1:1 AOPI stain	Nexcelom	Cat# CS2-0106
Critical commercial assays		
QIAGEN RNeasy Mini kit	QIAGEN	Cat# 74104
KAPA Stranded RNA-seq kit with RiboErase	Kapa Biosystems	Cat# 07962304001
NEXTflex RNA-seq Barcodes	PerkinElmer	Cat# NOVA-512915
Single Cell 3' Gene Expression kit v3.1	10x Genomics	Cat# 1000268
Deposited data		
NYGC postmortem frontal/motor cortex and spinal cord RNA-seq	NYGC ALS Consortium	GEO: GSE137810
Single nuclei RNA-seq dataset from C9 and sporadic ALS patients	Pineda et al.[38]	BioProject: PRJNA1073234
Single nuclei RNA-seq dataset from sporadic ALS patients	This study	GEO: GSE271156
Software and algorithms		
Cell Ranger v5.0.1	10x Genomics	https://www.10xgenomics.com/support/software/cell-ranger/latest
CellRangerTE (custom CellRanger reference)	This study	https://github.com/mhammell-laboratory/CellRangerTE https://doi.org/10.5281/zenodo.14750680
DANCer/scDANCer	This study	https://github.com/mhammell-laboratory/DANcer https://doi.org/10.5281/zenodo.14750751
DESeq2 v1.38.3	Love et al.[51]	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
glmGamPoi v1.10.2	Ahlmann-Eltze and Huber[52]	https://www.bioconductor.org/packages/release/bioc/html/glmGamPoi.html
GSEA v4.3.3	Subramanian et al.[23]	https://www.gsea-msigdb.org/gsea/index.jsp
Leafcutter v0.2.7	Li et al.[36]	https://github.com/davidaknowles/leafcutter
Leiden v0.9.1	Traag et al.[43]	https://pypi.org/project/leidenalg/
Scran v1.26.2	Lun et al.[53]	https://bioconductor.org/packages/release/bioc/html/scran.html
Seurat v4.0.1	Butler et al.[54]	https://cloud.r-project.org/web/packages/Seurat/index.html
STAR v2.7.6	Dobin et al.[55]	https://github.com/alexdobin/STAR

REAGENT or RESOURCE	SOURCE	IDENTIFIER
TEtranscripts v2.2.3	Jin et al.[56]	https://github.com/mhammell-laboratory/TEtranscripts
WGCNA v1.72-1	Langfelder and Horvarth[22]	https://cran.r-project.org/web/packages/WGCNA/index.html