Description of Additional Supplementary Files

File name: Supplementary Data 1

Description: Differentially expressed genes (DEGs) in ALS cases (n=422) versus controls (n=272). AvExpr, average $\log_2$ gene expression; bp, base pair; FC, fold-change; P-value, moderated t-test p-values unadjusted for multiple comparisons; FDR, Benjamini-Hochberg-adjusted p-value for false discovery rate.

File name: Supplementary Data 2

Description: Overlap of DEGs in ALS cases (n=422) versus controls (n=272) with literature. Sheet 1: Summary of literature studies. **Sheet 2**: DEG_consistent_up, upregulated ALS DEGs shared with Swindell et al.[1] and Grima et al.[2] **Sheet 3**: DEG_consistent_down, downregulated ALS DEGs shared with Swindell and Grima. **Sheet 4**: DEG_unqiue_up, upregulated ALS DEGs unique from Swindell and Grima. **Sheet 5**: DEG_unique_down, downregulated ALS DEGs unique from Swindell and Grima. **Sheet 6**: Our_study_DEG_up, upregulated ALS DEGs from this study. **Sheet 7**: Our_study_DEG_down, downregulated ALS DEGs from this study. **Sheet 8**: 2019_DEG_up, upregulated ALS DEGs from Swindell. **Sheet 9**: 2019_DEG_down, downregulated ALS DEGs from Swindell. **Sheet 10**: 2023_DEG_up, upregulated ALS DEGs from Grima. **Sheet 11**: 2023_DEG_down, downregulated ALS DEGs from Grima. AvExpr, average $\log_2$ gene expression; CTL, control; DEGs, differentially expressed genes; FC, fold-change; P-values, unadjusted for multiple comparisons; FDR, Benjamini-Hochberg-adjusted p-value for false discovery rate; LFC, log fold-change.

File name: Supplementary Data 3

Description: Clinical characteristics of ALS participants in the primary study cohort for survival prediction. Median (25%,75%); n (%); two-sided Wilcoxon rank sum test; Pearson's Chi-squared test.

File name: Supplementary Data 4

Description: Feature selection by prediction models for ALS case survival (n=420). Sheet 1: Stepwise model. **Sheet 2**: XGBoost model. P-values, unadjusted for multiple comparisons.

File name: Supplementary Data 5

Description: Performance of prediction models for ALS case survival in internal dataset (n=420). Areas under the receiver operating characteristic curves for the clinical variables only, stepwise (18-stepwise selected genes + clinical variables), XGBoost (8-XGBoost selected genes + clinical variables), gene only stepwise (18-stepwise selected genes only), and gene only XGBoost (8-XGBoost selected genes only) models in thirty random train-test splits of our internal dataset at 2, 3, 4, 5, 6, 7, and 8 years. **Sheet 1**: Summary results. **Sheet 2**: Detailed results.

File name: Supplementary Data 6

Description: Pathways enrichment analysis of DEGs in ALS cases (n=422) versus controls (n=272) by Kyoto Encyclopedia of Genes and Genomes (KEGG). Sheet 1: KEGG_both_primary_adjusted, KEGG pathways following enrichment of primary DEGs and DEGs adjusted for cell type proportions. **Sheet 2**: KEGG_only_primary, KEGG pathways following enrichment only of primary DEGs. **Sheet 3**: KEGG_only_adjusted, KEGG pathways following enrichment only of DEGs adjusted for cell type proportions. P-values, unadjusted for multiple comparisons; FDR, Benjamini-Hochberg-adjusted p-value for false discovery rate; log10p, $-\log_{10}$(p-value).

File name: Supplementary Data 7

Description: Pathways enrichment analysis of DEGs in ALS cases (n=422) versus controls (n=272) by Gene Ontology (GO). Sheet 1: Summary of excel sheets. **Sheet 2**: BP_Common, BP terms commonly enriched from primary DEGs and DEGs adjusted for cell type proportions. **Sheet 3**: BP_Disappear, BP terms enriched only from primary DEGs. **Sheet 4**: BP_Show, BP terms enriched only from DEGs adjusted for cell type proportions. **Sheet 5**: Primary_BPMFCC, all GO terms (including BP, MF, CC) enriched from primary DEGs. **Sheet 6**: Adjusted_BPMFCC, all GO terms (including BP, MF, CC) enriched from DEGs adjusted for cell type proportions. **Sheet 7**: BPMFCC_Common, GO terms commonly enriched from primary DEGs and DEGs adjusted for cell type proportions. **Sheet 8**: BPMFCC_Disappear, GO terms enriched only from primary DEGs. **Sheet 9**: BPMFCC_Show, GO terms enriched only from DEGs adjusted for cell type proportions. BP, biological process; CC, cellular component; FDR, Benjamini-Hochberg-adjusted p-value for false discovery rate; MF, molecular function.

File name: Supplementary Data 8

Description: Overlapping and unique candidates selected by drug perturbation analysis across this study, iPSC-neuron, and postmortem spinal cord datasets. Subset, overlap among groups; detail, drug name; CommonAdj, “core genes” from DEGs adjusted for cell type composition; CommonNoAdj, “core genes” from primary DEGs; PM, postmortem spinal cord dataset; TDP, iPSC-neuron dataset; cell_iname, cell culture used in LINCS perturbation experiment; pert_idose, concentration of perturbagen used in LINCS perturbation experiment; pert_itime, time culture was treated with the perturbagen used in LINCS perturbation experiment; MOA, perturbagen mechanism of action; target_name, perturbagen target in MOA; raw_cs, raw connectivity score; fdr_q_nlog10, $-\log_{10}$(false discovery rate q-value); id, LINCS database perturbation ID number; norm_cs, normalized connectivity score.

References

1. Swindell WR, Kruse CPS, List EO, Berryman DE, Kopchick JJ. ALS blood expression profiling identifies new biomarkers, patient subgroups, and evidence for neutrophilia and hypoxia. *J Transl Med*. May 22 2019;17(1):170. doi:10.1186/s12967-019-1909-0
2. Grima N, Liu S, Southwood D, et al. RNA sequencing of peripheral blood in amyotrophic lateral sclerosis reveals distinct molecular subtypes: Considerations for biomarker discovery. *Neuropathol Appl Neurobiol*. Dec 2023;49(6):e12943. doi:10.1111/nan.12943