A Cross-Organ Single-Cell Atlas of ALS Reveals Tissue-Specific Remodeling, and Druggable Biomarkers

Jianyu Pang#[1,2], Mingye Wang#[1,2], Jiahao Chen[3,4], Yining Liu[1], Jing Zhang[2,6], Hao Zhang*[7], Lei Gao*[3,4,5], Chong Gao*[3,4], Limin Xu*[1,2]

Author's Note:

(#): First author with equal contribution.

(*) Corresponding Author: Hao Zhang (zhanghao@hospital.westlake.edu.cn); Lei Gao (gao_smu@126.com); Chong Gao (chandergao@hotmail.com); Limin Xu (xulimin@cuhk.edu.cn).

Institution:

1. School of Medicine, The Chinese University of Hong Kong, Shenzhen, China.
2. The Chinese University of Hong Kong, Shenzhen Futian Biomedical Innovation R&D Centre, Shenzhen, China.
3. School of Basic Medical Science, Guangzhou University of Chinese Medicine, Guangzhou, China.
4. Institute of Traditional Chinese Medicine Model Organisms, Guangzhou University of Chinese Medicine, Guangzhou, China.
5. Guangdong Provincial Key Laboratory of Chinese Medicine Pharmaceutics, Southern Medical University, Guangzhou, China.
6. HaploX Biotechnology, Shenzhen, China.
7. Motor Neuron Disease Diagnosis and Treatment Center, Affiliated Hangzhou First People's Hospital, School of Medicine, Westlake University, Hangzhou, China.

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.04.28.000208. This version posted April 28, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All Rights Reserved. No reuse allowed without permission.

Abstract

Background: Amyotrophic lateral sclerosis (ALS) is a highly heterogeneous neurodegenerative disorder. Its pathological process involves not only motor neuron degeneration but also glial abnormalities, disrupted intercellular communication, and remodeling of peripheral tissues such as skeletal muscle. However, the cross-tissue pathological features of ALS across the brain, spinal cord, and skeletal muscle, as well as their molecular basis, remain insufficiently characterized.

Methods: In this study, we integrated single-nucleus RNA sequencing (snRNA-seq) data from the brain, spinal cord, and skeletal muscle of SOD1, TDP43, and wild-type mice to construct a cross-tissue cellular atlas. Based on stringent quality control, we performed cell type annotation, proportion analysis, irGSEA functional scoring, CellChat-based intercellular communication inference, pseudotime trajectory analysis, pySCENIC-based transcriptional regulatory network reconstruction, and multi-model machine learning screening. Subsequently, DeepPurpose and AutoDock were used to perform drug-target prediction and molecular docking analyses for candidate biomarkers.

Results: A total of 109,875 high-quality cells were obtained, and 18 cell types were identified. The results showed that different ALS genotypes induced pronounced and tissue-specific remodeling of cellular composition and functional states in the brain, spinal cord, and skeletal muscle. Functional scoring revealed marked tissue-dependent differences in neuroinflammation, oxidative metabolism, proteostasis imbalance, and aberrant cell proliferation. Cell communication analysis indicated that NRXN, NEGR, NRG, and PTPRM signaling pathways play central roles in different tissues, suggesting remodeling of synaptic maintenance and cell adhesion networks in ALS. Pseudotime analysis further demonstrated distinct temporal reprogramming patterns in neuronal and stromal cell differentiation trajectories between the SOD1 and TDP43 models.

After integrating multidimensional features, machine learning identified four high-confidence biomarkers: *Tac2*, *Gsn*, *Taco1*, and *Sod1*. Drug screening and molecular docking further revealed potential high-affinity interactions between sulforaphane, phenylhydrazine, and several candidate compounds with key targets.

Conclusion: This study systematically reveals the pathological heterogeneity of ALS across tissues, cell types, and molecular layers, elucidates its core intercellular communication networks and key candidate biomarkers, and provides a potential molecular foundation and candidate drug leads for targeted intervention and precision therapy in ALS.

Keywords: Amyotrophic lateral sclerosis; Single-nucleus RNA sequencing; Cross-tissue atlas; Biomarkers; Potential therapeutic compounds.

Introduction

Amyotrophic lateral sclerosis (ALS) is a progressive neurodegenerative disorder characterized by motor neuron degeneration, ultimately leading to respiratory failure [1]. Although the antisense oligonucleotide therapy Tofersen, which targets SOD1 mutations, has demonstrated efficacy in reducing neurofilament light chain levels [2], the pathophysiological network of ALS remains exceedingly complex, involving multilayered cascades such as proteostasis imbalance, aberrant RNA metabolism, and neuroinflammation [1]. In recent years, investigations into ALS pathogenesis have extended to axonal transport defects, mitochondrial dysfunction, and astrocyte-mediated neurotoxicity [3, 4]. Nevertheless, its pathological basis has not yet been fully elucidated, and effective interventions remain limited. The currently recognized core ALS-associated genes include *SOD1*, *TARDBP* (encoding TDP-43), *FUS*, and *C9orf72*. Systematic reviews have delineated the molecular pathways by which these pathogenic mutations drive disease onset, primarily through oxidative stress, RNA metabolic dysregulation, and proteostasis impairment [5]. Moreover, recent studies have shown that loss of TDP-43 function can lead to aberrant splicing of *KCNQ2* mRNA, and the resulting neuronal hyperexcitability may represent an important pathogenic mechanism in ALS [6]. Notably, ALS is not merely a consequence of motor neuron degeneration; rather, it is a multicellular and multiorgan pathological process involving neurons, microglia, astrocytes, peripheral monocytes/macrophages, and T lymphocytes, with neuroinflammatory responses spanning both the central and peripheral nervous systems [7, 8]. In particular, the role of cross-tissue interactions among the brain, spinal cord, and skeletal muscle in disease progression and phenotypic heterogeneity remains insufficiently investigated.

In recent years, the advent of single-nucleus RNA sequencing (snRNA-seq) has provided new opportunities for resolving the heterogeneity of complex tissues. Using this technology, an ALS motor cortex atlas has been established, clearly revealing disease-associated cell-state transitions, remodeling of cell proportions, and aberrant ligand-receptor (LR) interaction networks [9]. Meanwhile, by integrating cell communication inference, pseudotime analysis, and machine learning-based strategies, researchers have identified key nodes such as a glial subpopulation characterized by RIPK1 inflammatory signaling in the spinal cord [10], providing a powerful computational framework for identifying disease-relevant cell populations, core signaling pathways, and high-confidence biomarkers. However, most existing ALS studies have focused on a single tissue or a single cell population, and systematic cross-scale integrative analyses spanning the central nervous system and peripheral muscle tissue are still lacking. Therefore, constructing an ALS cellular atlas from a cross-tissue perspective and further elucidating its molecular regulatory networks and potentially actionable targets are of important theoretical and clinical significance.

Based on this rationale, the present study integrated snRNA-seq data from the brain, spinal cord, and skeletal muscle of classical ALS SOD1 and TDP43 mouse models, together with wild-type controls, to construct a cross-tissue cellular atlas. We systematically analyzed genotype-dependent changes in cell composition, functional-state differences, cell communication networks, and differentiation trajectory features. On this basis, by combining differential expression analysis, cell communication analysis, and transcriptional regulatory network results, we further applied machine learning to identify high-confidence ALS-related biomarkers and explored their potential therapeutic value through drug-target prediction and molecular docking. This study aims to elucidate the systemic pathological remodeling of ALS from multiorgan, multilayered, and multidimensional perspectives, thereby providing new theoretical foundations for understanding its clinical heterogeneity and developing precision therapeutic strategies.

Methods

1. Single-cell data acquisition and analysis

In this study, two classical ALS transgenic mouse models were used: TDP43 (B6. Cg-Tg (Prnp-TARDBPA315T) 95 Balo / J) and SOD1 (B6SJL-Tg (SOD1-G93A) 1 Gur / J), with age-matched wild-type littermates serving as controls. All animals were housed under specific pathogen-free (SPF) conditions, and all experimental procedures were conducted in accordance with standard animal ethical guidelines.

During sample processing, brain, spinal cord, and skeletal muscle tissues were rapidly dissected and preserved at low temperature. The tissues were mechanically minced and enzymatically digested under appropriate conditions to obtain single-cell suspensions, which were subsequently filtered through a density gradient to remove tissue debris. After red blood cell lysis and dead-cell removal, the cells were purified and adjusted to the appropriate concentration for downstream library preparation. Cell viability, assessed by trypan blue staining, was consistently above 85%.

Single-cell libraries were prepared using the 10× Genomics Chromium platform (Single Cell 3′ v3 kit), with approximately 5,000 cells captured per sample. Libraries were amplified and constructed according to the standard protocol and sequenced on an Illumina NovaSeq 6000 platform using paired-end sequencing (PE150), with a sequencing depth of no less than 20,000 reads per cell.

Raw sequencing data were aligned and quantified using Cell Ranger software [11] to generate gene expression matrices. Based on these quantification results, Seurat (v4.4.0) [12] was used for subsequent data analysis and integration. During quality control, genes expressed in fewer than 10 cells and low-quality cells with fewer than 500 detected genes were removed. In addition, only cells with more than 200 detected genes and a mitochondrial gene proportion below 10% were retained for downstream analyses.

2. Pathway activity assessment

All gene sets used in the analysis were obtained from the MSigDB mouse database [13] via the R package msigdbr, encompassing Hallmark gene sets (inflammatory response, oxidative phosphorylation, apoptosis, and reactive oxygen species) as well as Gene Ontology (GO) biological process categories (oxidative stress, autophagy, endoplasmic reticulum stress, neuronal death, cell proliferation, and microglial activation). These genes were further consolidated into four major functional modules: Neuroinflammation, Oxidative_Mitochondria, CellDeath_Proteostasis, and CellCycle_Proliferation for scoring. Detailed information on the gene sets is provided in Supplementary Table 1.

To minimize potential systematic bias introduced by any single algorithm, functional pathway activity at the single-cell level was evaluated based on the snRNA-seq expression matrix using the irGSEA framework [14] (v3.3.4). irGSEA integrates multiple commonly used gene set enrichment analysis methods; in this study, seven algorithms were applied independently, namely AUCell, UCell, singscore, ssGSEA, AddModuleScore, JASMINE, and viper. The normalized data matrix from the Seurat object was used as input, and each algorithm was executed through the `irGSEA.score` function with the following parameters: `custom=TRUE`, `minGSSize=1`, `maxGSSize=2000`, and `seed=123`. The aforementioned four functional modules were used as the gene sets for scoring. To eliminate scale differences among the outputs of different algorithms, the single-cell scoring matrix from each algorithm was separately subjected to Min-Max normalization, linearly rescaling scores to the 0-1 range. Finally, the normalized scores from the seven algorithms were averaged to obtain a composite activity score for each cell in each functional module, and the resulting scores were added to the metadata of the Seurat object for subsequent statistical analyses.

3. Cell-cell communication analysis

Cell-cell communication networks in the snRNA-seq data were systematically inferred using CellChat [15] (v2.2.0). Based on a LR interaction database and a mass-action-law model, CellChat takes the normalized gene expression matrix and cell type annotations as input to estimate communication probabilities between cell populations.

During the analysis, the mouse LR database (CellChatDB.mouse) was loaded according to species origin. For pathway-specific analyses, the database was further subsetted based on the functional classification field. The trimean method was used for communication probability estimation to improve robustness against outliers. Communication intensities were then summarized at the pathway level to generate intercellular interaction networks and enable visualization. Notably, before running CellChat, snRNA-seq data from each group were filtered to exclude cell types containing fewer than 100 cells, thereby reducing noise and potential artifacts.

4. Pseudotime trajectory inference and pathway enrichment

Monocle [16] (v2.34.0) was used to infer pseudotemporal differentiation trajectories for each cell subpopulation in the snRNA-seq dataset. Following the standard workflow, data preprocessing and dimensionality reduction were first performed. During trajectory construction, differentially expressed genes (DEGs) among experimental groups within each subpopulation were used as ordering genes. For example, in the trajectory analysis of the neuronal subpopulation in brain tissue, DEGs identified among the SOD1, TDP43, and WT groups were selected. Differential expression analysis was performed using the `FindAllMarkers` function in the Seurat package with the Wilcoxon rank-sum test. The selection criteria were adjusted P value (BH) < 0.05, $|\log FC| \geq 0.25$, and min.pct ≥ 0.1. Trajectory embedding and visualization were ultimately performed using the DDRTree algorithm.

In addition, clusterProfiler [17] (v4.16.6) was used to conduct GO enrichment analysis for DEGs from each group. Significantly enriched terms with adjusted P value (Benjamini-Hochberg, BH) < 0.05 were retained and ranked in descending order by GeneRatio. The results were visualized using DotPlot to display the biological process (BP), molecular function (MF), and cellular component (CC) categories.

5. Multidimensional candidate gene screening strategy

To systematically identify transcriptionally active ALS-associated genes, we established a three-dimensional integrative analytical framework. In the first dimension, differential expression analysis was performed across all cell types at the tissue level (brain, spinal cord, and skeletal muscle) using the `FindAllMarkers` function (Wilcoxon test with BH correction), and DEGs with adjusted P value < 0.05 were retained. In the second dimension, based on CellChat-inferred intercellular communication results, significantly interacting LR pairs were extracted, and signaling pathways with adjusted P value < 0.05 were identified using the BH method, with emphasis placed on biologically meaningful communication axes. In the third dimension, pySCENIC [18] (v0.12.0) was applied to reconstruct transcription factor regulatory networks. This workflow included three steps: GRNBoost2-based co-expression network inference, cis-regulatory motif enrichment validation, and AUCell-based regulon activity scoring. By integrating the expression matrix with species-specific motif databases, key transcription factors and their downstream target-gene regulatory networks were systematically identified.

Finally, the results from the three dimensions were integrated and deduplicated, and the source information of each gene across different dimensions was recorded for subsequent functional analysis and prioritization.

6. Machine learning modeling and SHAP-based interpretation for biomarker discovery

To improve the robustness of candidate biomarker selection and reduce bias introduced by any single algorithm, we incorporated SHapley Additive exPlanations (SHAP), a game-theory-based model interpretation method, to quantify the contribution of each feature to model predictions. On this basis, three classification models were constructed using scikit-learn (v1.7.1), xgboost (v3.0.3), and lightgbm (v4.6.0), namely random forest classifier (RFC), XGBoost (XGB), and LightGBM (LGB). For each model, SHAP values were calculated and the top 10 most important features were extracted. High-confidence biomarkers shared across models were then identified using an intersection-based strategy.

Given that snRNA-seq data are characterized by high dimensionality, sparsity, and technical noise, a pseudo-bulk strategy was applied for denoising prior to model training. Specifically, for the candidate genes obtained from the three-dimensional screening, average expression values were calculated according to each `experimental group × cell type` combination, yielding a total of 123 pseudo-bulk samples. This approach preserved biological differences while effectively reducing noise. The samples were then stratified and split into a training set (N = 86) and a test set (N = 37) at a 7:3 ratio. Using the experimental group (N = 9) as the classification label, each model was tuned by 5-fold cross-validation combined with Bayesian optimization (N = 30 iterations), with the area under the receiver operating characteristic curve (AUC) used as the primary optimization metric to address class imbalance.

After model optimization, the multiclass models were systematically evaluated on an independent test set using accuracy (ACC), AUC, F1 score, precision-recall (PR) curve, recall (sensitivity), and precision (positive predictive value). In addition, class-specific AUC values were calculated for the nine experimental groups to comprehensively assess model discrimination performance under different biological conditions.

7. Protein and small-molecule structure acquisition

The protein sequences encoded by the ALS biomarker genes in this study were obtained from the UniProt database [19]. To ensure that the drug-target interaction (DTI) analysis retained cross-species relevance, both human and mouse protein sequences were downloaded. In addition, experimentally resolved X-ray crystal structures of proteins were retrieved from the RCSB PDB database [20], whereas predicted structures were obtained from the AlphaFold database [21].

For small-molecule compounds, we first retrieved literature-supported compounds associated with ALS biomarker genes from the CTD database [22] and obtained their SMILES and SDF structures from PubChem. Furthermore, to more comprehensively explore potential DTIs involving ALS biomarkers, we also incorporated a previously reported compound collection [23] (N = 804,959).

All small-molecule structures were ultimately converted into MOL2 format for subsequent molecular docking analyses. Specifically, compounds in SMILES format were first converted to SDF files using RDKit (v2023.2), followed by hydrogen addition, three-dimensional conformer generation using the ETKDG algorithm, and energy minimization with the MMFF94 force field. The resulting SDF files were then converted to MOL2 format using Open Babel [24]. For three-dimensional SDF structures retrieved from PubChem, Open Babel was used directly for format conversion, with preprocessing steps consistent with the RDKit workflow, including hydrogen addition and energy minimization.

8. Integrative deep learning and molecular docking for virtual screening

In this study, DeepPurpose [25] pre-trained models were used to perform multi-feature encoding and DTI prediction for compound-protein pairs. Specifically, six encoding combinations were applied: CNN-CNN, Daylight-AAC, Morgan-AAC, Morgan-CNN, MPNN-CNN, and Transformer-CNN. In the DeepPurpose framework, compound SMILES strings and protein sequences were used as inputs, and the resulting binding scores were used to quantify DTI strength. To reduce the potential bias introduced by any single encoder, the prediction results from the six encoding strategies were aggregated using a combined mean- and maximum-based approach. For each ALS biomarker gene, the top 10 compounds were selected for subsequent molecular docking analyses.

Molecular docking was performed using AutoDock Vina [26] (v4.2). First, the official script `prepare_ligand4.py` was used to process compound MOL2 structures and generate PDBQT files. Next, `prepare_receptor4.py` was applied to preprocess protein PDB structures, including the addition of missing hydrogen atoms and removal of water molecules, followed by generation of PDBQT files. Subsequently, `prepare_gpf4.py` was used to integrate ligand and receptor information into grid parameter files (GPF) and automatically calculate the grid box center coordinates. Then, `prepare_dpf4.py` was run with default parameters to generate docking parameter files (DPF), including 10 independent Lamarckian genetic algorithm search runs, selection of the best global conformation as the final output, and an RMSD clustering tolerance of 2.0 Å. Docking tasks were ultimately executed in parallel to enable large-scale virtual screening.

For result integration, an algorithmic framework developed by our team was used to combine outputs from DeepPurpose and AutoDock Vina. First, failed docking results were removed, and only candidate interactions with docking energies < -5 kcal/mol were retained. When multiple protein structures were available (e.g., X-ray and AlphaFold-predicted structures), the mean docking energy was calculated and then converted to its absolute value. A composite scoring system was subsequently constructed as follows:

Weight score = 0.5 × docking energy (X-ray structure) + 0.3 × docking energy (AlphaFold-predicted structure) + 0.2 × DeepPurpose binding score + 0.1 × log(number of protein PDB structures)

This scoring scheme accounts for differences in scale among algorithmic outputs and assigns different weights to experimentally resolved and predicted structures, thereby enabling an integrated evaluation of DTI results. In addition, the open-source version of PyMOL (v2.6.0a0) was used to visualize the AutoDock docking results.

9. Drug-likeness and safety assessment of small molecules

The physicochemical properties related to absorption, distribution, metabolism, and excretion (ADME) of small-molecule compounds were predicted using the SwissADME database [27] to evaluate their drug-likeness and safety profiles. After the compound SMILES strings were submitted as input, SwissADME returned key parameters including molecular weight, bioavailability, and pharmacokinetic properties. Compounds were then filtered according to the following criteria: molecular weight ≤ 500 Da, hydrogen bond donors ≤ 5, hydrogen bond acceptors ≤ 10, consensus Log P ≤ 5, topological polar surface area (TPSA) ≤ 140 Å², number of rotatable bonds ≤ 10, and ESOL Log S ≥ -6. In addition, compounds were required to satisfy both the Veber and Egan rules with zero violations, have a bioavailability score ≥ 0.55, show predicted gastrointestinal absorption as “High,” and not serve as P-glycoprotein substrates. Furthermore, they were required to show no inhibitory activity against CYP1A2, CYP2C19, CYP2C9, CYP2D6, or CYP3A4, and to have no more than one PAINS alert.

10. Statistical analysis

All statistical analyses were performed using R (v4.5.2). Comparisons between two groups were conducted using the Wilcoxon rank-sum test. For analyses involving multiple P values, such as DEG screening and pathway enrichment analysis, the BH method was applied for multiple-testing correction to reduce the false-positive rate**.**

Results

1. Construction of a cross-tissue cellular atlas of ALS disease models and overall features

SnRNA-seq data from the brain, spinal cord, and skeletal muscle of TDP43, SOD1, and WT mice were integrated for analysis. Considering that cross-tissue integration may introduce additional noise, no excessive batch correction was applied after stringent quality control. A total of 109,875 high-quality cells were retained (Fig. 1a). UMAP results showed that cells clustered primarily by cell type rather than by sample origin, indicating that this strategy effectively preserved biologically meaningful cross-tissue features. Using canonical marker genes for annotation, 18 cell types were identified (Fig. 1b-c).

Cell proportion analysis revealed significant tissue-dependent differences (Fig. 1d). In the central nervous system, astrocytes and microglia showed distinct trends in the brain and spinal cord. In the brain, astrocytes followed the pattern SOD1 < TDP43 < WT, whereas microglia followed TDP43 < WT < SOD1. In the spinal cord, astrocytes followed TDP43 < SOD1 < WT, while microglia followed SOD1 < WT < TDP43, suggesting that glial remodeling in ALS models is tissue-specific. Oligodendrocytes and Oligodendrocyte precursor cells (OPCs) were mainly enriched in the spinal cord and were relatively more abundant in the TDP43 model. Among neuronal subpopulations, excitatory neurons were primarily distributed in the brain (SOD1 < WT < TDP43), whereas inhibitory neurons were more enriched in the spinal cord (WT < TDP43 < SOD1), with model-dependent differences in relative abundance. In skeletal muscle, the proportions of immune cells (macrophages/monocytes) and muscle satellite cells were generally lower in ALS models than in WT. Muscle fiber types also underwent remodeling: *Runx1+* regeneration-associated fibers increased in the SOD1 model, whereas *Myh2+* (IIa/IIx) fibers were elevated in the TDP43 model; in contrast, WT mice were dominated by *Myh4+* (IIb) mature fibers. Among stromal cells, adipocytes were mainly distributed in the brain and skeletal muscle and showed an overall decreasing trend in the SOD1 model, while fibroblasts in skeletal muscle followed the pattern SOD1 < TDP43 < WT. Collectively, these findings indicate that distinct ALS genotypes drive characteristic cellular remodeling across the brain, spinal cord, and skeletal muscle, underscoring the role of the tissue microenvironment in shaping disease phenotypes.

We further used the integrated algorithm irGSEA to assess the activity scores of four functional modules at the single-cell level: Neuroinflammation, Oxidative Mitochondria, CellDeath Proteostasis, and CellCycle Proliferation (Fig. 1e-f). The results showed marked tissue- and genotype-dependent differences in module activity.

In brain tissue, the SOD1 model exhibited higher activities of Neuroinflammation, CellDeath Proteostasis, and CellCycle Proliferation than the WT and TDP43 groups, with the relative strength following the order Neuroinflammation > CellDeath Proteostasis > CellCycle Proliferation, whereas Oxidative Mitochondria scores were lower than those in WT. In spinal cord tissue, the SOD1 model showed significantly increased Oxidative Mitochondria activity, while Neuroinflammation and CellCycle Proliferation scores were markedly reduced compared with the other groups, presenting a regulatory pattern opposite to that observed in the brain. The TDP43 model also exhibited elevated Oxidative Mitochondria activity in the spinal cord, accompanied by reduced Neuroinflammation and CellCycle Proliferation activity. In skeletal muscle, the SOD1 model showed the highest CellDeath Proteostasis score, suggesting that apoptosis and proteostasis dysregulation are particularly prominent in muscle tissue; the trends of the other three modules were broadly consistent with those in the brain.

The TDP43 model likewise was characterized by increased Oxidative Mitochondria and relatively decreased Neuroinflammation and CellCycle Proliferation activity in skeletal muscle.

Taken together, this multi-algorithm integrative functional scoring approach quantitatively revealed genotype-specific regulatory patterns of core pathological pathways, including neuroinflammation, oxidative metabolism, cell death, and proliferation, in the brain, spinal cord, and skeletal muscle microenvironments of ALS models, providing strong evidence for the molecular basis of clinical heterogeneity in ALS.

2. Comparative analysis of inter-tissue cell-cell communication networks

To characterize differences in intercellular communication across tissues and genotypic backgrounds, we extracted information for nine major cell types according to “tissue × model group” combinations and systematically analyzed the cell-cell communication networks in each group based on the CellChat framework.

Analysis of communication intensity (Fig. 2a) showed that in brain tissue, communication between excitatory neurons and neurons was the most abundant, with an overall decreasing trend in communication intensity of SOD1 ≈ WT > TDP43. Compared with the WT and TDP43 groups, the SOD1 group exhibited a marked reduction in the diversity of intercellular communication types, manifested by the loss of interaction signals between excitatory neurons and astrocytes (TDP43 < WT), as well as communication connections between OPCs and other cell populations (TDP43 < WT). In addition, a unique vascular smooth muscle cell communication node appeared in the TDP43 group, although its intensity was relatively weak.

In spinal cord tissue, communication between inhibitory neurons and neurons was highly active, and astrocytes and OPCs also participated in prominent signaling events. The overall communication intensity followed the pattern SOD1 > WT > TDP43.

Notably, vascular smooth muscle cell-mediated communication was detected in both the TDP43 and WT groups, whereas a schwann cell communication pattern was uniquely observed in the WT group.

In skeletal muscle, endothelial cells and fibroblasts formed the core communication hub, with particularly strong signaling interactions between fibroblasts and endothelial Cells, macrophages/monocytes, and pericytes. In the SOD1 group, a muscle satellite cell communication pattern emerged specifically, and these cells showed strong interactions with endothelial cells, pericytes, and Type I Myofibers (*Runx1*+).

To further clarify the specificity of key communication signals, we intersected significantly enriched LR pairs across groups and then analyzed the union of the top three LR pairs ranked by communication strength within each group based on the intersection set (Fig. 2b). The results indicated that the dominant communication pathways in brain and spinal cord tissues were mainly NRXN, NEGR, and NRG signaling, which are involved in neuronal synapse formation, maturation, and stability; in contrast, skeletal muscle tissue was dominated by PTPRM, LAMININ, NCAM, and APP signaling pathways that regulate intercellular adhesion.

Among these, *Nrxn1-Nlgn1* was the LR pair with the highest communication strength in the brain tissue of the SOD1 and WT groups, as well as in all spinal cord groups (Fig. 2c-d). In the SOD1 brain, this pair mediated communication only between excitatory neurons and neurons; in the WT brain, its communication range was expanded to interactions between neurons / excitatory neurons and astrocytes / OPCs. In the spinal cord, inhibitory neurons replaced excitatory neurons as the major source of this signal. In the TDP43 group, the top-ranked NRXN signal was *Nrxn3-Nlgn1* (Fig. 2c), which not only mediated intercellular communication but also participated in autocrine regulation in OPCs and excitatory neurons. These findings indicate that signaling pathways related to synaptic maintenance and glia-neuron support are markedly perturbed in ALS models.

In the skeletal muscle communication network, endothelial cells primarily established connections with other cell types through the PTPRM pathway via *Ptprm-Ptprm* homophilic interactions (Fig. 2e). In both the SOD1 and TDP43 groups, adipocytes were absent from this signaling axis. The SOD1 group uniquely exhibited communication between endothelial cells and Type I Myofibers (*Runx1*+), whereas the TDP43 group showed a communication pattern between endothelial cells and vascular smooth muscle cells. These observations suggest that ALS muscle tissue may influence degenerative changes in muscle fibers by regulating vascular permeability or intercellular mechanotransduction.

3. Cross-tissue pseudotime trajectory analysis

Based on the identification of highly communicating cell types described above, pseudotime differentiation trajectory inference was further performed to systematically characterize the dynamic differentiation features of key cell subpopulations in the brain, spinal cord, and skeletal muscle. Specifically, excitatory neurons and neurons in the brain, inhibitory neurons and neurons in the spinal cord, and endothelial cells and fibroblasts in skeletal muscle were analyzed separately.

In the brain, trajectory analysis of excitatory neurons (Fig. 3a-b) showed that cells from the SOD1 model were mainly enriched at the terminal stage of differentiation (Fig. 3b), accompanied by significant enhancement of signals related to synaptic structure assembly and ion channel activity (Fig. 3c), including dendritic spine development, postsynaptic structure organization, and voltage-gated cation channel activity. These findings suggest that neurons at the terminal differentiation stage undergo critical remodeling associated with synaptic structural maturation and electrophysiological function. In contrast, the TDP43 model was mainly distributed in the mid-differentiation stage (Fig. 3b), where enrichment of excitatory synaptic components and glutamatergic signaling pathways was observed (Fig. 3d), including ionotropic glutamate receptor complexes, ligand-gated ion channel activity, and postsynaptic density membrane structures, indicating that these cells were at an early stage of excitatory synapse assembly and functional establishment. Further analysis of the overall neuronal population in the brain (Fig. S1a-b) also revealed marked temporal differences between the two models: SOD1 cells were mainly concentrated in the early and late differentiation stages, with a relative reduction in the middle stage, whereas TDP43 cells peaked at the terminal stage. Functional enrichment results showed that the SOD1 model (Fig. S1c) was mainly associated with RNA metabolic regulation, cytoskeletal binding, and basement membrane-related processes, whereas the TDP43 model (Fig. S1d) was enriched in synapse assembly, axon guidance, and transmembrane ion channel activity. Collectively, these results indicate that in the brain, SOD1 and TDP43 pathogenic genotypes drive temporal reprogramming of neuronal differentiation trajectories through differential regulation of RNA processing-cytoskeletal remodeling and synaptic structure-electrophysiological assembly processes.

Similar temporal differences were also observed in the spinal cord. Trajectory analysis of inhibitory neurons (Fig. 3e-f) showed that the SOD1 model was mainly enriched in the early-to-mid differentiation stages (Fig. 3f), accompanied by enhanced signaling related to synaptic transmission and axonal ensheathment (Fig. 3g), including glutamate receptor pathways, action potential regulation, and myelin sheath components, suggesting that these cells were undergoing synaptic connection establishment and initial axonal wrapping. By contrast, the TDP43 model was mainly distributed in the mid-to-late stages and peaked at the terminal stage (Fig. 3f), corresponding to pronounced enrichment of oligodendrocyte-mediated myelination and endoplasmic reticulum protein processing functions (Fig. 3h), indicating sustained activity in axonal myelination and proteostasis regulation. Further analysis of the overall neuronal population in the spinal cord (Fig. S1e-f) revealed opposite distribution patterns between the two models: SOD1 cells were mainly concentrated at the terminal stage, whereas TDP43 cells were concentrated at the early stage. Functional enrichment results showed that the SOD1 model was mainly associated with microtubule organization and vesicle transport (Fig. S1g), whereas the TDP43 model was enriched in endoplasmic reticulum stress response and myelin-related structures (Fig. S1h). Together, these findings suggest that SOD1 mutations tend to activate cytoskeleton-dependent synaptic transport and structural maintenance processes at later stages of differentiation, whereas TDP43 mutations induce endoplasmic reticulum stress and myelin assembly-related signaling at earlier stages, reflecting distinct temporal and functional differentiation patterns in spinal cord neurons.

In skeletal muscle, we further analyzed the differentiation trajectories of highly communicative endothelial cells (Fig. 3i-l) and fibroblasts (Fig. S1i-l). The results showed that both cell types in the SOD1 model were mainly enriched in the mid-to-late stages of differentiation, whereas cells from the TDP43 model were distributed more evenly across all stages. Functional enrichment analysis indicated that both endothelial cells and fibroblasts in the SOD1 model exhibited signals related to sarcomere assembly, actin cytoskeleton binding, and regulation of muscle contraction (Fig. 3k; Fig. S1k), suggesting that SOD1 may drive non-muscle cells toward a muscle-like phenotype. In contrast, endothelial cells in the TDP43 model were mainly enriched in pathways related to regulation of angiogenesis and immune pathways involved in antigen processing and presentation (Fig. 3l), while fibroblasts were more strongly associated with extracellular matrix organization and growth factor binding (Fig. S1l). These results suggest that SOD1 and TDP43 mutations reshape the muscle microenvironment through distinct stromal cell response mechanisms, promoting muscle-like remodeling or immune-vascular regulation, respectively, thereby contributing to divergent pathological features in peripheral ALS tissues.

4. Multidimensional integrative machine learning identifies high-confidence ALS biomarkers

To systematically screen key ALS biomarkers, we constructed a multidimensional feature integration and machine learning framework (Fig. 4a). First, candidate gene sets were generated from three layers. At the tissue × cell type level, differential expression analysis identified 5,395 DEGs (adjusted $P < 0.05$). CellChat-based cell communication analysis yielded 208 significantly interacting LR pairs (adjusted $P < 0.05$). pySCENIC-based transcriptional regulatory network analysis identified 11,997 active transcription factors and their target genes. After integrating these three result sets, a total of 12,886 non-redundant candidate genes were obtained for downstream analysis.

For model construction, three classification models, RFC, XGB, and LGB, were built using the denoised pseudo-bulk dataset. The results showed that all three models performed well in the nine-class classification task (Fig. 4b), with a mean ROC-AUC of 98.2% and a mean F1 score of 80.3% on the test set. Among them, the XGB model achieved a PR-AUC of 0.931. Further class-specific evaluation (Fig. 4c-e) revealed that all three models showed strong discriminative ability for the SOD1 group, suggesting that this model possesses particularly distinctive molecular features.

Using SHAP-based interpretability analysis, we extracted the top 10 features from each model and took their intersection, ultimately identifying four high-confidence ALS-associated biomarkers (Fig. 4f): *Tac2*, *Gsn*, *Taco1*, and *Sod1*. SHAP results showed that these four genes exhibited highly consistent feature-importance rankings across different models and cell-type contexts (Fig. 4g), with *Tac2* contributing the most (mean SHAP = 0.468), followed by *Gsn* (0.371), *Taco1* (0.216), and *Sod1* (0.197). Notably, *Sod1*, a canonical disease-causing gene, was successfully recovered by the model, further supporting the biological validity of this strategy.

Further validation using snRNA-seq expression data (Fig. 4h) showed that *Tac2* was mainly highly expressed in inhibitory neurons of the spinal cord in the TDP43 model; *Gsn* was significantly upregulated in skeletal muscle fibroblasts of the TDP43 model; *Taco1* showed marked abnormalities in the SOD1 model, especially with high expression in multiple skeletal muscle fiber subtypes such as Type I myofibers and *Runx1*+ fibers; and *Sod1* was significantly elevated in spinal cord oligodendrocytes and inhibitory neurons in the SOD1 model. These results further support the robustness and biological relevance of the selected biomarkers at the level of multiple tissues and cell types.

5. Targeted drug screening by integrating deep learning and molecular docking

To further explore the potential drug-development value of key ALS biomarkers, we applied two mainstream DTI prediction methods, DeepPurpose and AutoDock, and combined them with an integrated weighted algorithm to comprehensively evaluate candidate compounds and identify the most promising targeting ligands. To enhance cross-species robustness, we conducted virtual screening using both mouse and human proteins in parallel, and subsequently retained only the intersection of DTI results from the two species for downstream analysis. In addition, ADME and other pharmacokinetic properties of the compounds were incorporated to comprehensively assess their safety and drug-likeness. Beyond large-scale virtual screening based on reported DTIs, we also systematically mined more than 800,000 small-molecule compounds to identify potential targeted drugs (Fig. 5a).

A total of 1,119 known DTI pairs involving the four ALS biomarkers recorded in the CTD database were included, among which SOD1 had 876 pairs, GSN had 174 pairs, TAC2 had 33 pairs, and TACO1 had 36 pairs. Notably, after stringent filtering, only two compounds showed potential drug-likeness (Table 1, Supplementary Table 2): sulforaphane (mean Weight Score = 7.71) and phenylhydrazine (mean Weight Score = 5.36). Both compounds targeted GSN. Sulforaphane has been reported to increase GSN gene methylation and thereby downregulate its mRNA expression [28]. Phenylhydrazine, as a hemolytic agent, can induce oxidative damage and membrane disruption in red blood cells, and GSN may participate in regulating the host sensitivity to phenylhydrazine toxicity [29]. In addition, pharmacokinetic analysis showed that phenylhydrazine had a significantly lower TPSA than sulforaphane, suggesting better intestinal absorption and possibly greater drug-development potential.

We further analyzed the optimal binding conformations of these two DTI pairs to resolve their interaction sites at the molecular level (Fig. 5b-c). The best docking binding energy for sulforaphane-GSN was -6.09 kcal/mol (Fig. 5b), and the interaction was mainly mediated by residues ASN-115, ASN-111, and ILE-76 in GSN. The best docking binding energy for phenylhydrazine-GSN was -6.02 kcal/mol (Fig. 5c), with the interaction primarily mediated by ASP-12.

In addition, screening of literature-reported compounds (N = 804,959) yielded six potential DTI pairs (Table 2, Supplementary Table 3). Among them, SOD1 and TAC2 each identified two high-affinity candidate compounds. Both human and mouse SOD1 proteins showed favorable interaction potential with C15H11FN4O2 (mean Weight Score = 9.23) and C15H26O (mean Weight Score = 9.02), whereas TAC2 yielded corresponding DTI results only for the human protein. Meanwhile, GSN showed the highest weighted interaction score with C9H12N2OS (mean Weight Score = 9.7), while TACO1 showed a relatively lower interaction score with C9H13NO2 (mean Weight Score = 8.2).

Pharmacokinetic results indicated that the interaction between C15H26O and SOD1 had favorable TPSA and a relatively high consensus Log P, suggesting good lipophilicity and membrane permeability and thus potentially high drug-development value. This compound interacted with SOD1 via the VAL-148 residue, with a docking binding energy of -6.64 kcal/mol (Fig. 5e). In addition, C15H11FN4O2 showed the best docking binding energy with SOD1, reaching -7.59 kcal/mol (Fig. 5f), and interacted mainly through ILE-17, ASN-19, and SER-34 in SOD1.

Overall, structural interpretation of DTI binding sites based on molecular docking results further elucidated the interaction patterns between key ALS biomarkers and compounds with potential drug-likeness (Fig. 5b-i), providing an important basis for subsequent targeted drug design and optimization.

Table 1. DTI results based on the CTD database

Chemical Name	**Target Gene**	**Interaction**	**MW (Da)**	**TPSA (Å^2)**	**Weight Score Mus**	**Weight Score Homo**
sulforaphane	*Gsn*	Sulforaphane results in increased methylation of the GSN gene and decreased expression of GSN mRNA.	177.29	80.73	7.37	8.05
phenylhydrazine	*Gsn*	GSN protein affects the susceptibility to phenylhydrazine	108.14	38.05	5.48	5.24

Table 2. DTI results based on literature-derived small-molecule compounds

Compound	**Target Gene**	**MW (Da)**	**Consensus.Log.P**	**TPSA (Å^2)**	**Weight Score Mus**	**Weight Score Homo**
$C_9H_{13}NO_2$	*Taco1*	167.21	0.6	66.48	8.96	7.45
$C_9H_{12}N_2OS$	*Gsn*	196.27	1.49	81.28	8.08	11.31
$C_9H_{11}FN_2O_5$	*Tac2*	246.19	-0.63	104.55	/	8.85
$C_{15}H_{26}O$	*Sod1*	222.37	3.42	20.23	8.63	9.41
$C_{15}H_{20}O_6$	*Tac2*	296.32	1.22	93.06	/	8.57
$C_{15}H_{11}FN_4O_2$	*Sod1*	298.27	1.13	104.42	8.65	9.81

Discussion

By integrating snRNA-seq data from the brain, spinal cord, and skeletal muscle, this study systematically delineated the cellular composition, intercellular communication networks, differentiation trajectories, and molecular regulatory features of ALS across multiple tissue microenvironments. Our findings indicate that ALS pathology is not confined to motor neurons, but rather represents a systemic degenerative process involving coordinated dysregulation of neuron-glia-peripheral tissue interactions [30]. ALS is increasingly recognized as a multifactorial neurodegenerative disorder driven by complex interactions among genetic, environmental, developmental, and resiliencerelated factors. Emerging evidence, including hypothalamic atrophy, further supports the concept that ALS extends well beyond the motor system [31]. Compared with traditional studies focused on a single tissue or cell type, the heterogeneous neuroinflammatory phenotypes exhibited by microglia and astrocytes across different CNS regions underscore the necessity of examining ALS from a cross-tissue, multilayered perspective [32]. In this context, our work provides a more comprehensive characterization of the heterogeneous pathological landscapes associated with the canonical ALS models SOD1 and TDP43.

At the cellular level, we observed pronounced tissue-dependent alterations in the proportions of glial cells, neuronal subtypes, immune cells, and stromal cells, suggesting that ALS-associated cellular remodeling is strongly shaped by local microenvironments. This observation is consistent with the marked clinical heterogeneity of ALS with respect to site of onset, disease progression rate, and pathogenic mechanisms [33], and further implies that distinct genotypes may drive disease progression through tissue-specific pathological routes.

Functionally, irGSEA-based scoring of key modules revealed that neuroinflammation, oxidative metabolic imbalance, proteostasis disruption, and aberrant proliferation exhibited distinct regulatory patterns across tissues. Such genotype-dependent and tissue-specific responses can be interpreted from the perspective of regional glial heterogeneity and differential neuronal vulnerability. In the glial compartment, cortical astrocytes in SOD1-G93A mice display a prominent reactive pro-inflammatory phenotype, characterized by increased *miR-155*/*miR-21*/*miR-146a* expression, whereas spinal astrocytes tend toward a more non-inflammatory transcriptional and proteomic repression state; intriguingly, exosomal *miR-155*/*miR-21*/*miR-146a* depletion is shared by both regions [34]. At the neuronal level, fast-fatigable motor neurons in the rNLS8 TDP-43 model undergo selective degeneration, while slow motor neurons show intrinsic resistance to pathological TDP-43 and can partially compensate through axonal sprouting and reinnervation of denervated neuromuscular junctions [35]. In a broader disease context, aging-associated immune cells, including microglia and T cells, undergo immunosenescence accompanied by mitochondrial dysfunction, redox imbalance, defective autophagy, and impaired nutrient sensing, thereby sustaining a pro-inflammatory state that exacerbates neuroinflammation [36]. The spinal cord, as a central site of motor neuron degeneration, may therefore harbor a pathogenic microenvironment shaped by both oxidative stress and immunosenescence-associated inflammation, contributing to the coexistence of metabolic reprogramming and immune-suppressive-like states during disease progression.

Cell-cell communication analysis further demonstrated that ALS pathology is not merely reflected by changes in cell abundance, but also by remodeling of functional interaction networks. In the brain and spinal cord, signaling pathways centered on NRXN, NEGR, and NRG were strongly enriched, indicating that synapse formation, stability, and neuron-glia support circuits are among the earliest processes affected in ALS. NRGs, as EGF-like extracellular ligands, play key roles in development, maintenance, and repair of the central and peripheral nervous systems through ErbB signaling [37]. In particular, abnormality of the *Nrxn1-Nlgn1* and *Nrxn3-Nlgn1* LR axes has been implicated in ALS-related synaptic dysfunction [38]. Recent studies have further shown that NRXNs regulate inhibitory synapse maintenance and function in a region- and synapse-specific manner, with distinct GABAergic neuronal populations expressing different NRXN isoform combinations and corresponding ligands, including NRGs, thereby forming a complex trans-synaptic molecular code that defines inhibitory synaptic function and circuit specificity [38]. Given the central roles of the NRXN-NLGN axis in synaptic maturation, neural plasticity, and excitatory/inhibitory balance, its disruption may directly contribute to motor circuit dysfunction in ALS.

In skeletal muscle, communication networks dominated by PTPRM, LAMININ, NCAM, and APP suggest that peripheral muscle pathology in ALS is not simply secondary to denervation, but is accompanied by intrinsic defects involving metabolism, proteostasis, regeneration, and RNA metabolism [39]. Notably, endothelial cells and fibroblasts formed a communication hub in muscle, indicating that vascular niche remodeling and fibrotic processes may be important contributors to muscle atrophy and defective regeneration in ALS.

Pseudotime trajectory analysis further supported these conclusions from a dynamic perspective. In the brain, excitatory neurons occupied distinct differentiation stages between SOD1 and TDP43 models, suggesting that these genotypes perturb neuronal maturation within different temporal windows. SOD1 was more associated with late-stage differentiation and synaptic/cytoskeletal remodeling, whereas TDP43 more prominently affected RNA metabolism, early synapse assembly, and endoplasmic reticulum stress. Mechanistically, *SOD1* is primarily linked to oxidative stress and metal homeostasis imbalance, whereas pathological *TDP43* aggregation can sequester specific microRNAs and proteins, disrupt the expression of mitochondrial nuclear-encoded proteins, induce global mitochondrial dysfunction, and promote a self-amplifying “stress-aggregation” cycle [40]. Thus, the distinct pseudotime enrichments observed in our analysis are biologically plausible. Similarly, in the spinal cord, SOD1 preferentially activated later-stage cytoskeleton-dependent synaptic transport and structural maintenance programs, while TDP43 induced early ER stress and myelin assembly-related signaling. In skeletal muscle, SOD1 favored a muscle-like remodeling program in endothelial cells and fibroblasts, whereas TDP43 was more closely linked to immune-vascular pathways, suggesting genotype-specific stromal responses in peripheral tissues.

By integrating differential expression, cell-cell communication, and transcriptional regulatory networks, and then applying machine learning, we identified *Tac2*, *Gsn*, *Taco1*, and *Sod1* as high-confidence biomarkers with translational relevance. *Tac2* and *Gsn* showed marked aberrant expression in specific cell subsets and disease contexts, implying that they may capture genotype- and tissue-specific molecular responses at different disease stages. As expected, *Sod1*, a canonical pathogenic gene, was successfully recovered, supporting the robustness of the analytical pipeline. *Taco1*, a mitochondrial translation factor, may impair mitochondrial protein synthesis by causing mitoribosome stalling at polyproline sequences, thereby affecting COX1 and COX3 translation efficiency and highlighting mitochondrial translational defects as a potential contributor to ALS bioenergetic failure [41]. *Gsn*, one of the most abundant actin-binding proteins involved in cell motility, morphology, and metabolism, showed a strong association with muscle fibroblasts and the TDP43 model, suggesting that it may participate in peripheral pathology through actin filament severing, cytoskeletal remodeling, and interactions with immune cells [42]. This makes *Gsn* a particularly compelling candidate for functional validation.

Finally, DeepPurpose- and AutoDock-based virtual screening identified sulforaphane and phenylhydrazine as potential GSN-targeting compounds, as well as several high-affinity candidates targeting SOD1 and TAC2, offering new leads for ALS-targeted intervention. Although these findings require experimental validation, the computational results suggest that GSN-associated compounds possess favorable binding energies and pharmacokinetic properties, supporting the hypothesis that modulation of actin-binding protein-mediated cytoskeletal and adhesion pathways may represent a promising strategy for ALS intervention in both muscle and neural tissues [42]. Overall, this study establishes a relatively complete systems biology framework for ALS, spanning cross-tissue cellular atlases, communication networks, lineage trajectories, biomarker discovery, and drug prediction. It not only deepens our understanding of ALS tissue heterogeneity and genotype-specific pathology, but also provides testable targets and mechanistic hypotheses for future precision therapies.

Nevertheless, several limitations should be acknowledged. First, the analysis was primarily based on mouse models. Although the SOD1 and TDP43 models capture key pathological features of ALS, they do not fully recapitulate the complexity of sporadic human ALS and clinical heterogeneity; therefore, cross-species generalizability remains to be validated in human samples. Second, to avoid overcorrection and preserve true biological signals, strong batch correction was not applied during cross-tissue snRNA-seq integration. While this helps retain tissue-specific features, some technical bias may remain. Third, cell-cell communication, pseudotime, and transcriptional regulatory analyses are computational inferences and reflect potential associations rather than direct causality. Thus, the roles of key pathways such as NRXN and PTPRM, as well as candidate biomarkers such as *Tac2*, *Gsn*, and *Taco1*, require further experimental confirmation.

Future studies should incorporate human ALS tissues, iPSC-derived neuronal and muscle models, as well as spatial transcriptomics and multi-omics approaches to validate the interaction patterns and regulatory networks proposed here, and to resolve their spatial organization and dynamic evolution. In parallel, gain- and loss-of-function experiments together with pharmacological assessments are needed to define the mechanistic roles and therapeutic potential of the identified biomarkers and candidate compounds. Collectively, this study provides a new analytical framework for the cross-tissue systems-level dissection of ALS and offers actionable targets and theoretical foundations for mechanistic studies and precision intervention.

Conclusion

This study systematically constructed a cellular atlas of the ALS brain, spinal cord, and skeletal muscle based on cross-tissue snRNA-seq data, revealing genotype-driven remodeling of cellular composition, disruption of functional pathways, and abnormalities in cell-cell communication networks. In addition, high-confidence ALS-associated biomarkers, including *Tac2*, *Gsn*, *Taco1*, and *Sod1*, were identified, and several potential target compounds were further proposed, providing new theoretical insights for ALS molecular subtyping, mechanistic investigation, and drug development.

Declarations

Ethics approval and consent to participate: Not applicable.

Consent for publication: Not applicable.

Availability of data and materials: All data used in this study are subject to policy restrictions and are therefore not publicly available. Access to the data can be obtained by contacting the corresponding author. The analysis code is also not publicly available due to policy restrictions, but it may be requested from the corresponding author.

Competing interests: The authors declare that they have no competing interests in this section.

Funding: This work was supported by Shenzhen Medical Academy of Research and Translation (Grant No.: D2401026); Yunnan Provincial Science and Technology Department (Grant No.:202402AA310033); Shenzhen-Hong Kong Cooperation Zone for Technology and Innovation (Grant No.: HZQB - KCZYB - 2020056).

Authors' contributions: Jianyu Pang and Wang Mingye contributed equally to this work. Limin Xu, Chong Gao, Lei Gao, and Hao Zhang conceived and designed the study. Jianyu Pang, Wang Mingye, Jiahao Chen, Yining Liu, and Jing Zhang collected and analyzed the data. Jianyu Pang and Wang Mingye drafted the manuscript. All authors contributed to manuscript revision and approved the final version.

Acknowledgments: Thank all the open-source software and method used in this study.

References

[1] Dey B, Kumar A, Patel AB. Pathomechanistic Networks of Motor System Injury in Amyotrophic Lateral Sclerosis. Curr Neuropharmacol. 2024;22(11):1778-1806.

[2] Miller TM, Cudkowicz ME, Shaw PJ, Genge A, Sobue G, Bucelli RC, et al. Long-Term Tofersen in SOD1 Amyotrophic Lateral Sclerosis. JAMA Neurol. 2026;83(2):115-125.

[3] Raoufinia R, Alyari G, Nia AT, Abbaszadegan MR, Mahmoudi A, Shafaeibajestan S, et al. Cutting-edge treatments in amyotrophic lateral sclerosis: the role of molecular pathogenesis in targeted therapies. Stem Cell Res Ther. 2025;16(1):689.

[4] Chia R, Moaddel R, Kwan JY, Rasheed M, Ruffo P, Landeck N, et al. A plasma proteomics-based candidate biomarker panel predictive of amyotrophic lateral sclerosis. Nat Med. 2025;31(10):3440-3450.

[5] Mizielinska S, Hautbergue GM, Gendron TF, van Blitterswijk M, Hardiman O, Ravits J, et al. Amyotrophic lateral sclerosis caused by hexanucleotide repeat expansions in C9orf72: from genetics to therapeutics. Lancet Neurol. 2025;24(3):261-274.

[6] Joseph BJ, Marshall KA, Harley P, Mann JR, Alessandrini F, Vanoye CG, et al. TDP-43-dependent mis-splicing of KCNQ2 triggers intrinsic neuronal hyperexcitability in ALS/FTD. Nat Neurosci. 2025;28(12):2476-2492.

[7] Majewski S, Klein P, Boillee S, Clarke BE, Patani R. Towards an integrated approach for understanding glia in Amyotrophic Lateral Sclerosis. Glia. 2025;73(3):591-607.

[8] Liu Z, Cheng X, Zhong S, Zhang X, Liu C, Liu F, et al. Peripheral and Central Nervous System Immune Response Crosstalk in Amyotrophic Lateral Sclerosis. Front Neurosci. 2020;14:575.

[9] Ruf WP, Kuhlwein JK, Meier L, Brockmann SJ, LeeBae J, Sadri-Vakili G, et al. Multi-modal dissection of cell-type specific TDP-43 pathology in the motor cortex. Nat Commun. 2026;17(1).

[10] Zelic M, Blazier A, Pontarelli F, LaMorte M, Huang J, Tasdemir-Yilmaz OE, et al. Single-cell transcriptomic and functional studies identify glial state changes and a role for inflammatory RIPK1 signaling in ALS pathogenesis. Immunity. 2025;58(4):961-979 e968.

[11] Zheng GX, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, et al. Massively parallel digital transcriptional profiling of single cells. Nat Commun. 2017;8:14049.

[12] Hao Y, Hao S, Andersen-Nissen E, Mauck WM, 3rd, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. Cell. 2021;184(13):3573-3587 e3529.

[13] Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. 2015;1(6):417-425.

[14] Fan C, Chen F, Chen Y, Huang L, Wang M, Liu Y, et al. irGSEA: the integration of single-cell rank-based gene set enrichment analysis. Brief Bioinform. 2024;25(4).

[15] Jin S, Plikus MV, Nie Q. CellChat for systematic analysis of cell-cell communication from single-cell transcriptomics. Nat Protoc. 2025;20(1):180-219.

[16] Qiu X, Mao Q, Tang Y, Wang L, Chawla R, Pliner HA, et al. Reversed graph embedding resolves complex single-cell trajectories. Nat Methods. 2017;14(10):979-982.

[17] Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012;16(5):284-287.

[18] Van de Sande B, Flerin C, Davie K, De Waegeneer M, Hulselmans G, Aibar S, et al. A scalable SCENIC workflow for single-cell gene regulatory network analysis. Nat Protoc. 2020;15(7):2247-2276.

[19] UniProt C. UniProt: the Universal Protein Knowledgebase in 2023. Nucleic Acids Res. 2023;51(D1):D523-D531.

[20] Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. Nucleic Acids Res. 2000;28(1):235-242.

[21] Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596(7873):583-589.

[22] Davis AP, Grondin CJ, Johnson RJ, Sciaky D, Wiegers J, Wiegers TC, et al. Comparative Toxicogenomics Database (CTD): update 2021. Nucleic Acids Res. 2021;49(D1):D1138-D1143.

[23] Wong F, Omori S, Donghia NM, Zheng EJ, Collins JJ. Discovering small-molecule senolytics with deep neural networks. Nat Aging. 2023;3(6):734-750.

[24] O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR. Open Babel: An open chemical toolbox. J Cheminform. 2011;3:33.

[25] Huang K, Fu T, Glass LM, Zitnik M, Xiao C, Sun J. DeepPurpose: a deep learning library for drug-target interaction prediction. Bioinformatics. 2021;36(22-23):5545-5547.

[26] Eberhardt J, Santos-Martins D, Tillack AF, Forli S. AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. J Chem Inf Model. 2021;61(8):3891-3898.

[27] Daina A, Michielin O, Zoete V. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. Sci Rep. 2017;7:42717.

[28] Dos Santos P, Machado ART, De Grandis RA, Ribeiro DL, Tuttis K, Morselli M, et al. Transcriptome and DNA methylation changes modulated by sulforaphane induce cell cycle arrest, apoptosis, DNA damage, and suppression of proliferation in human liver cancer cells. Food Chem Toxicol. 2020;136:111047.

[29] Cantu C, Bose F, Bianchi P, Reali E, Colzani MT, Cantu I, et al. Defective erythroid maturation in gelsolin mutant mice. Haematologica. 2012;97(7):980-988.

[30] Pandya VA, Patani R. The role of glial cells in amyotrophic lateral sclerosis. Int Rev Neurobiol. 2024;176:381-450.

[31] Cauchi RJ, Tosolini AP. ALS: a field in motion. Sci Rep. 2025;15(1):44791.

[32] Cipollina G, Davari Serej A, Di Nolfi G, Gazzano A, Marsala A, Spatafora MG, et al. Heterogeneity of Neuroinflammatory Responses in Amyotrophic Lateral Sclerosis: A Challenge or an Opportunity? Int J Mol Sci. 2020;21(21).

[33] Gonzalez-Sanchez M, Ramirez-Exposito MJ, Martinez-Martos JM. Pathophysiology, Clinical Heterogeneity, and Therapeutic Advances in Amyotrophic Lateral Sclerosis: A Comprehensive Review of Molecular Mechanisms, Diagnostic Challenges, and Multidisciplinary Management Strategies. Life (Basel). 2025;15(4).

[34] Gomes C, Sequeira C, Barbosa M, Cunha C, Vaz AR, Brites D. Astrocyte regional diversity in ALS includes distinct aberrant phenotypes with common and causal pathological processes. Exp Cell Res. 2020;395(2):112209.

[35] Hur SK, Hunter M, Dominique MA, Farag M, Cotton-Samuel D, Khan T, et al. Slow motor neurons resist pathological TDP-43 and mediate motor recovery in the rNLS8 model of amyotrophic lateral sclerosis. Acta Neuropathol Commun. 2022;10(1):75.

[36] Tsang VSK, Malaspina A, Henson SM. The metabolic intersection between immunosenescence and neuroinflammation in amyotrophic lateral sclerosis. J Inflamm (Lond). 2025;22(1):36.

[37] Ou GY, Lin WW, Zhao WJ. Neuregulins in Neurodegenerative Diseases. Front Aging Neurosci. 2021;13:662474.

[38] Boxer EE, Aoto J. Neurexins and their ligands at inhibitory synapses. Front Synaptic Neurosci. 2022;14:1087238.

[39] Anakor E, Duddy WJ, Duguez S. The Cellular and Molecular Signature of ALS in Muscle. J Pers Med. 2022;12(11).

[40] Zuo X, Zhou J, Li Y, Wu K, Chen Z, Luo Z, et al. TDP-43 aggregation induced by oxidative stress causes global mitochondrial imbalance in ALS. Nat Struct Mol Biol. 2021;28(2):132-142.

[41] Brischigliaro M, Kruger A, Moran JC, Antonicka H, Ahn A, Shoubridge EA, et al. The human mitochondrial translation factor TACO1 alleviates mitoribosome stalling at polyproline stretches. Nucleic Acids Res. 2024;52(16):9710-9726.

[42] Feldt J, Schicht M, Garreis F, Welss J, Schneider UW, Paulsen F. Structure, regulation and related diseases of the actin-binding protein gelsolin. Expert Rev Mol Med. 2019;20:e7.

Figure

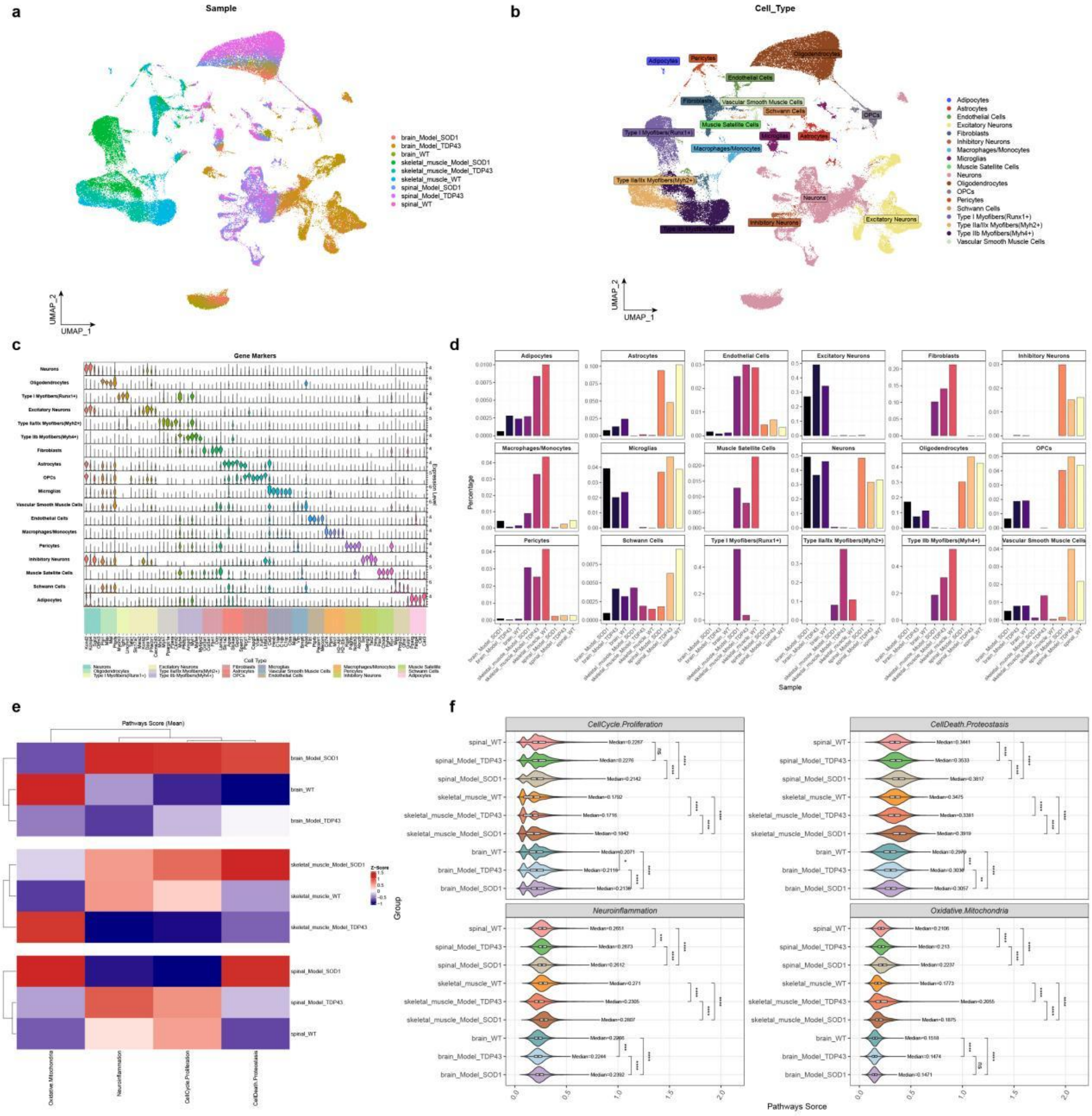


Fig. 1 Cross-tissue single-nucleus atlas of ALS model mice. (**a**) UMAP visualization of the integrated dataset comprising nine sample groups across different models and tissues. (**b**) Manual annotation based on canonical marker genes identified 18 distinct cell types. (**c**) Expression patterns of marker genes for

each cell type, used to validate the cell annotations. (**d**) Relative proportions of the 18 cell types across different model and tissue groups. (**e**) Functional pathway activity scores for four gene sets, including neuroinflammation, apoptosis, oxidative mitochondrial function, and cell proliferation, were estimated using irGSEA. Mean scores were calculated for each group, normalized, and displayed as a heatmap. (**f**) Group-wise comparisons of integrated scores were performed using the Wilcoxon rank-sum test. Significance was defined as *P < 0.05, **P < 0.01, ***P < 0.001, and ****P < 0.0001.

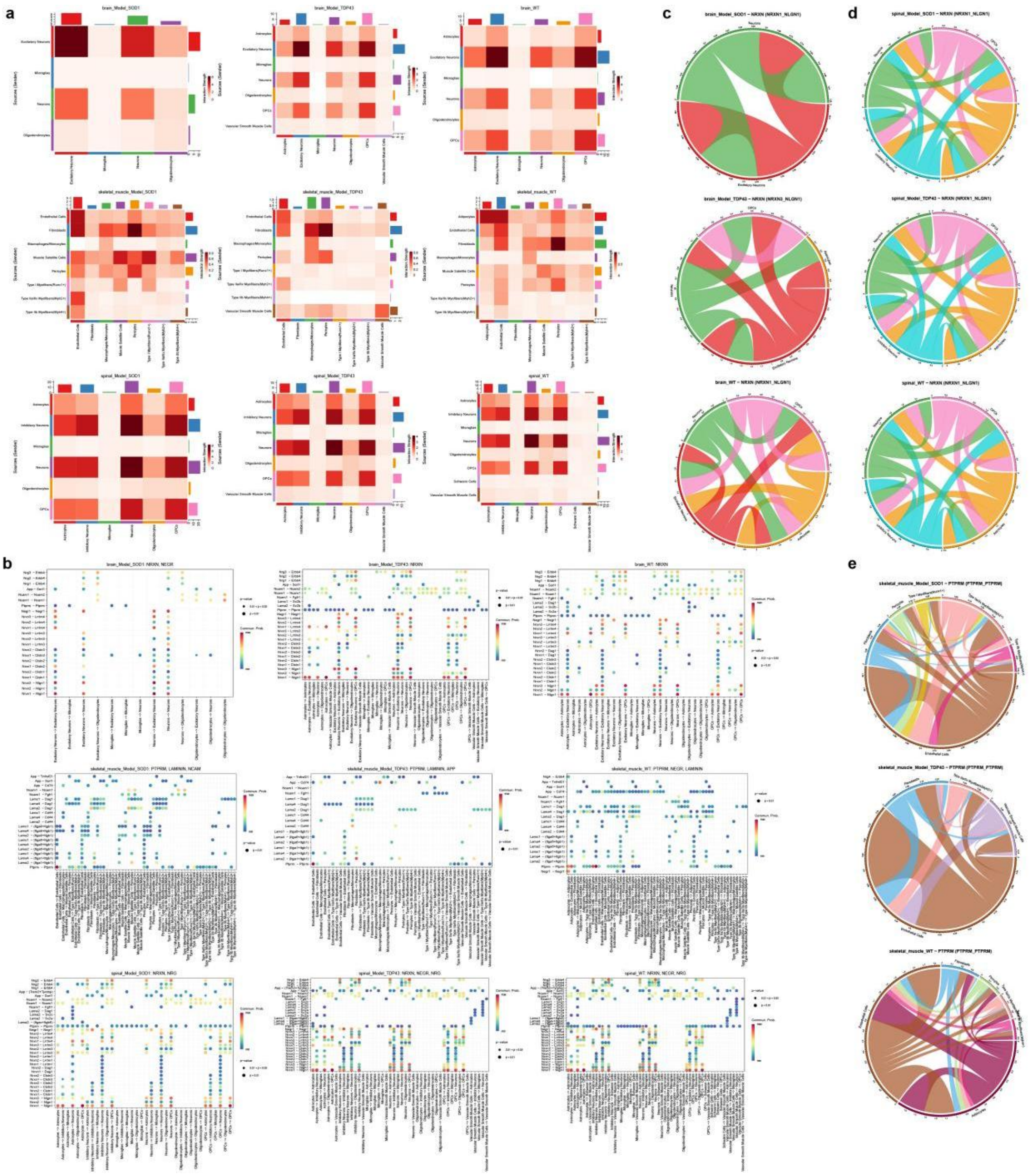


Fig. 2 Cross-tissue cell-cell communication analysis. (**a**) Heatmap of intercellular communication strength across brain, skeletal muscle, and spinal cord tissues in the SOD1, TDP43, and WT groups. Color intensity indicates communication strength, with rows and columns representing sender and

receiver cell types, respectively. (**b**) Visualization of the top three ligand-receptor (LR) interactions in each sample. Color intensity reflects interaction strength, and dot size indicates statistical significance. (**c-e**) The strongest LR interactions in the brain (**c**), spinal cord (**d**), and skeletal muscle (**e**), respectively. Colored arcs denote distinct cell types, arc thickness reflects interaction strength, and arc color indicates the sender cell type.

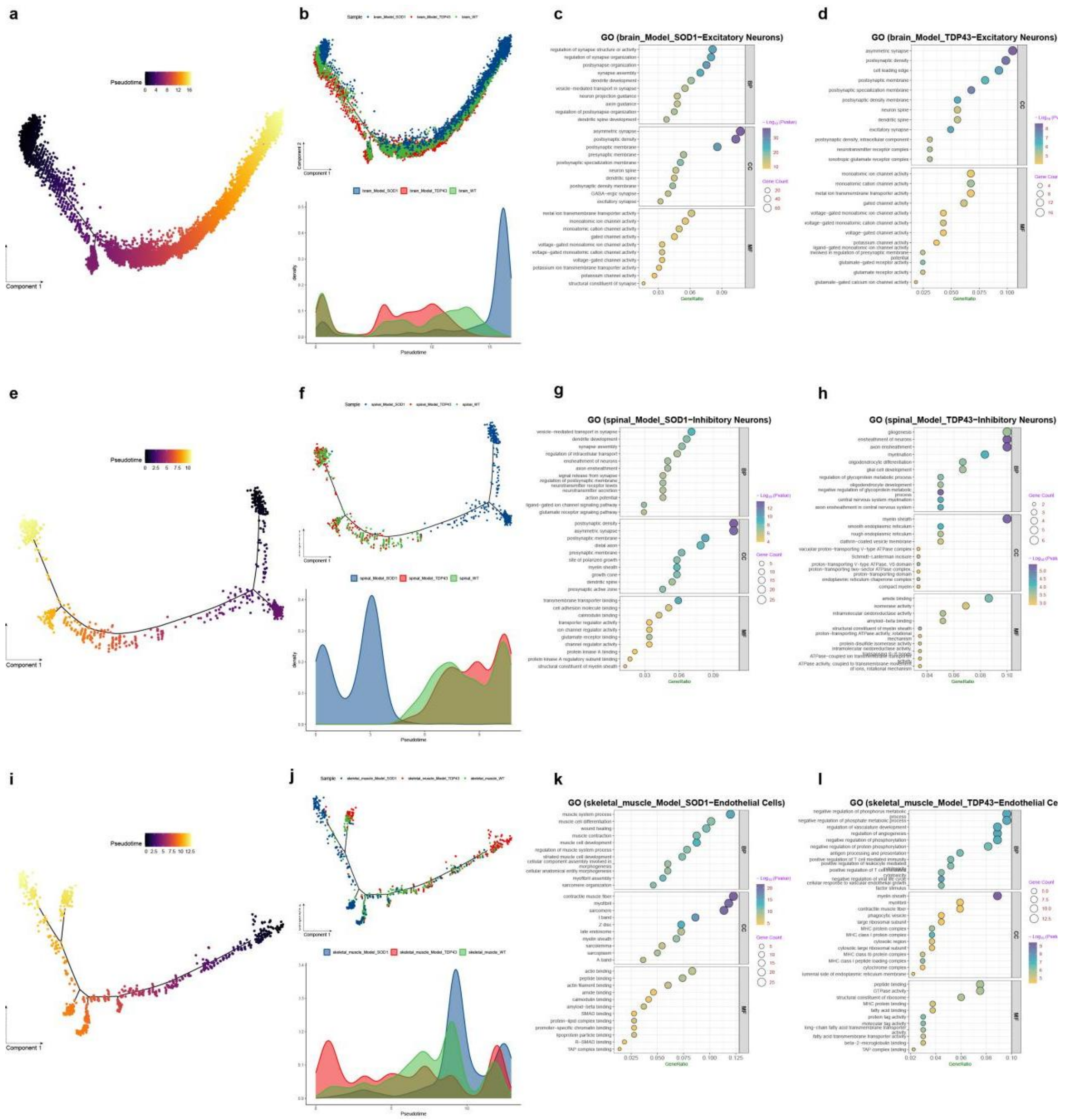


Fig. 3 Cross-tissue pseudotime trajectory inference. (**a**) Differentiation trajectory of excitatory neurons in the brain. (**b**) Mapping of SOD1, TDP43, and WT cells onto the excitatory neuron trajectory inferred from brain tissue. (**c**) GO enrichment analysis of differentially expressed genes (DEGs) in brain excitatory neurons from the SOD1 model. (d) GO enrichment analysis of DEGs in

brain excitatory neurons from the TDP43 model. (**e**) Differentiation trajectory of inhibitory neurons in the spinal cord. (**f**) Mapping of SOD1, TDP43, and WT cells onto the inhibitory neuron trajectory inferred from spinal cord tissue. (**g**) GO enrichment analysis of DEGs in spinal inhibitory neurons from the SOD1 model. (**h**) GO enrichment analysis of DEGs in spinal inhibitory neurons from the TDP43 model. (**i**) Differentiation trajectory of endothelial cells in skeletal muscle. (**j**) Mapping of SOD1, TDP43, and WT cells onto the endothelial cell trajectory inferred from skeletal muscle tissue. (**k**) GO enrichment analysis of DEGs in skeletal muscle endothelial cells from the SOD1 model. (**l**) GO enrichment analysis of DEGs in skeletal muscle endothelial cells from the TDP43 model.

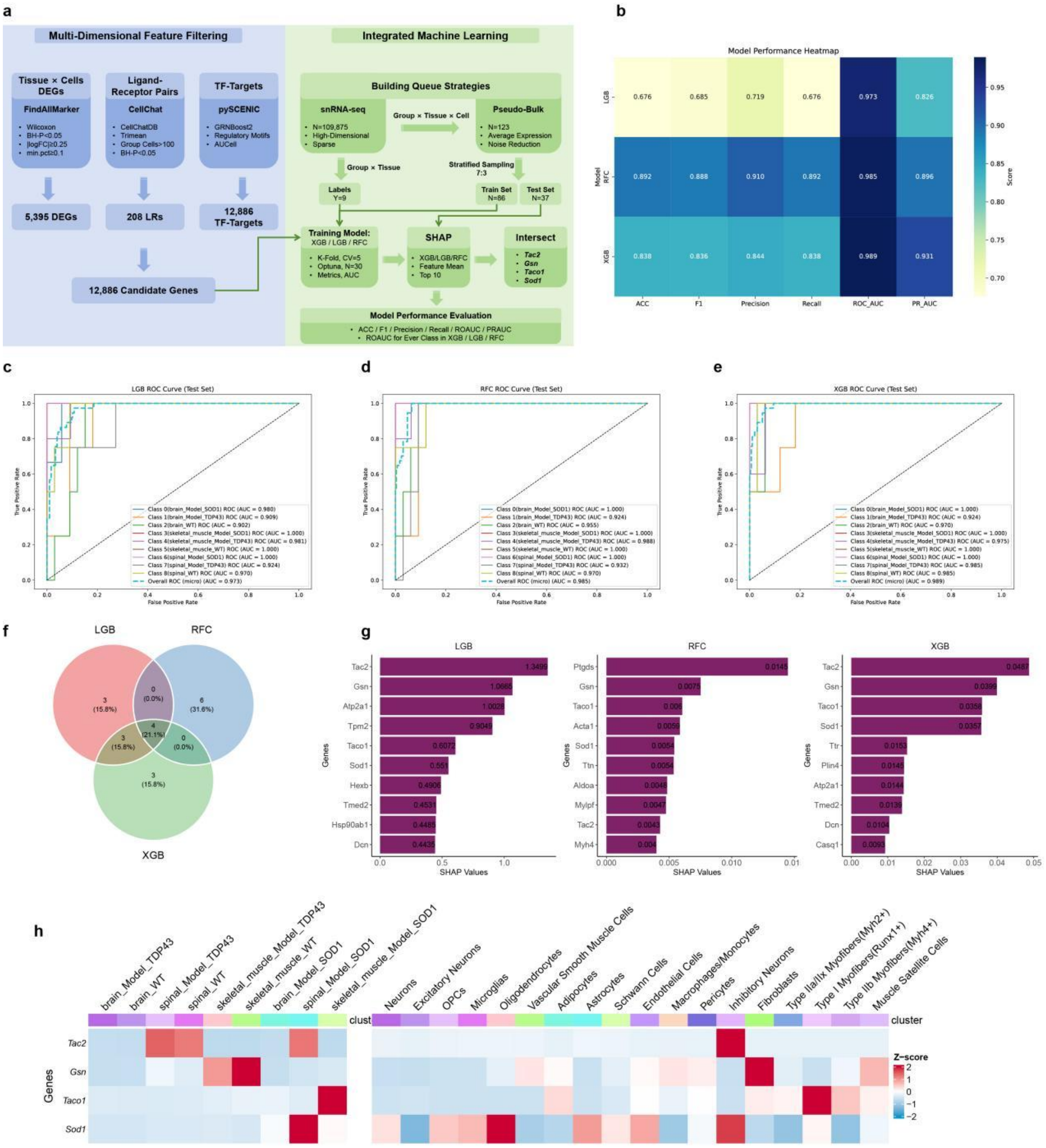


Fig. 4 Integrated machine learning-based identification of ALS biomarkers. (**a**) Workflow for

screening cross-tissue and cross-model biomarkers in ALS. (**b**) XGB, LGB, and RFC models were trained on pseudo-bulk datasets, with AUC used as the hyperparameter optimization metric; model performance was evaluated using ACC, F1, Precision, Recall, PR-AUC, and ROC-AUC. (**c-e**) Class-specific and overall ROC curves for the LGB (**c**), RFC (**d**), and XGB (**e**) models in the test set. (**f**) Feature importance of the three models was assessed using SHAP, and the intersection of the top 10 SHAP-ranked genes from each model was selected. (**g**) SHAP value distribution of the top 10 genes from each model. (**h**) Heatmap showing the mean expression of the final four candidate genes across tissue × model strata and at the cellular atlas level.

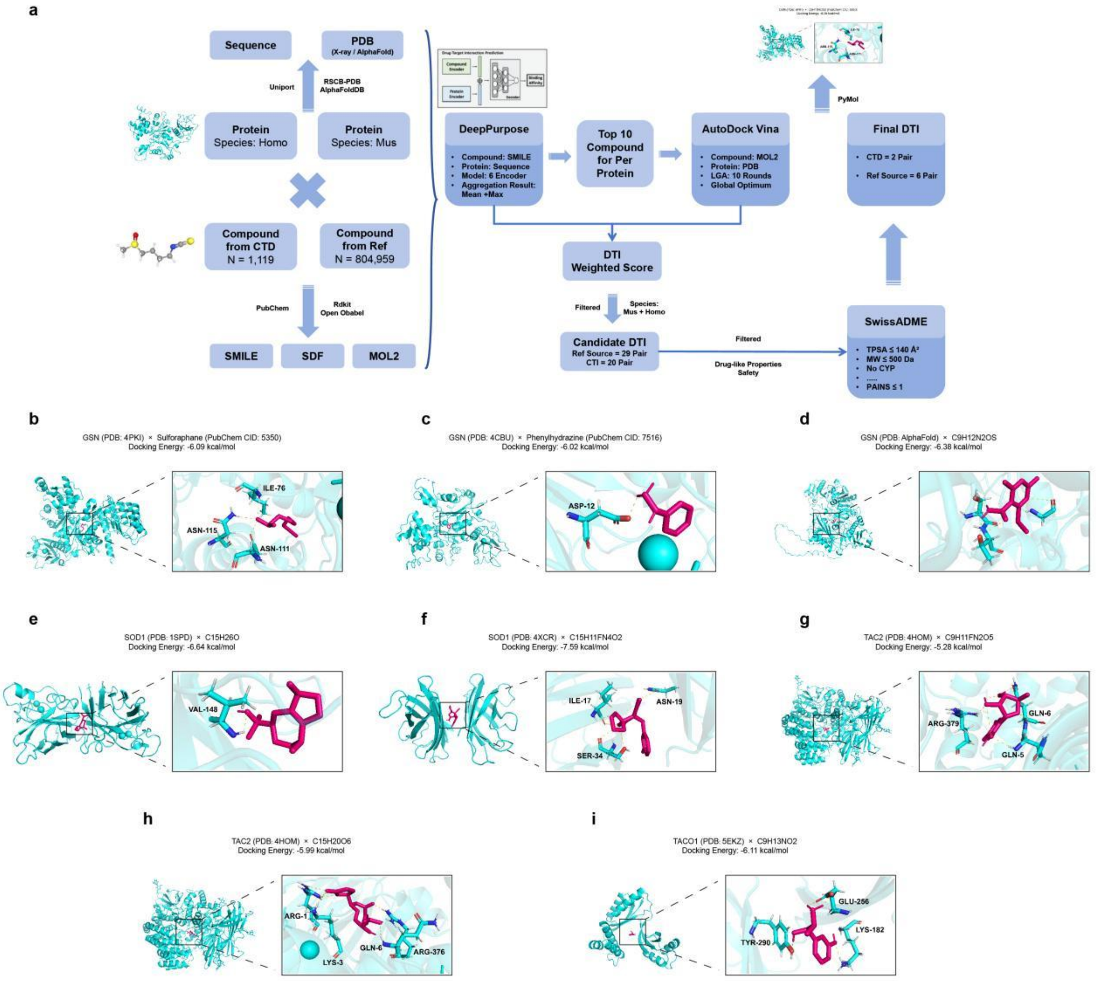


Fig. 5 Small-molecule compounds with potential druggability based on ALS biomarkers. (**a**) Workflow for potential drug screening and evaluation based on four cross-tissue, cross-model ALS biomarkers, integrating large-scale datasets of reported interacting compounds and candidate lead compounds. (**b-i**) Following stringent screening by DeepPurpose and AutoDock, and after assessing

compound safety and drug-likeness, the optimal binding conformations of target-drug interactions were visualized and analyzed using PyMOL. These include GSN with the natural product sulforaphane (**b**), GSN with phenylhydrazine (**c**), GSN with the lead compound C9H12N2OS (**d**), SOD1 with the lead compound C15H26O (**e**), SOD1 with the lead compound C15H11FN4O2 (**f**), TAC2 with the lead compound C9H11FN2O5 (**g**), TAC2 with the lead compound C15H20O6 (**h**), and TACO1 with the lead compound C9H13NO2 (**i**).

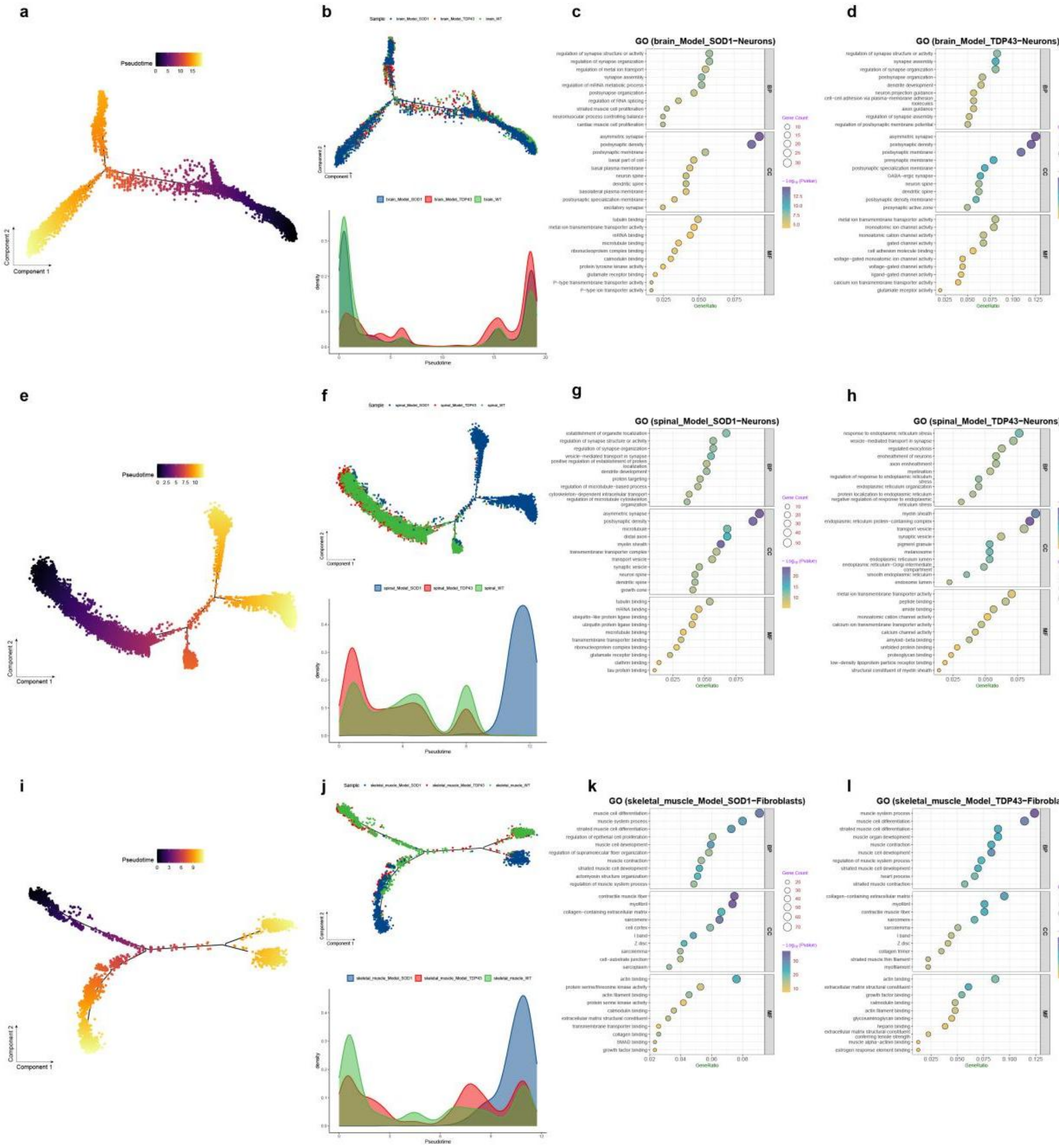


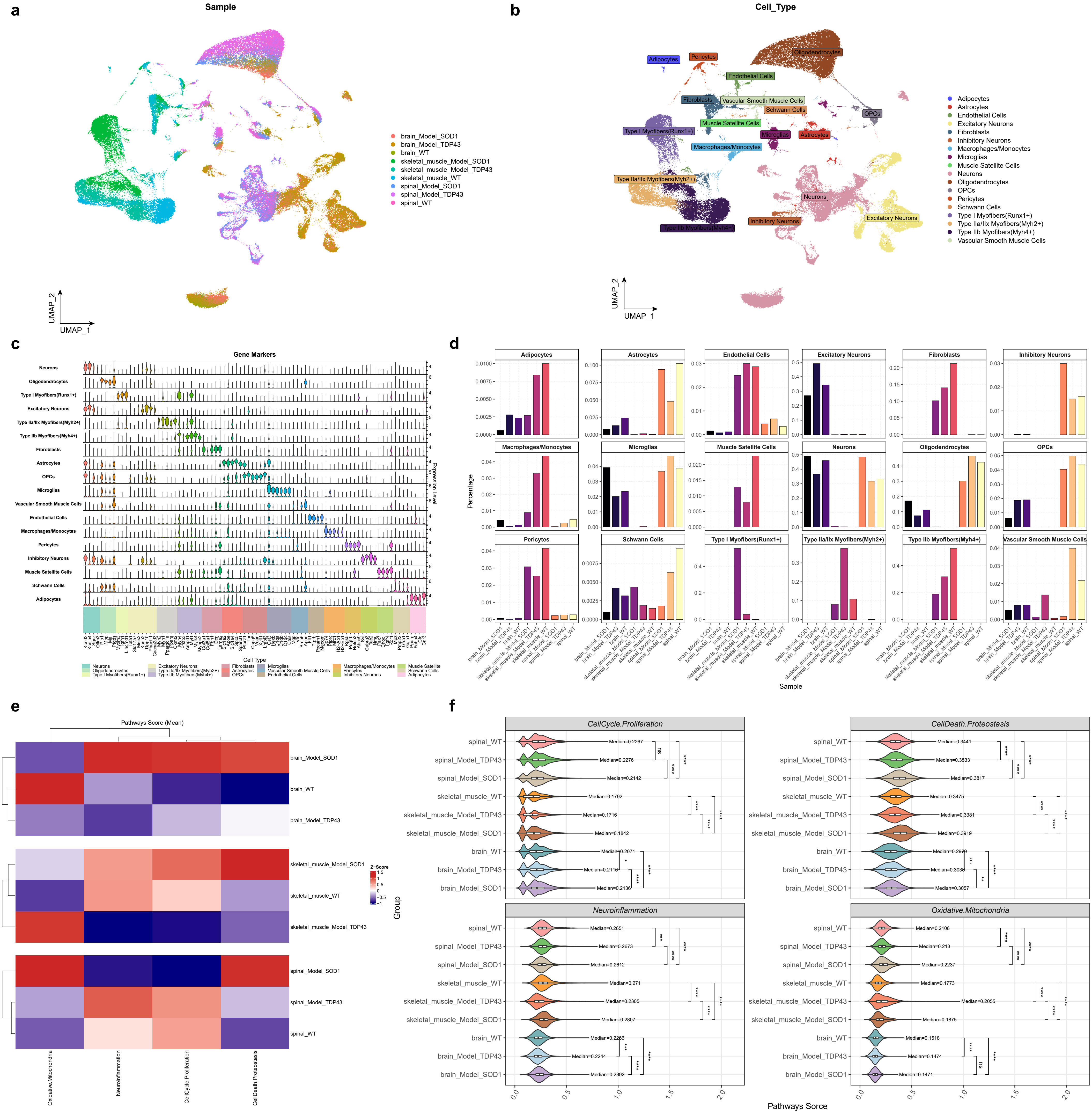
Fig. S1 Cross-tissue pseudotime trajectory inference. (**a**) Differentiation trajectory of neurons in the brain. (**b**) Mapping of SOD1, TDP43, and WT cells onto the neuronal trajectory inferred from brain tissue. (**c**) GO enrichment analysis of DEGs in brain neurons from the SOD1 model. (**d**) GO enrichment analysis of DEGs in brain neurons from the TDP43 model. (**e**) Differentiation trajectory of neurons in the spinal cord. (**f**) Mapping of SOD1, TDP43, and WT cells onto the neuronal trajectory inferred from spinal cord tissue. (**g**) GO enrichment analysis of DEGs in spinal neurons from the SOD1 model. (**h**) GO enrichment analysis of DEGs in spinal neurons from the TDP43 model. (**i**) Differentiation trajectory of fibroblasts in skeletal muscle. (**j**) Mapping of SOD1, TDP43, and WT cells onto the fibroblast trajectory inferred from skeletal muscle tissue. (**k**) GO enrichment analysis of DEGs in skeletal muscle fibroblasts from the SOD1 model. (**l**) GO enrichment analysis of DEGs in skeletal muscle endothelial cells from the TDP43 model.

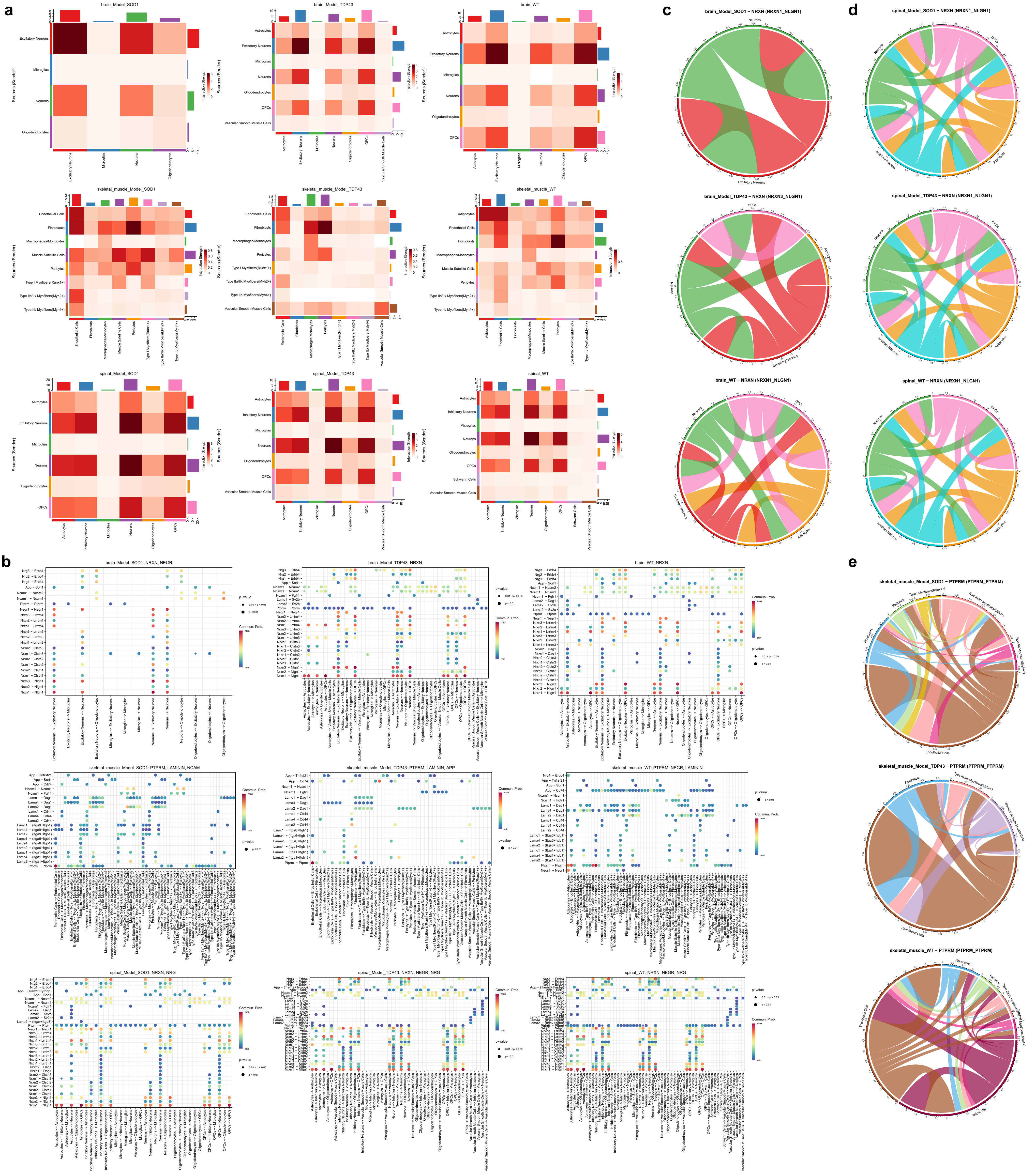
Supplementary Material

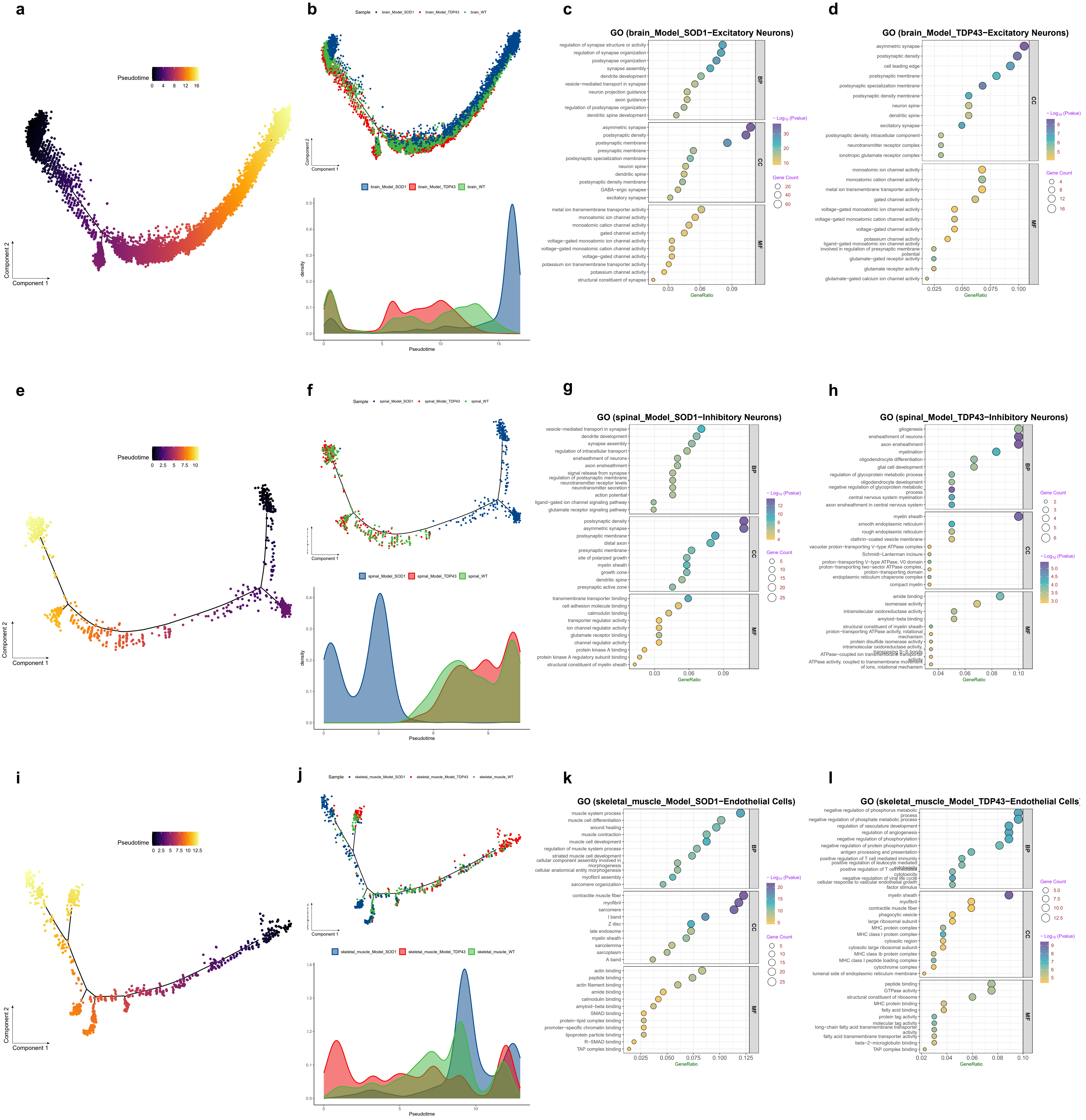
Supplementary Table 1. Gene sets. (**a**) Marker genes used for cell-type annotation. (**b**) Four manually curated gene sets used for functional assessment: Neuroinflammation, Oxidative Mitochondria, CellDeath Proteostasis, and CellCycle Proliferation.

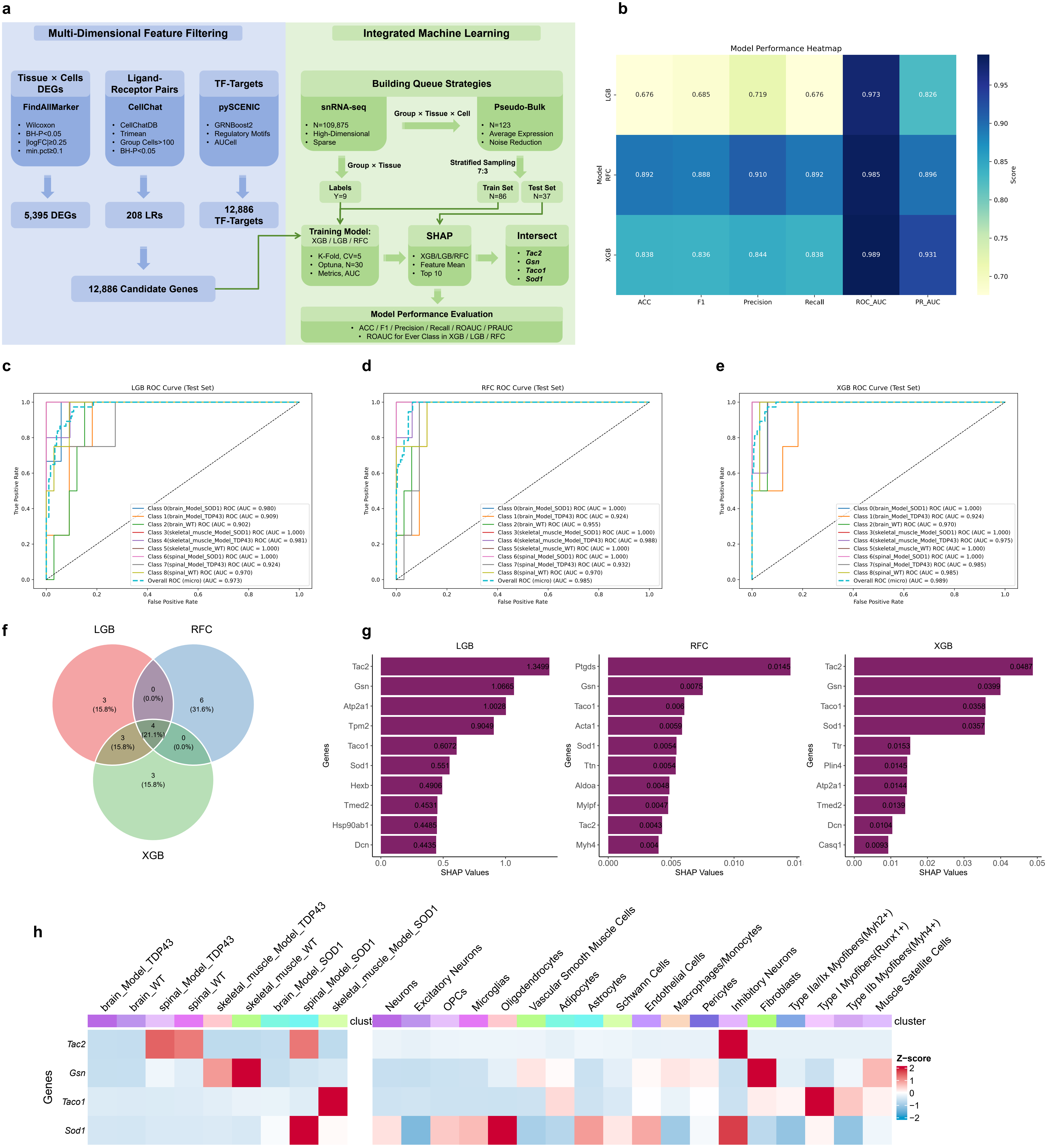
Supplementary Table 2. Screening results of CTD-derived compounds against the four ALS biomarkers.

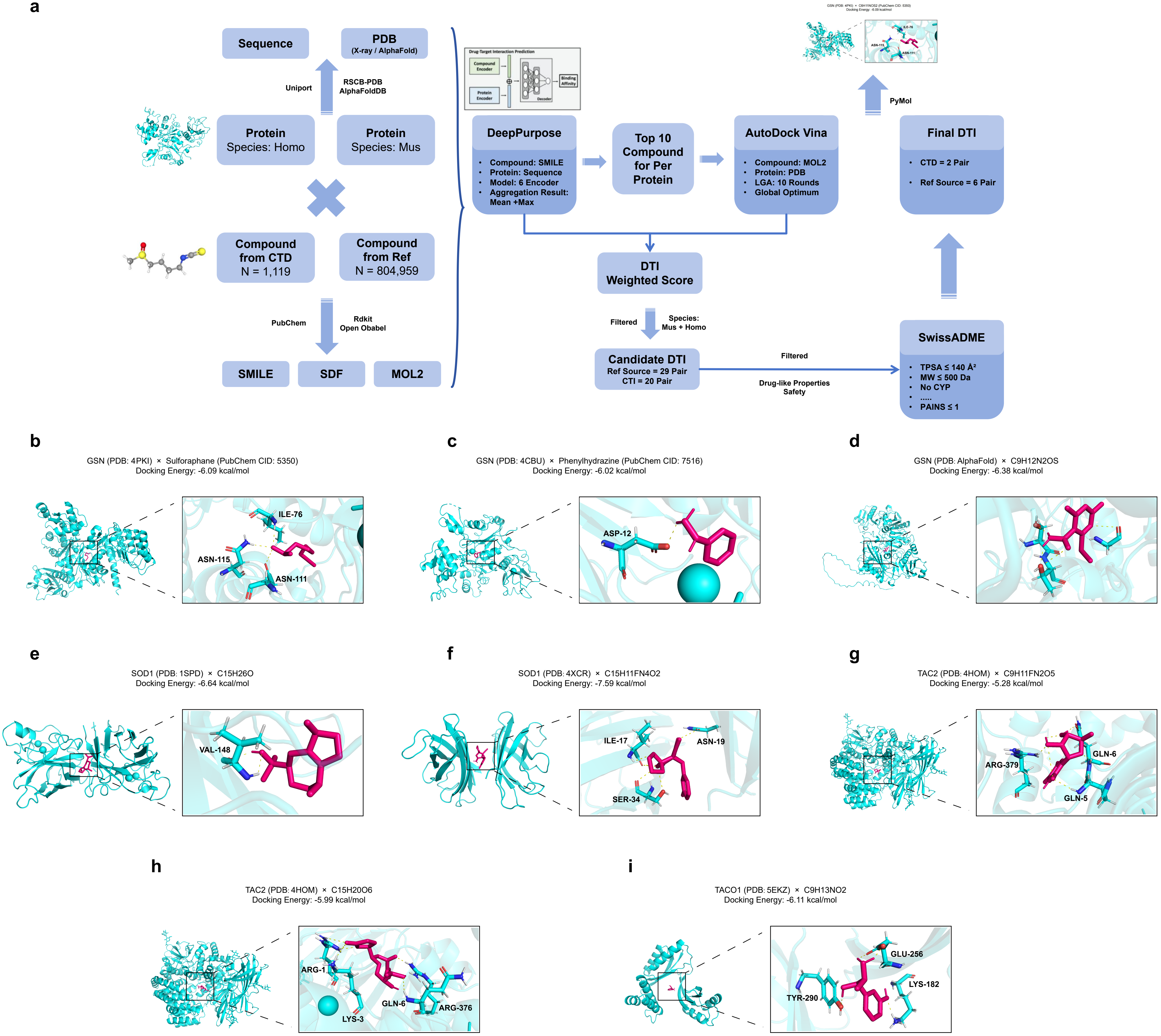
Supplementary Table 3. Screening results of literature-derived compounds against the four ALS biomarkers.


a
Sample
brain_Model_SOD1
brain_Model_TDP43
brain_WT
skeletal_muscle_Model_SOD1
skeletal_muscle_Model_TDP43
skeletal_muscle_WT
spinal_Model_SOD1
spinal_Model_TDP43
spinal_WT
UMAP_1
UMAP_2
b
Cell_Type
Adipocytes
Astrocytes
Endothelial Cells
Excitatory Neurons
Fibroblasts
Inhibitory Neurons
Macrophages/Monocytes
Microglias
Muscle Satellite Cells
Neurons
Oligodendrocytes
OPCs
Pericytes
Schwann Cells
Type I Myofibers(Runx1+)
Type IIa/IIx Myofibers(Myh2+)
Type IIb Myofibers(Myh4+)
Vascular Smooth Muscle Cells
c
Gene Markers
Expression Level
Cell Type
d
Percentage
Sample
e
Pathways Score (Mean)
Z-Score
1.5
1
0.5
0
−0.5
−1
Oxidative.Mitochondria
Neuroinflammation
CellCycle.Proliferation
CellDeath.Proteostasis
f
Group
Pathways Sorce
CellCycle.Proliferation
Median=0.2267
Median=0.2276
Median=0.2142
Median=0.1792
Median=0.1716
Median=0.1842
Median=0.2071
Median=0.2116
Median=0.2136
CellDeath.Proteostasis
Median=0.3441
Median=0.3533
Median=0.3817
Median=0.3475
Median=0.3381
Median=0.3919
Median=0.2979
Median=0.3036
Median=0.3057
Neuroinflammation
Median=0.2651
Median=0.2673
Median=0.2612
Median=0.271
Median=0.2305
Median=0.2807
Median=0.2266
Median=0.2244
Median=0.2392
Oxidative.Mitochondria
Median=0.2106
Median=0.213
Median=0.2237
Median=0.1773
Median=0.2055
Median=0.1875
Median=0.1518
Median=0.1474
Median=0.1471
ns
0.0
0.5
1.0
1.5
2.0


a
brain_Model_SOD1
brain_Model_TDP43
brain_WT
skeletal_muscle_Model_SOD1
skeletal_muscle_Model_TDP43
skeletal_muscle_WT
spinal_Model_SOD1
spinal_Model_TDP43
spinal_WT
Sources (Sender)
Interaction Strength
b
brain_Model_SOD1: NRXN, NEGR
brain_Model_TDP43: NRXN
brain_WT: NRXN
skeletal_muscle_Model_SOD1: PTPRM, LAMININ, NCAM
skeletal_muscle_Model_TDP43: PTPRM, LAMININ, APP
skeletal_muscle_WT: PTPRM, NEGR, LAMININ
spinal_Model_SOD1: NRXN, NRG
spinal_Model_TDP43: NRXN, NEGR, NRG
spinal_WT: NRXN, NEGR, NRG
Commun. Prob.
p-value
c
brain_Model_SOD1 - NRXN (NRXN1_NLGN1)
brain_Model_TDP43 - NRXN (NRXN3_NLGN1)
brain_WT - NRXN (NRXN1_NLGN1)
d
spinal_Model_SOD1 - NRXN (NRXN1_NLGN1)
spinal_Model_TDP43 - NRXN (NRXN1_NLGN1)
spinal_WT - NRXN (NRXN1_NLGN1)
e
skeletal_muscle_Model_SOD1 - PTPRM (PTPRM_PTPRM)
skeletal_muscle_Model_TDP43 - PTPRM (PTPRM_PTPRM)
skeletal_muscle_WT - PTPRM (PTPRM_PTPRM)


a
Pseudotime
0 4 8 12 16
Component 2
Component 1
b
Sample
brain_Model_SOD1
brain_Model_TDP43
brain_WT
Component 2
Component 1
brain_Model_SOD1
brain_Model_TDP43
brain_WT
density
0.5
0.4
0.3
0.2
0.1
0.0
0
5
10
15
Pseudotime
c
GO (brain_Model_SOD1−Excitatory Neurons)
regulation of synapse structure or activity
regulation of synapse organization
postsynapse organization
synapse assembly
dendrite development
vesicle−mediated transport in synapse
neuron projection guidance
axon guidance
regulation of postsynapse organization
dendritic spine development
BP
asymmetric synapse
postsynaptic density
postsynaptic membrane
presynaptic membrane
postsynaptic specialization membrane
neuron spine
dendritic spine
postsynaptic density membrane
GABA−ergic synapse
excitatory synapse
CC
metal ion transmembrane transporter activity
monoatomic ion channel activity
monoatomic cation channel activity
gated channel activity
voltage−gated monoatomic ion channel activity
voltage−gated monoatomic cation channel activity
voltage−gated channel activity
potassium ion transmembrane transporter activity
potassium channel activity
structural constituent of synapse
MF
0.03
0.06
0.09
GeneRatio
− Log10 (Pvalue)
30
20
10
Gene Count
20
40
60
d
GO (brain_Model_TDP43−Excitatory Neurons)
asymmetric synapse
postsynaptic density
cell leading edge
postsynaptic membrane
postsynaptic specialization membrane
postsynaptic density membrane
neuron spine
dendritic spine
excitatory synapse
postsynaptic density, intracellular component
neurotransmitter receptor complex
ionotropic glutamate receptor complex
CC
monoatomic ion channel activity
monoatomic cation channel activity
metal ion transmembrane transporter activity
gated channel activity
voltage−gated monoatomic ion channel activity
voltage−gated monoatomic cation channel activity
voltage−gated channel activity
potassium channel activity
ligand−gated monoatomic ion channel activity involved in regulation of presynaptic membrane potential
glutamate−gated receptor activity
glutamate receptor activity
glutamate−gated calcium ion channel activity
MF
0.025
0.050
0.075
0.100
GeneRatio
− Log10 (Pvalue)
8
7
6
5
Gene Count
4
8
12
16
e
Pseudotime
0 2.5 5 7.5 10
Component 2
Component 1
f
Sample
spinal_Model_SOD1
spinal_Model_TDP43
spinal_WT
Component 1
spinal_Model_SOD1
spinal_Model_TDP43
spinal_WT
density
0.4
0.3
0.2
0.1
0.0
0
3
6
9
Pseudotime
g
GO (spinal_Model_SOD1−Inhibitory Neurons)
vesicle−mediated transport in synapse
dendrite development
synapse assembly
regulation of intracellular transport
ensheathment of neurons
axon ensheathment
signal release from synapse
regulation of postsynaptic membrane neurotransmitter receptor levels
neurotransmitter secretion
action potential
ligand−gated ion channel signaling pathway
glutamate receptor signaling pathway
BP
postsynaptic density
asymmetric synapse
postsynaptic membrane
distal axon
presynaptic membrane
site of polarized growth
myelin sheath
growth cone
dendritic spine
presynaptic active zone
CC
transmembrane transporter binding
cell adhesion molecule binding
calmodulin binding
transporter regulator activity
ion channel regulator activity
glutamate receptor binding
channel regulator activity
protein kinase A binding
protein kinase A regulatory subunit binding
structural constituent of myelin sheath
MF
0.03
0.06
0.09
GeneRatio
− Log10 (Pvalue)
12
10
8
6
4
Gene Count
5
10
15
20
25
h
GO (spinal_Model_TDP43−Inhibitory Neurons)
gliogenesis
ensheathment of neurons
axon ensheathment
myelination
oligodendrocyte differentiation
glial cell development
regulation of glycoprotein metabolic process
oligodendrocyte development
negative regulation of glycoprotein metabolic process
central nervous system myelination
axon ensheathment in central nervous system
BP
myelin sheath
smooth endoplasmic reticulum
rough endoplasmic reticulum
clathrin−coated vesicle membrane
vacuolar proton−transporting V−type ATPase complex
Schmidt−Lanterman incisure
proton−transporting V−type ATPase, V0 domain
proton−transporting two−sector ATPase complex, proton−transporting domain
endoplasmic reticulum chaperone complex
compact myelin
CC
amide binding
isomerase activity
intramolecular oxidoreductase activity
amyloid−beta binding
structural constituent of myelin sheath
proton−transporting ATPase activity, rotational mechanism
protein disulfide isomerase activity
intramolecular oxidoreductase activity, transposing S−S bonds
ATPase−coupled ion transmembrane transporter activity
ATPase activity, coupled to transmembrane movement of ions, rotational mechanism
MF
0.04
0.06
0.08
0.10
GeneRatio
Gene Count
2
3
4
5
6
− Log10 (Pvalue)
5.0
4.5
4.0
3.5
3.0
i
Pseudotime
0 2.5 5 7.5 10 12.5
Component 2
Component 1
j
Sample
skeletal_muscle_Model_SOD1
skeletal_muscle_Model_TDP43
skeletal_muscle_WT
Component 1
skeletal_muscle_Model_SOD1
skeletal_muscle_Model_TDP43
skeletal_muscle_WT
0.1
0.0
0
5
10
Pseudotime
k
GO (skeletal_muscle_Model_SOD1−Endothelial Cells)
muscle system process
muscle cell differentiation
wound healing
muscle contraction
muscle cell development
regulation of muscle system process
striated muscle cell development
cellular component assembly involved in morphogenesis
cellular anatomical entity morphogenesis
myofibril assembly
sarcomere organization
BP
contractile muscle fiber
myofibril
sarcomere
I band
Z disc
late endosome
myelin sheath
sarcolemma
sarcoplasm
A band
CC
actin binding
peptide binding
actin filament binding
amide binding
calmodulin binding
amyloid−beta binding
SMAD binding
protein−lipid complex binding
promoter−specific chromatin binding
lipoprotein particle binding
R−SMAD binding
TAP complex binding
MF
0.025
0.050
0.075
0.100
0.125
GeneRatio
− Log10 (Pvalue)
20
15
10
5
Gene Count
5
10
15
20
25
l
GO (skeletal_muscle_Model_TDP43−Endothelial Cells
negative regulation of phosphorus metabolic process
negative regulation of phosphate metabolic process
regulation of vasculature development
regulation of angiogenesis
negative regulation of phosphorylation
negative regulation of protein phosphorylation
antigen processing and presentation
positive regulation of T cell mediated immunity
positive regulation of leukocyte mediated cytotoxicity
positive regulation of T cell mediated cytotoxicity
negative regulation of viral life cycle
cellular response to vascular endothelial growth factor stimulus
BP
myelin sheath
myofibril
contractile muscle fiber
phagocytic vesicle
large ribosomal subunit
MHC protein complex
MHC class I protein complex
cytosolic region
cytosolic large ribosomal subunit
MHC class Ib protein complex
MHC class I peptide loading complex
cytochrome complex
lumenal side of endoplasmic reticulum membrane
CC
peptide binding
GTPase activity
structural constituent of ribosome
MHC protein binding
fatty acid binding
protein tag activity
molecular tag activity
long−chain fatty acid transmembrane transporter activity
fatty acid transmembrane transporter activity
beta−2−microglobulin binding
TAP complex binding
MF
0.02
0.04
0.06
0.08
0.10
GeneRatio
Gene Count
5.0
7.5
10.0
12.5
− Log10 (Pvalue)
9
8
7
6
5


a
Multi-Dimensional Feature Filtering
Tissue × Cells DEGs
FindAllMarker
Wilcoxon
BH-P<0.05
|logFC|≥0.25
min.pct≥0.1
Ligand-Receptor Pairs
CellChat
CellChatDB
Trimean
Group Cells>100
BH-P<0.05
TF-Targets
pySCENIC
GRNBoost2
Regulatory Motifs
AUCell
5,395 DEGs
208 LRs
12,886 TF-Targets
12,886 Candidate Genes
Integrated Machine Learning
Building Queue Strategies
snRNA-seq
N=109,875
High-Dimensional
Sparse
Group × Tissue × Cell
Pseudo-Bulk
N=123
Average Expression
Noise Reduction
Group × Tissue
Stratified Sampling 7:3
Labels Y=9
Train Set N=86
Test Set N=37
Training Model: XGB / LGB / RFC
K-Fold, CV=5
Optuna, N=30
Metrics, AUC
SHAP
XGB/LGB/RFC
Feature Mean
Top 10
Intersect
Tac2
Gsn
Taco1
Sod1
Model Performance Evaluation
ACC / F1 / Precision / Recall / ROAUC / PRAUC
ROAUC for Ever Class in XGB / LGB / RFC
b
Model Performance Heatmap
Model
LGB 0.676 0.685 0.719 0.676 0.973 0.826
RFC 0.892 0.888 0.910 0.892 0.985 0.896
XGB 0.838 0.836 0.844 0.838 0.989 0.931
ACC F1 Precision Recall ROC_AUC PR_AUC
Score 0.95 0.90 0.85 0.80 0.75 0.70
c
LGB ROC Curve (Test Set)
True Positive Rate
False Positive Rate
Class 0(brain_Model_SOD1) ROC (AUC = 0.980)
Class 1(brain_Model_TDP43) ROC (AUC = 0.909)
Class 2(brain_WT) ROC (AUC = 0.902)
Class 3(skeletal_muscle_Model_SOD1) ROC (AUC = 1.000)
Class 4(skeletal_muscle_Model_TDP43) ROC (AUC = 0.981)
Class 5(skeletal_muscle_WT) ROC (AUC = 1.000)
Class 6(spinal_Model_SOD1) ROC (AUC = 1.000)
Class 7(spinal_Model_TDP43) ROC (AUC = 0.924)
Class 8(spinal_WT) ROC (AUC = 0.970)
Overall ROC (micro) (AUC = 0.973)
d
RFC ROC Curve (Test Set)
Class 0(brain_Model_SOD1) ROC (AUC = 1.000)
Class 1(brain_Model_TDP43) ROC (AUC = 0.924)
Class 2(brain_WT) ROC (AUC = 0.955)
Class 3(skeletal_muscle_Model_SOD1) ROC (AUC = 1.000)
Class 4(skeletal_muscle_Model_TDP43) ROC (AUC = 0.988)
Class 5(skeletal_muscle_WT) ROC (AUC = 1.000)
Class 6(spinal_Model_SOD1) ROC (AUC = 1.000)
Class 7(spinal_Model_TDP43) ROC (AUC = 0.932)
Class 8(spinal_WT) ROC (AUC = 0.970)
Overall ROC (micro) (AUC = 0.985)
e
XGB ROC Curve (Test Set)
Class 0(brain_Model_SOD1) ROC (AUC = 1.000)
Class 1(brain_Model_TDP43) ROC (AUC = 0.924)
Class 2(brain_WT) ROC (AUC = 0.970)
Class 3(skeletal_muscle_Model_SOD1) ROC (AUC = 1.000)
Class 4(skeletal_muscle_Model_TDP43) ROC (AUC = 0.975)
Class 5(skeletal_muscle_WT) ROC (AUC = 1.000)
Class 6(spinal_Model_SOD1) ROC (AUC = 1.000)
Class 7(spinal_Model_TDP43) ROC (AUC = 0.985)
Class 8(spinal_WT) ROC (AUC = 0.985)
Overall ROC (micro) (AUC = 0.989)
f
LGB
RFC
XGB
3 (15.8%)
0 (0.0%)
6 (31.6%)
4 (21.1%)
3 (15.8%)
0 (0.0%)
3 (15.8%)
g
LGB
Tac2 1.3499
Gsn 1.0665
Atp2a1 1.0028
Tpm2 0.9049
Taco1 0.6072
Sod1 0.551
Hexb 0.4906
Tmed2 0.4531
Hsp90ab1 0.4485
Dcn 0.4435
RFC
Ptgds 0.0145
Gsn 0.0075
Taco1 0.006
Acta1 0.0059
Sod1 0.0054
Ttn 0.0054
Aldoa 0.0048
Mylpf 0.0047
Tac2 0.0043
Myh4 0.004
XGB
Tac2 0.0487
Gsn 0.0399
Taco1 0.0358
Sod1 0.0357
Ttr 0.0153
Plin4 0.0145
Atp2a1 0.0144
Tmed2 0.0139
Dcn 0.0104
Casq1 0.0093
Genes
SHAP Values
h
brain_Model_TDP43
brain_WT
spinal_Model_TDP43
spinal_WT
skeletal_muscle_Model_TDP43
skeletal_muscle_WT
brain_Model_SOD1
spinal_Model_SOD1
skeletal_muscle_Model_SOD1
Neurons
Excitatory Neurons
OPCs
Microglias
Oligodendrocytes
Vascular Smooth Muscle Cells
Adipocytes
Astrocytes
Schwann Cells
Endothelial Cells
Macrophages/Monocytes
Pericytes
Inhibitory Neurons
Fibroblasts
Type IIa/IIx Myofibers(Myh2+)
Type I Myofibers(Runx1+)
Type IIb Myofibers(Myh4+)
Muscle Satellite Cells
clust
cluster
Tac2
Gsn
Taco1
Sod1
Genes
Z-score
2 1 0 −1 −2


a
Sequence
PDB
(X-ray / AlphaFold)
Uniport
RSCB-PDB
AlphaFoldDB
Protein
Species: Homo
Protein
Species: Mus
Compound
from CTD
N = 1,119
Compound
from Ref
N = 804,959
PubChem
Rdkit
Open Obabel
SMILE
SDF
MOL2
Drug-Target Interaction Prediction
Compound Encoder
Protein Encoder
Decoder
Binding Affinity
DeepPurpose
Compound: SMILE
Protein: Sequence
Model: 6 Encoder
Aggregation Result: Mean +Max
Top 10 Compound for Per Protein
AutoDock Vina
Compound: MOL2
Protein: PDB
LGA: 10 Rounds
Global Optimum
DTI Weighted Score
Filtered
Species: Mus + Homo
Candidate DTI
Ref Source = 29 Pair
CTI = 20 Pair
Filtered
Drug-like Properties
Safety
SwissADME
TPSA ≤ 140 Å²
MW ≤ 500 Da
No CYP
.....
PAINS ≤ 1
Final DTI
CTD = 2 Pair
Ref Source = 6 Pair
PyMol
GSN (PDB: 4PKI) × C6H11NOS2 (PubChem CID: 5350)
Docking Energy: -6.09 kcal/mol
ILE-76
ASN-115
ASN-111
b
GSN (PDB: 4PKI) × Sulforaphane (PubChem CID: 5350)
Docking Energy: -6.09 kcal/mol
ILE-76
ASN-115
ASN-111
c
GSN (PDB: 4CBU) × Phenylhydrazine (PubChem CID: 7516)
Docking Energy: -6.02 kcal/mol
ASP-12
d
GSN (PDB: AlphaFold) × C9H12N2OS
Docking Energy: -6.38 kcal/mol
e
SOD1 (PDB: 1SPD) × C15H26O
Docking Energy: -6.64 kcal/mol
VAL-148
f
SOD1 (PDB: 4XCR) × C15H11FN4O2
Docking Energy: -7.59 kcal/mol
ILE-17
ASN-19
SER-34
g
TAC2 (PDB: 4HOM) × C9H11FN2O5
Docking Energy: -5.28 kcal/mol
ARG-379
GLN-6
GLN-5
h
TAC2 (PDB: 4HOM) × C15H20O6
Docking Energy: -5.99 kcal/mol
ARG-1
LYS-3
GLN-6
ARG-376
i
TACO1 (PDB: 5EKZ) × C9H13NO2
Docking Energy: -6.11 kcal/mol
GLU-256
LYS-182
TYR-290