SUPPLEMENTARY DATA

Multi-omic machine learning overcomes transcriptomic blind spots to prioritize therapeutic targets in ALS

Megha Gupta[1], Paras Verma[1] and Monika Sharma[1*]

[1] Biological Systems Engineering, Plaksha University, Sector 101 alpha, IT City, SAS Nagar, Mohali, Punjab, 140306, India

Supplementary Results for detailed XAI results:

Supplementary Note 1. Spatial Resolution and Molecular Validation of the Myeloid Driver

To pinpoint the anatomical provenance of the immune dysregulation observed in our network analysis, we examined tissue-specific immune signatures (**Supplementary Figure S8**). This validation step confirms that the myeloid signal is a genuine manifestation of localized pathology at the site of neurodegeneration rather than a systemic artifact.

In **Supplementary Figure S8A**, we deconvoluted the immune signals across diverse tissue sources. The myeloid-specific signature (blue bars) does not appear ubiquitously; instead, it exhibits precise spatial fidelity. It reaches its highest magnitude in the anterior horn of the spinal cord and motor neurons, the primary anatomical places of ALS pathology. Notably, this myeloid signature is conspicuously downregulated in support cells like astrocytes and oligodendrocytes, as well as in circulating lymphocytes. This distribution suggests that the inflammatory driver we identified is not a general systemic feature but represents active myeloid infiltration or microglial activation specifically within the vulnerable motor unit[1].

Complementing this spatial resolution, **Supplementary Figure S8B** bridges these broad pathway signals to the prioritized candidate genes. The heatmap links the high-level immune signatures to specific gene expression profiles, stratified by their machine learning classification. Notably, the target annotation (left column) reveals that the novel AI-prioritized discoveries (red indicators) deeply mimic the immune profile of positive-class genes (green indicators). Rather than scattering randomly across immune pathways, the computationally prioritized targets cluster cleanly within the shared and myeloid-specific neuroinflammatory responses.

Together, these data support an myeloid-driven immune response as a localized and anatomically specific pathology. The pattern is enriched in the anterior-horn microenvironment and provides candidate molecular signals associated to local microglial and macrophage activation for therapeutic modulation.

Supplementary Note 2. Detailed analyses of XAI consensus pipeline

From Transcriptomic Prioritization to Multi-Source Genomic Integration

The transcriptomic meta-analysis identified a prioritized set of molecular targets through systematic integration of 13 independent ALS datasets spanning four biological domains. However, transcriptomic dysregulation alone, while necessary, is insufficient to establish a gene's candidacy as a therapeutic target. A gene may be differentially expressed as a downstream consequence of disease pathology rather than as a causal contributor, and expression-based prioritization cannot assess whether a candidate possesses the genomic, network, and disease-association properties characteristic of established ALS genes.

To address this limitation, we developed an explainable artificial intelligence (XAI) consensus pipeline that integrates the transcriptomic prioritization scores from the meta-analysis (Driver_Score, N_Datasets, N_Domains, N_Pathways, Mean_NES) with four independent external biological databases: evolutionary constraint metrics from gnomAD v4.1, tissue-specific expression profiles from GTEx v8, protein-protein interaction network topology from BioGRID, and Mendelian disease associations from OMIM.

The integration of OMIM neurological disease annotations with sporadic ALS transcriptomic data is justified by the extensive molecular convergence between sporadic and familial ALS subtypes. *TDP-43* proteinopathy, the pathological hallmark of sporadic ALS, is caused by mutations in *TARDBP*, a Mendelian ALS gene[2]. The *C9orf72* hexanucleotide repeat expansion, initially identified in familial kindreds, accounts for 5-10% of apparently sporadic cases[3]. Furthermore, the OMIM features employed here capture general neurological disease involvement including epilepsy, ataxia, neuropathy and neurodegeneration rather than ALS specific annotations, serving as indicators of a gene's broader relevance to nervous system function and disease susceptibility.

Twelve multiplicative cross-database interaction features were engineered to capture synergistic biological signals not evident in any single data source (Table 1). For example, Hub_Neuro_Brain integrates PPI network centrality × neurological disease count × brain expression, testing the hypothesis that genes simultaneously occupying central network positions, associated with neurological disease and expressed in brain tissue are more likely to be ALS-relevant than genes possessing any single property in isolation.

Explainable AI Reveals Network Topology as the Determinant of Biomarker Utility

To overcome the limitations of traditional fold-change ranking, we trained a Random Forest classifier to distinguish established ALS-associated genes (Positive Class; n=66) from a background/unlabeled set (n=13,914). The model achieved strong discriminatory power with an Area Under the Curve (AUC) of 0.85 and an overall Out-of-Bag (OOB) error rate of 2.52% (Main Figure 6B). To assess whether model performance depend solely on rank-derived features, we trained a secondary "Biology-Only" model, excluding mathematical rank scores, which maintained an accuracy of 90.7%, indicating that non-rank features retained predictive information within the curated classification task. For comparability across metrics, the biology-only (tautology-defense) model achieved ROC-AUC = 0.83 and precision-recall AUC = 0.07, while the full multi-omic ensemble achieved ROC-AUC = 0.85 and precision-recall AUC = 0.10 (approximately 21-fold the class base rate). The full imbalance-aware metrics, feature-ablation, bias-control and external-validation analyses added in revision are provided in Supplementary Note 3 (Supplementary Tables S4-S17 and Supplementary Figure S16, S17). To deconstruct the decision making logic of the XAI, we employed SHAP (SHapley Additive exPlanations) and global feature importance analysis. By applying a stringent decision threshold (>80% confidence), the XAI identified a prioritized set of genes classified as non-significant by traditional statistics but possessing the high-dimensional architecture of Positive Class**.** Main figure 7A ranks the top 15 Positive Class genes (green) alongside the top 15 Consensus XAI Discoveries (red), ordered by Ensemble Percentile Rank (1.0 = 100th Percentile). Novel AI prioritized candidates such as *ATP1A1, YWHAG, RAF1, CRYAB,* and *STUB1* are ranked alongside confirmed ALS associated genes including *SQSTM1, NEFL, CHCHD10, SOD1*, and *UBQLN2,* all clearing the stringent confidence threshold. This intermingling indicates that the prioritized candidates occupy model-score ranges similar to those of positive-class genes; it does not establish mechanistic equivalence.

To examine the model features contributing to candidate prioritization, we performed SHAP force plot decomposition, comparing the decision paths of established ALS-associated genes (positive class) against our top Consensus Candidates (Supplementary Figure S11). For known associated genes like *SQSTM1* and *SOD1* (Probability ~0.999), the model's confidence was overwhelmingly driven by their high transcriptomic consensus score (Driver_Score) (contributing +0.45 to the probability) and Bridging_Potential (+0.20), indicating their systemic connectivity across multiple pathological domains. Remarkably, the top novel candidates displayed an similar explanatory profile. *ATP1A1* (probability score of 0.978) was prioritized almost exclusively due to its Driver_Score (+0.428) and Bridging_Potential (+0.195), mirroring the profile of *SQSTM1*. Similarly, *YWHAG* (probability score of 0.974)

was driven by a high Driver_Score (+0.45) and Bridging_Potential (+0.201), consistent with high network centrality in the input features. *RAF1* (probability score of 0.969) received additional model contributions from Cross_Domain_Impact (+0.057) and Bridging_Potential (+0.184), identifying it as a critical signaling node linking distinct pathological modules.

Multi-Source Feature Integration Reveals Strong Biological Discrimination Across 40 Features

Integration of four external biological databases (gnomAD, GTEx, BioGRID, OMIM) with the primary transcriptomic meta-analysis features yielded a final feature matrix of 40 features across 13,980 genes (Supplementary Figure S14). Univariate analysis revealed that 39 of 40 features (97.5%) showed statistically significant differences between Positive Class and Background genes (Wilcoxon rank-sum test, $p < 0.05$), with 35 features (85.4%) achieving significance at the Bonferroni-corrected threshold ($p < 0.05/40 = 1.25 \times 10^{-3}$) (Supplementary Figure S9A).

The most discriminating individual features were the cross-database interaction terms (Figure S9A), with Hub_Neuro_Brain ($p = 1.94 \times 10^{-74}$), Brain_Neuro ($p = 1.09 \times 10^{-72}$), and Brain_Neuro_Constrained ($p = 5.12 \times 10^{-71}$) achieving the strongest statistical separation. Notably, all ten top-ranked features by p-value were either OMIM-derived neurological disease features or cross-database interaction terms incorporating neurological disease information, confirming that the engineered interaction features capture disease-relevant biological signal beyond what is available from any single data source.

Among the primary database features, OMIM_N_Neuro ($p = 4.09 \times 10^{-70}$) and OMIM_Neuro_Ratio ($p = 8.68 \times 10^{-70}$) showed the strongest univariate discrimination, followed by PPI network features (PPI_Degree_Log, $p = 1.33 \times 10^{-9}$), brain expression features (Log_Brain_TPM, $p = 7.12 \times 10^{-9}$), and genetic constraint features (Missense_Z, $p = 4.73 \times 10^{-4}$). Positive Class exhibited significantly higher brain expression (median Brain_Max_TPM: 71.7 vs. 22.1 TPM), greater network connectivity (median PPI_Degree: 254 vs. 73 interactions), stronger mutational constraint (median LOEUF: 0.765 vs. 0.909), and more extensive neurological disease associations (median OMIM_N_Neuro: 1 vs. 0) compared to Background genes.

ALS Disease Feature Vector Is Dominated by Neurological Disease History, Not Expression or Connectivity

Standardized effect size analysis (Cohen's *d*) revealed that the molecular distinction between known ALS genes (n = 66) and background genes (n = 13,914) is dominated by neurological disease history and cross-database interaction features (Figure 6C, Table 1).

Critically, the standardized analysis corrected a misleading impression from raw (unstandardized) feature differences (Figure 7C). PPI_Degree dominated the raw vector (Δ = 340 interactions), exceeding OMIM_Neuro_Ratio (Δ = 0.456) by 745-fold. After standardization, PPI_Degree exhibited only a medium effect (*d* = 0.722) while OMIM_Neuro_Ratio emerged as the most discriminating feature (*d* = 1.267), demonstrating that these raw disparities reflect measurement scale differences rather than biological importance.

All twelve features exhibiting large effects ($d \geq 0.8$) were derived from OMIM disease annotations or cross-database interaction terms incorporating OMIM neurological information. The largest effect was observed for OMIM_Neuro_Ratio (*d* = 1.267), followed by Disease_Hub (*d* = 1.256), Brain_Neuro (*d* = 1.248), and Hub_Neuro_Brain (*d* = 1.209). Overall, 12 features exhibited large effects, 15 medium, 11 small, and 1 negligible. Of 39 features with non-zero variance, 38 were elevated in ALS genes. LOEUF was the sole negatively directed feature (*d* = −0.351), reflecting stronger evolutionary constraint in ALS genes (lower LOEUF = greater loss-of-function intolerance). Among the 12 large-effect features, 6 were OMIM-derived (OMIM_Neuro_Ratio, OMIM_N_Neuro, Log_OMIM_Phenotypes, OMIM_Has_Disease, OMIM_N_Phenotypes, OMIM_Disease_Pleiotropy) and 6 were cross-database interaction terms incorporating OMIM information (Disease_Hub, Brain_Neuro, Hub_Neuro_Brain, Neuro_Constrained, Brain_Neuro_Constrained, Disease_Burden). These features only reached the large effect threshold through multiplicative interaction with OMIM disease annotations; no single feature from the brain expression, PPI network, genetic constraint, or transcriptome data sources produced a large effect in isolation. This finding validates the cross-database feature engineering approach and reveals that the ALS molecular signature is defined not by any single biological property but by the convergence of neurological disease history with brain expression, network centrality, and evolutionary constraint.

The medium-effect features ($0.5 \leq d < 0.8$) included brain expression metrics (Log_Brain_TPM, *d* = 0.766; Brain_Specificity, *d* = 0.621), PPI network features (Is_Hub_Gene, *d* = 0.759; PPI_Degree_Log, *d* = 0.748; PPI_Degree, *d* = 0.722), and transcriptomic features (N_Datasets, *d* = 0.691; N_Domains, *d* = 0.615; Bridging_Potential, *d* = 0.610; N_Pathways, *d* = 0.546). The small-effect features included

genetic constraint metrics (Constraint_Combined, $d = 0.423$; Missense_Z, $d = 0.413$; pLI, $d = 0.351$), transcriptomic features (Driver_Score, $d = 0.488$; Pathway_Density, $d = 0.379$), and brain expression features (Brain_Expressed, $d = 0.495$; Brain_Mean_TPM, $d = 0.331$; Brain_Max_TPM, $d = 0.280$). Mean_NES was the only feature with a negligible effect ($d = 0.197$), suggesting that the magnitude of pathway enrichment scores contributes minimally to ALS gene discrimination compared to the number and diversity of enriched pathways.

This finding carries two important implications. First, it strongly supports the cross-database feature engineering approach: the six interaction terms among the top twelve features capture synergistic biological signals not evident in any single data source, suggesting that the multiplicative combination of features from independent databases generates discriminative power exceeding that of any individual feature. Second, it indicates that the molecular identity of an ALS-relevant gene is fundamentally associated with its neurological disease history and the convergence of that history with brain expression, network centrality, and evolutionary constraint, a finding with direct implications for the in silico drug reversal simulation and therapeutic target selection.

SUPPLEMENTARY TABLES

Supplementary Table S1. Number of dysregulated genes found in differential expression analyses of different datasets studied in this study.

GEO Accession ID	Domain (Tissue)	Dysregulated Genes
GSE33855	Mesenchymal (Fibroblasts)	22494 genes
GSE112676	Myeloid (Blood)	21580 genes
GSE68607	Lymphoid (LCL)	21754 genes
GSE28253	Lymphoid (Lymphocytes)	30234 genes
GSE122261	Mesenchymal (Muscle Cells)	17327 genes
GSE18920	CNS (Anterior Horn)	119 genes
GSE18920	CNS (Motor Neuron)	119 genes
GSE833	CNS (Motor Cortex)	5868 genes
GSE87385	CNS (Oligodendrocyte)	23348 genes
GSE87385	CNS (Astrocyte)	23348 genes
GSE112680	Myeloid (Blood)	21642 genes
GSE39642	Myeloid (Monocytes)	372 genes
GSE56808	Mesenchymal (Fibroblasts)	22880 genes

Supplementary Table S2: **Distribution of Datasets and Sample Composition Across Biological Domains.**

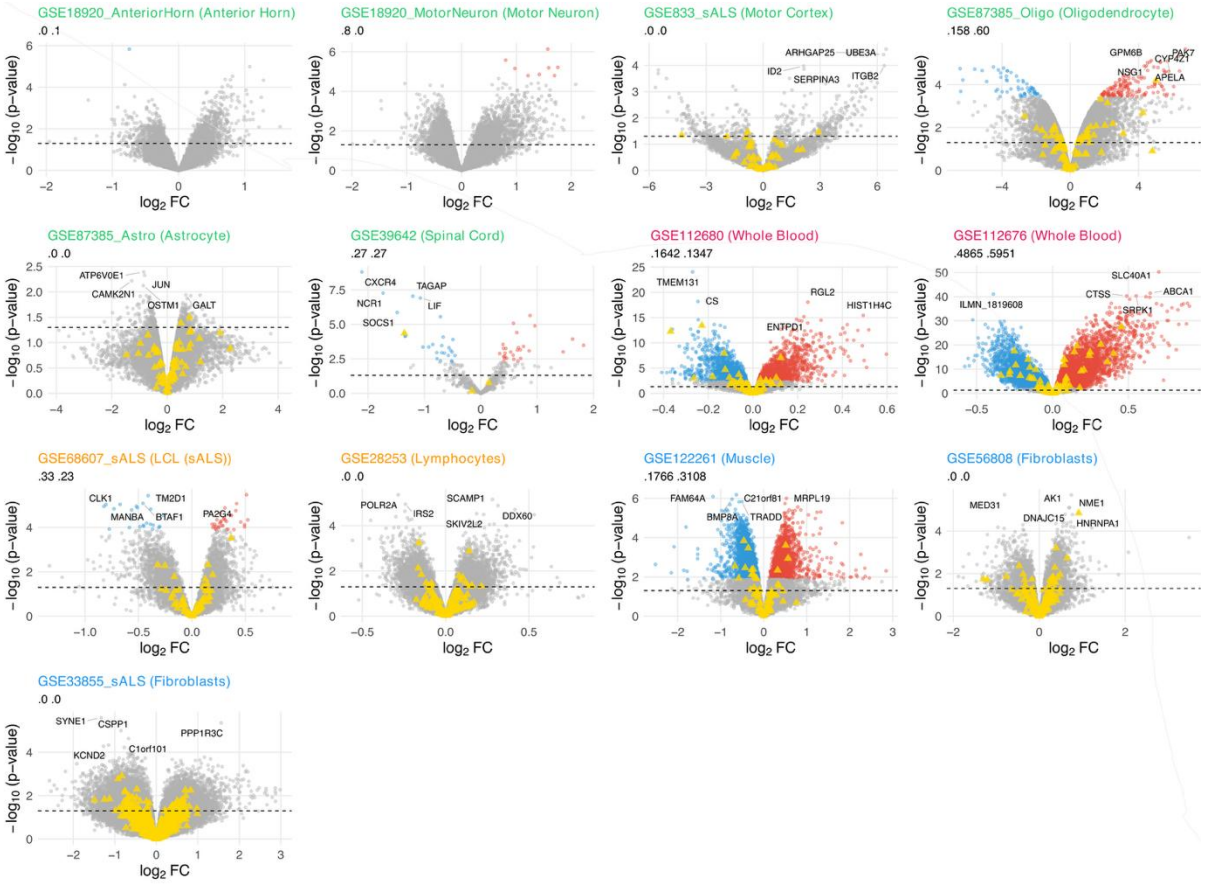
Domain	Datasets (n)	Total Samples	Tissues
CNS	5	61	Anterior Horn, Motor Neuron, Motor Cortex
Lymphoid	2	60	LCL, Lymphocytes
Mesenchymal	3	44	Fibroblasts, Muscle Cells
Myeloid	3	1062	Blood, Monocytes

Supplementary Table S3: Evidence Profiles of the top six candidates.

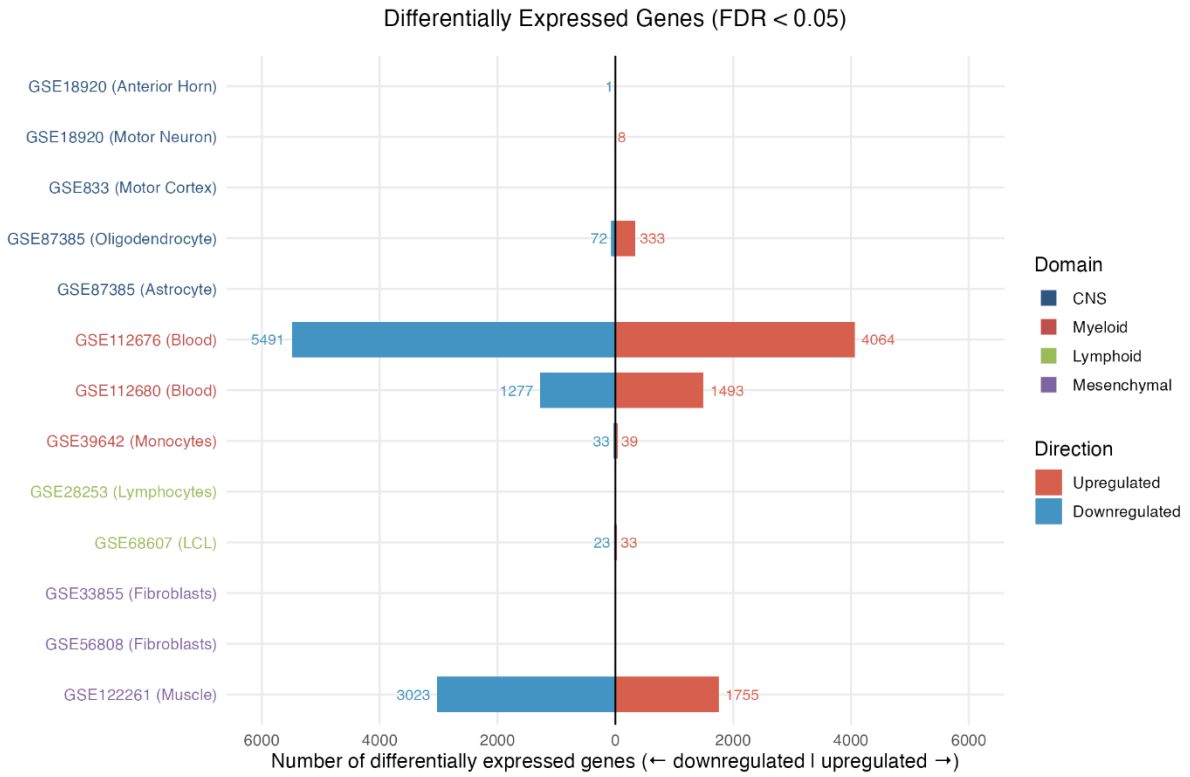
Gene Symbol	Composite Score	Primary Biological Function	Known ALS Interactors (1st-Degree)	Mechanistic Rationale & Translational Evidence
ATP1A1	0.873	Catalytic subunit of the Na^+/K^+-ATPase; maintains neuronal electrochemical gradients.	*SQSTM1, VCP, UBQLN2, TNIP1, CCNF, C9orf72* (8 total)	Extreme evolutionary constraint (LOEUF = 0.188). Lies at the intersection of autophagy, innate immunity, and repeat expansion axes. Highly druggable ion channel with established pharmacological modulators (cardiac glycosides)[4,5].
YWHAG	0.844	14-3-3γ isoform; regulates signal transduction, apoptosis, and protein trafficking.	*NEFL, KIF5A, SQSTM1, VCP, OPTN, NEK1* (13 total)	High constraint (pLI = 0.895) and profound brain expression (554.5 TPM). Reported as elevated in ALS patient cerebrospinal fluid (CSF), providing direct clinical relevance[6].
ANLN	0.825	Actin-binding protein; essential for cytokinesis and cytoskeletal organization.	*TARDBP, FUS, HNRNPA1, HNRNPA2B1, EWSR1, NEFL, TUBA4A, PFN1, PRPH, GFAP* (17 total)	Functions as a critical molecular integrator. Its dense interaction network directly bridges the RNA processing and cytoskeletal dysfunction axes identified in ALS pathogenesis[7].
RAF1	0.818	RAF proto-oncogene serine/threonine kinase.	*SQSTM1, VCP, UBQLN2, TARDBP, EWSR1, TAF15, HNRNPA2B1,*	Membership in the kinase druggable family; Exhibits the absolute highest ALS network connectivity among all candidates. Interacts with targets

			TUBA4A, PFN1, SETX, TBK1, CCNF, TNIP1, GRN, ATXN2, ANXA11	spanning virtually every central disease pathway[8].
STUB1	0.781	CHIP E3 ubiquitin ligase	*VCP, SOD1, SQSTM1, DNAJC7, PFN1, PRPH, GSK3B*	Target misfolded proteins for proteasomal degradation directly connects to the proteostatic stress signature identified in the myeloid domain and the proteasome subunit *PSMB2* prioritized in the transcriptomic analysis. Two OMIM neurological disease associations (spinocerebellar ataxia and Gordon Holmes syndrome) confirm its relevance to neurodegeneration[9,10].
FXR1	0.759	Fragile X-related protein 1	*EWSR1, SMN1, C9orf72*, *DCTN, HNRNP*	Its strong constraint (LOEUF = 0.374, pLI = 0.807) and two neurological disease associations position *FXR1* within the RNA processing dysregulation axis[11,12].

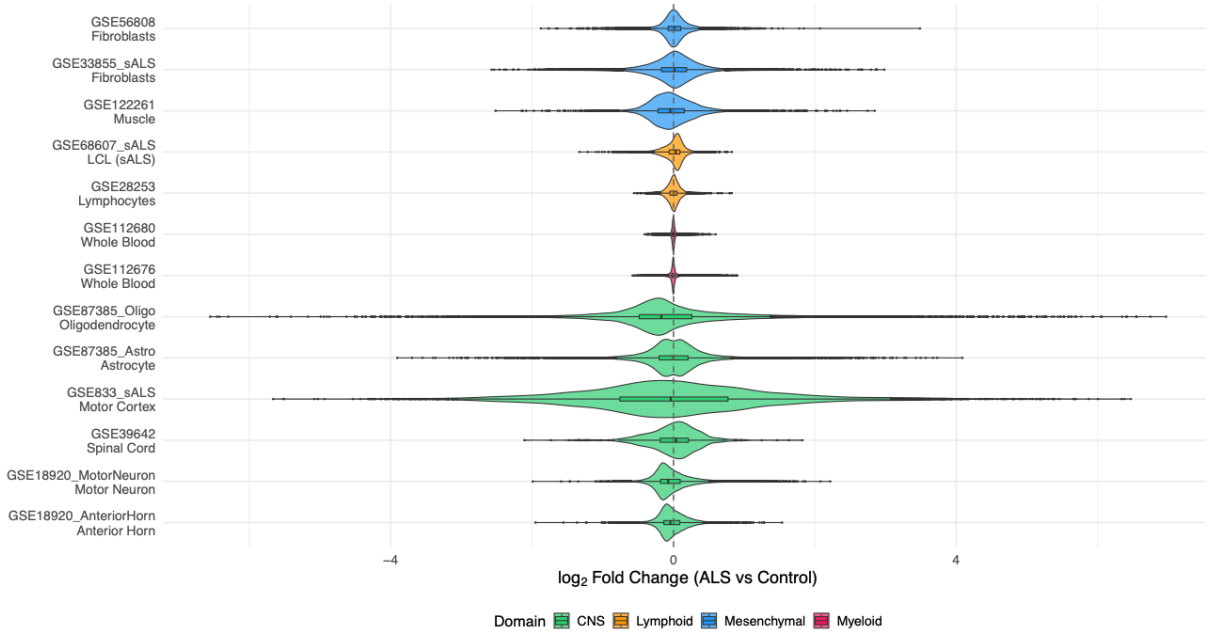
SUPPLEMENTARY FIGURES



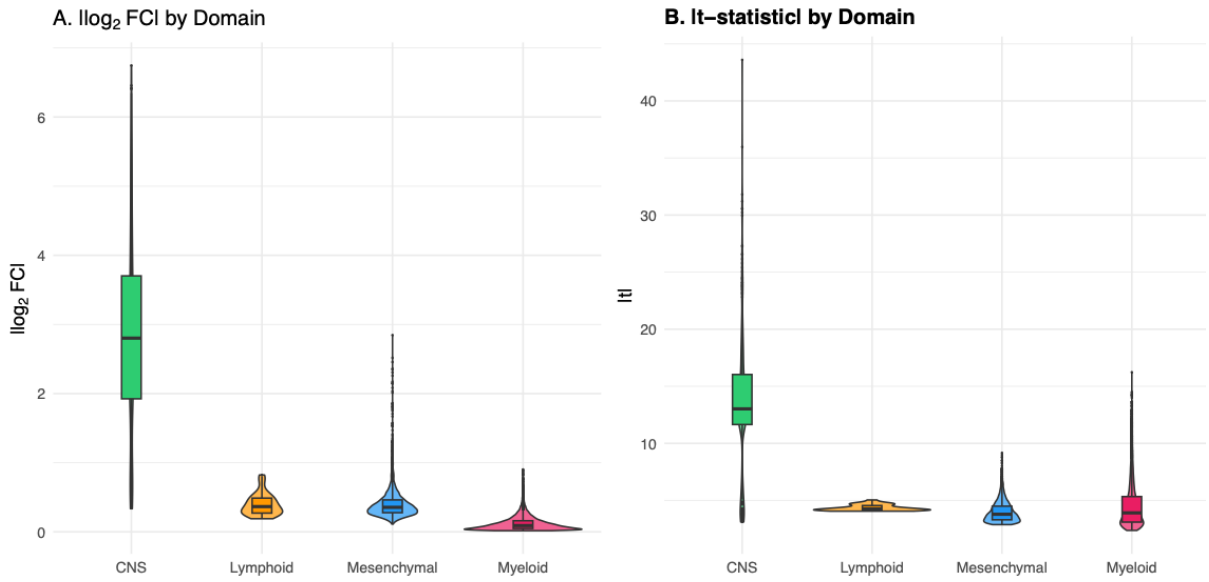
Supplementary Figure S1. Volcano plots for all 13 ALS transcriptomic datasets. Red: significantly upregulated (FDR < 0.05); blue: significantly downregulated; gray: non-significant. Gold triangles indicate genes from the curated ALS-gene positive set. Top 5 genes by p-value are labelled. Panel titles colored by biological domain; subtitles indicate the number of up and down-regulated significant genes.



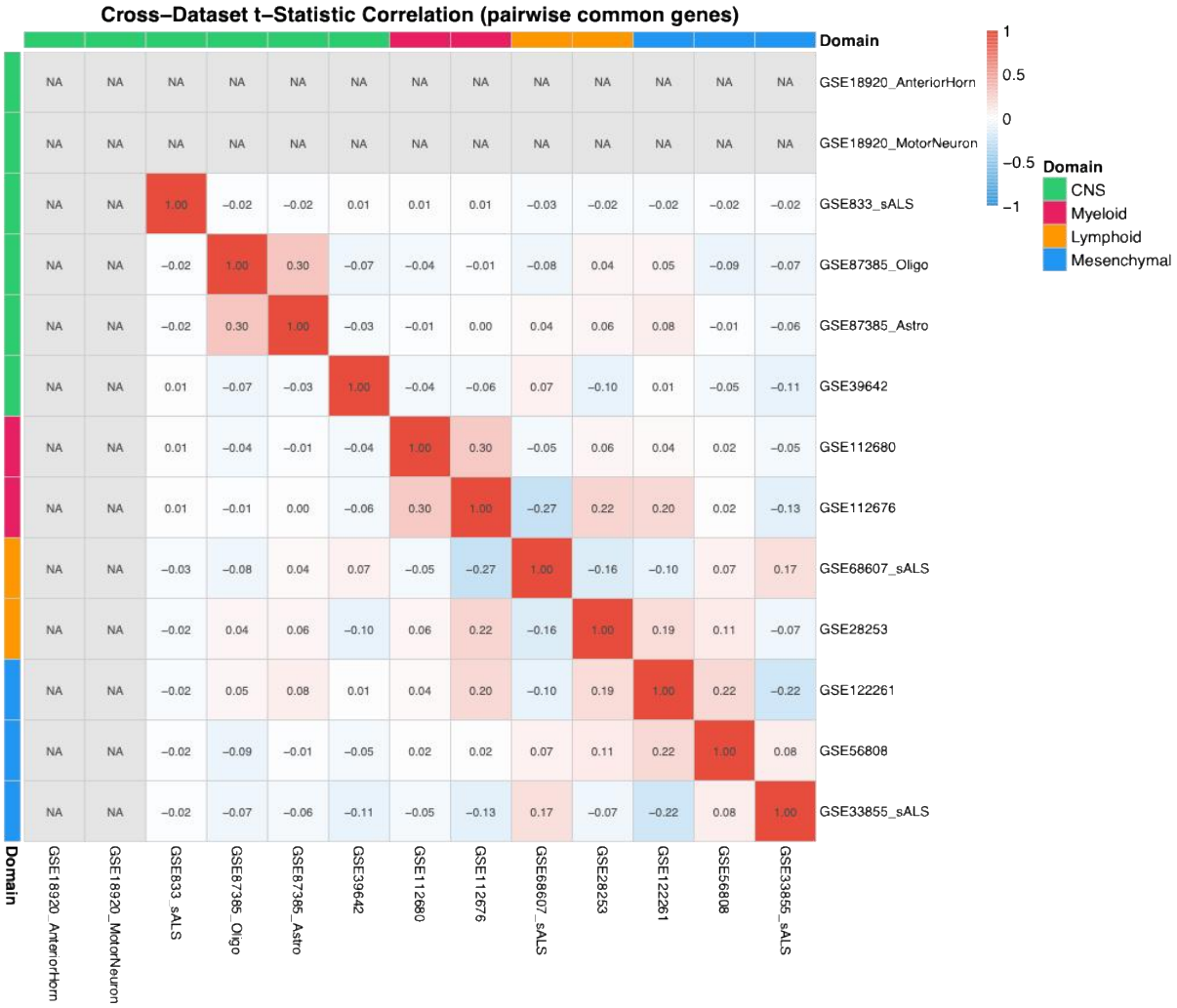
Supplementary Figure S2. Differentially expressed gene counts across all datasets. Horizontal bar plot showing the number of significantly upregulated (red) and downregulated (blue) genes at FDR < 0.05 for each dataset, ordered by biological domain.



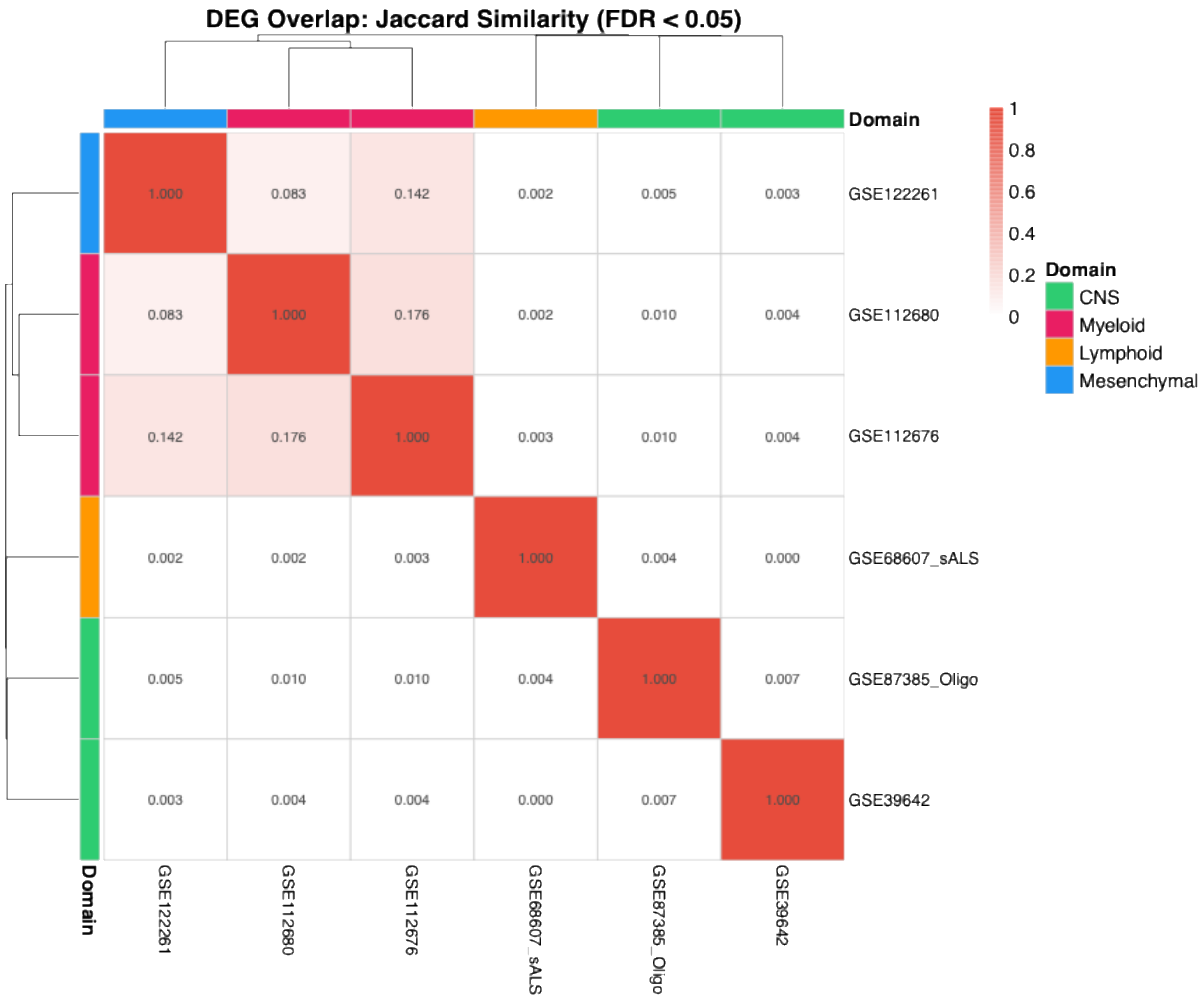
Supplementary Figure S3. Distribution of log2 fold changes across all datasets. Violin plots with embedded boxplots showing the distribution of effect sizes, colored by biological domain.



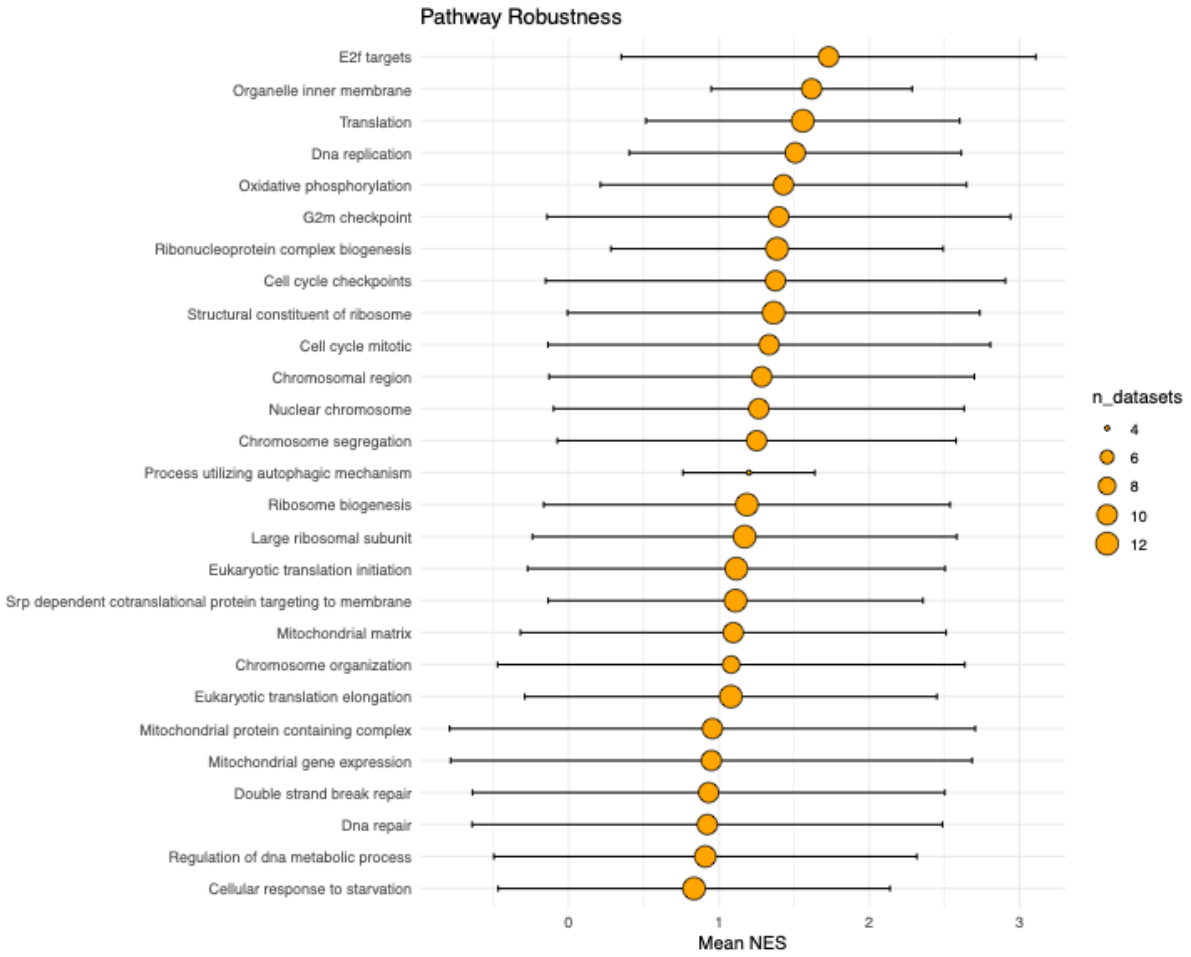
Supplementary Figure S4: Effect size distributions by biological domain. (A) Distribution of absolute log2 fold changes among significant DEGs (FDR < 0.05) across the four biological domains. (B) Distribution of absolute t-statistics among significant DEGs. Violin plots with embedded boxplots reveal domain-specific differences in effect size magnitude.



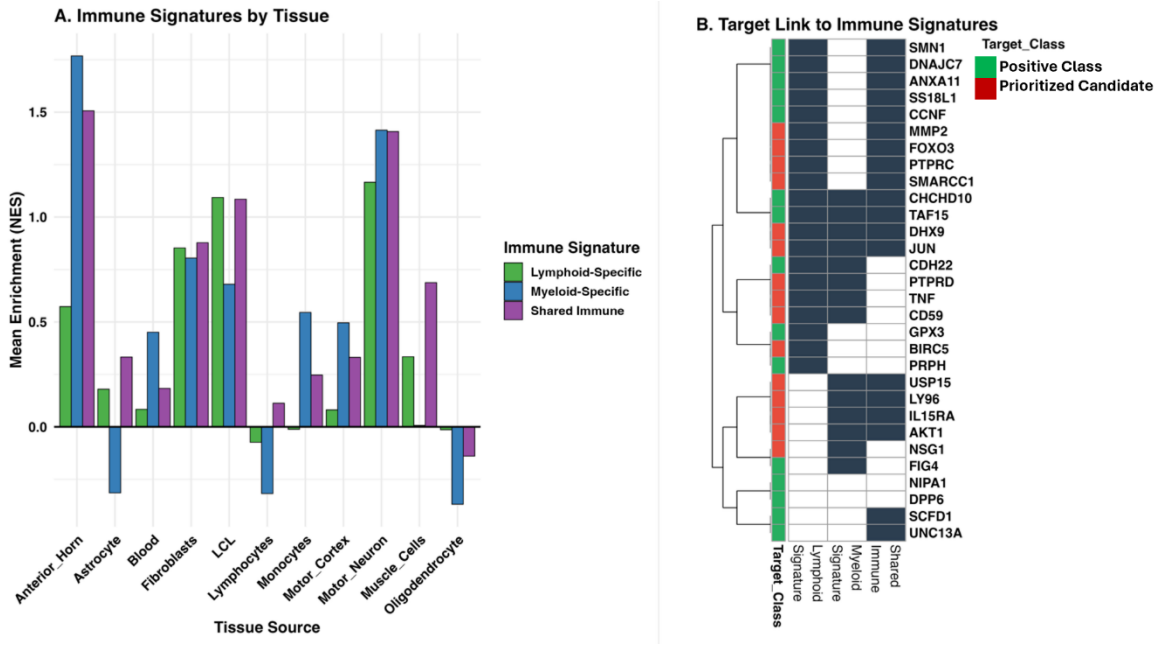
Supplementary Figure S5. Cross-dataset transcriptomic concordance. Pairwise Pearson correlation of gene-level t-statistics across all datasets. Higher correlation indicates greater transcriptomic concordance between datasets. Annotations indicate biological domain and tissue type. Datasets within the same domain generally show higher correlation, supporting the domain-stratified meta-analysis approach.



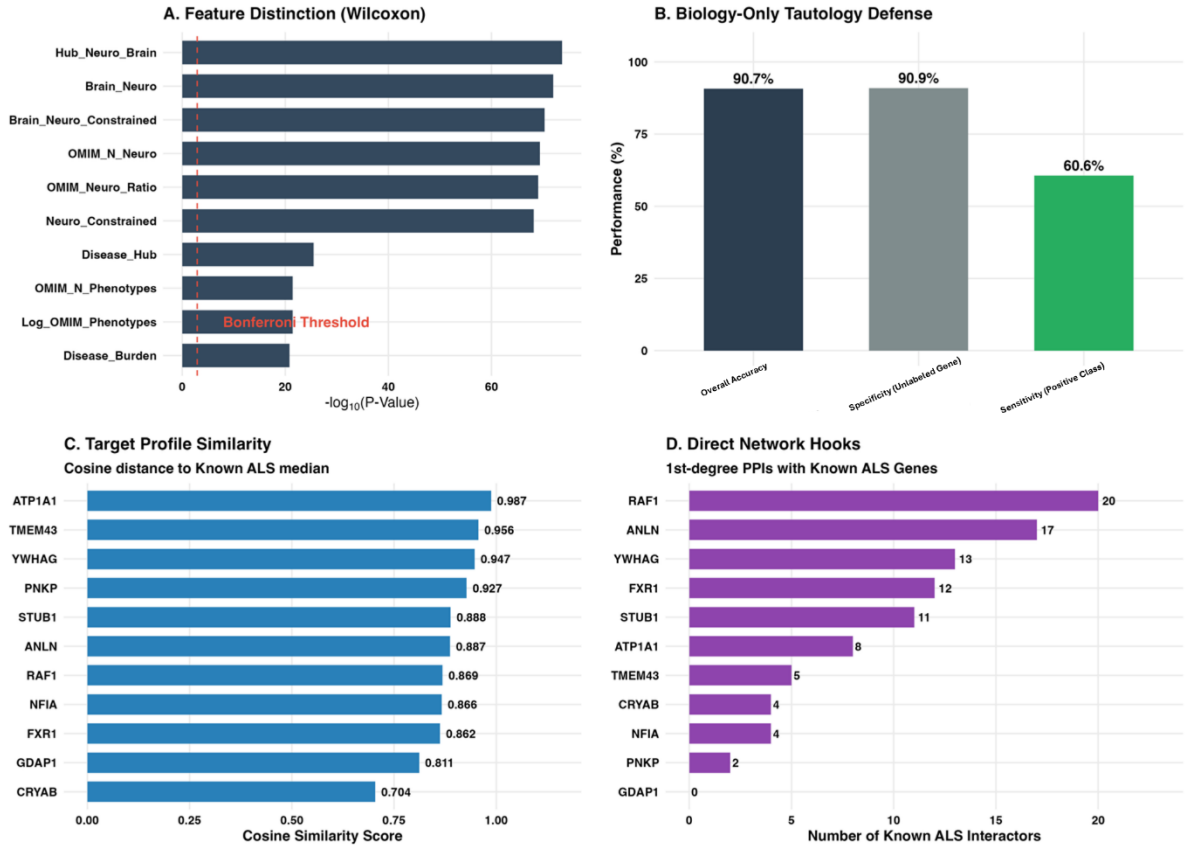
Supplementary Figure S6. Gene overlap across datasets. Jaccard similarity heatmap showing pairwise overlap coefficients between datasets.



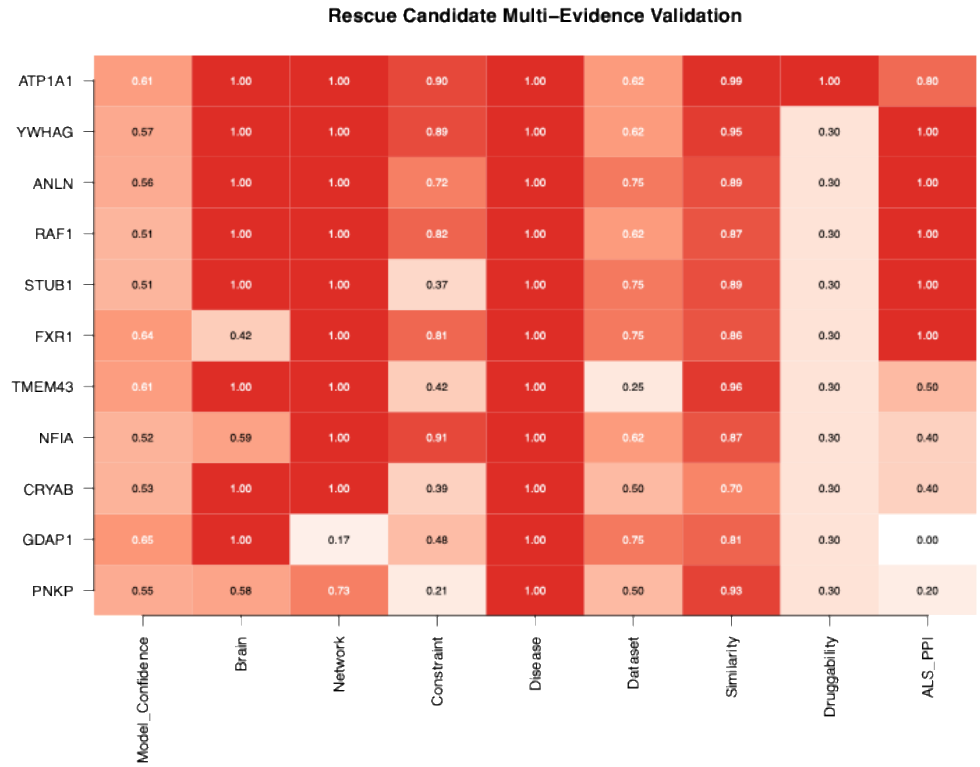
Supplementary Figure S7. **Robustness and consistency of pathway enrichment across ALS datasets studied**. Forest plot depicting the stability and magnitude of dysregulated biological pathways across the meta-analysis. The position of each central node on the x-axis indicates the Mean Normalized Enrichment Score (Mean NES) across all datasets where the pathway was detected. Horizontal error bars denote the standard deviation of the NES across these distinct datasets. The size of the node is proportional to the number of independent datasets (n_datasets) in which the pathway reached statistical significance, ranging from 4 to 12 datasets.



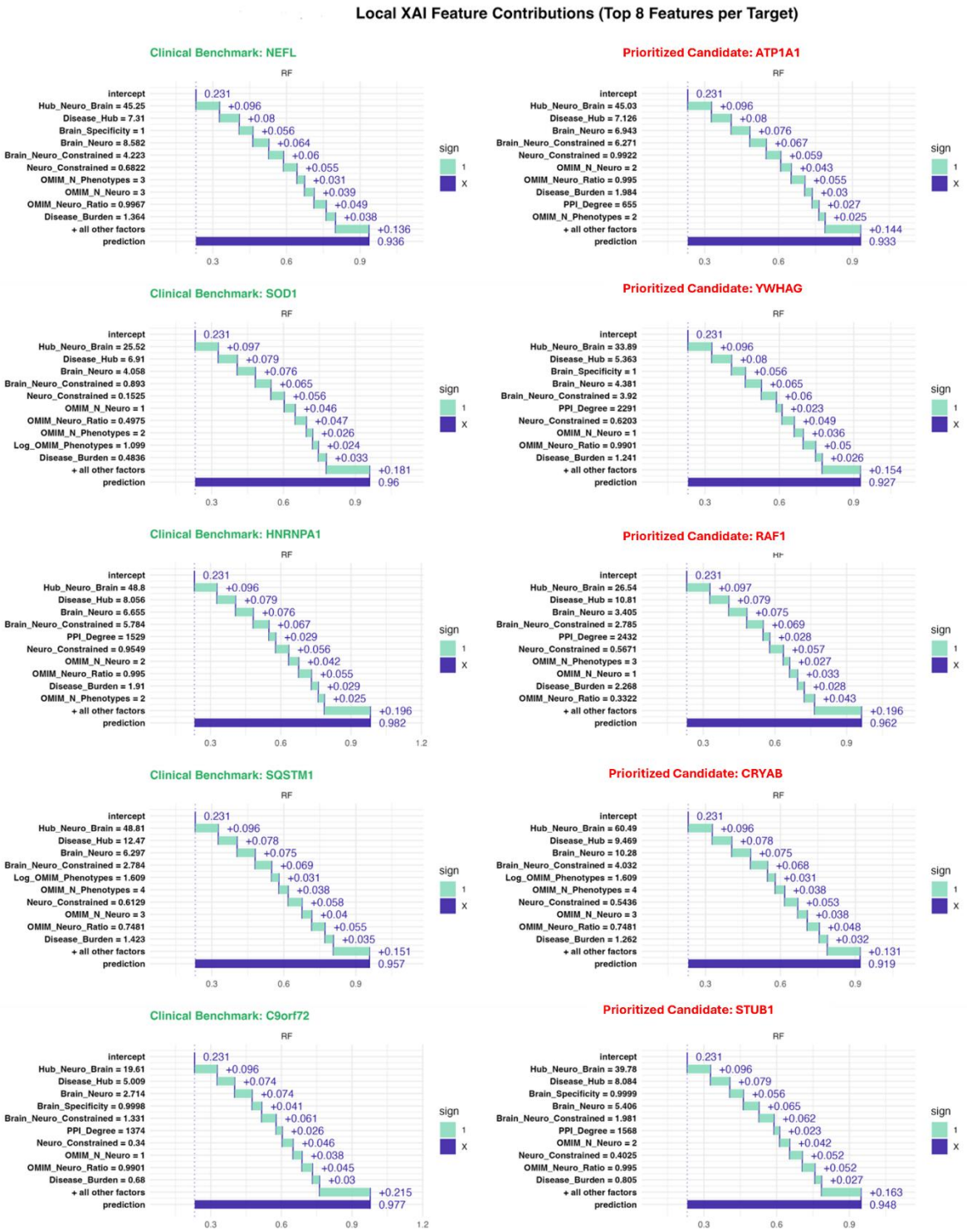
Supplementary Figure S8. Spatial resolution and molecular validation of ALS immune signatures. (A) Immune Signatures by Tissue: A bar chart displaying the Mean Normalized Enrichment Scores (NES) for Lymphoid-Specific (green), Myeloid-Specific (blue), and Shared Immune (purple) signatures across various anatomical and cellular sources. (B) Link to Immune Signatures: A binary heatmap mapping specific genes to the identified immune signatures. Dark blue cells indicate gene membership within a specific signature (Lymphoid, Myeloid, or Shared).



Supplementary Figure S9. (A) **Global feature distinction**. The top 10 most discriminating features between known positive class genes and background genes, ranked by univariate Wilcoxon rank-sum significance. All top features survive rigorous Bonferroni correction for multiple hypothesis testing. (B) **Biology-Only Tautology Defense** (C) **Target profile similarity**. Cosine similarity scores mapping the 40-dimensional multi-omic feature vectors of the 11 novel prioritized candidates against the median profile of the positive-class genes. Top candidate ATP1A1 exhibits high model-feature-profile similarity (Cosine = 0.987). (D) **Direct network hooks.** The number of curated first-degree BioGRID protein-protein interactions connecting the novel prioritized candidates directly to clinically established ALS disease associated genes.

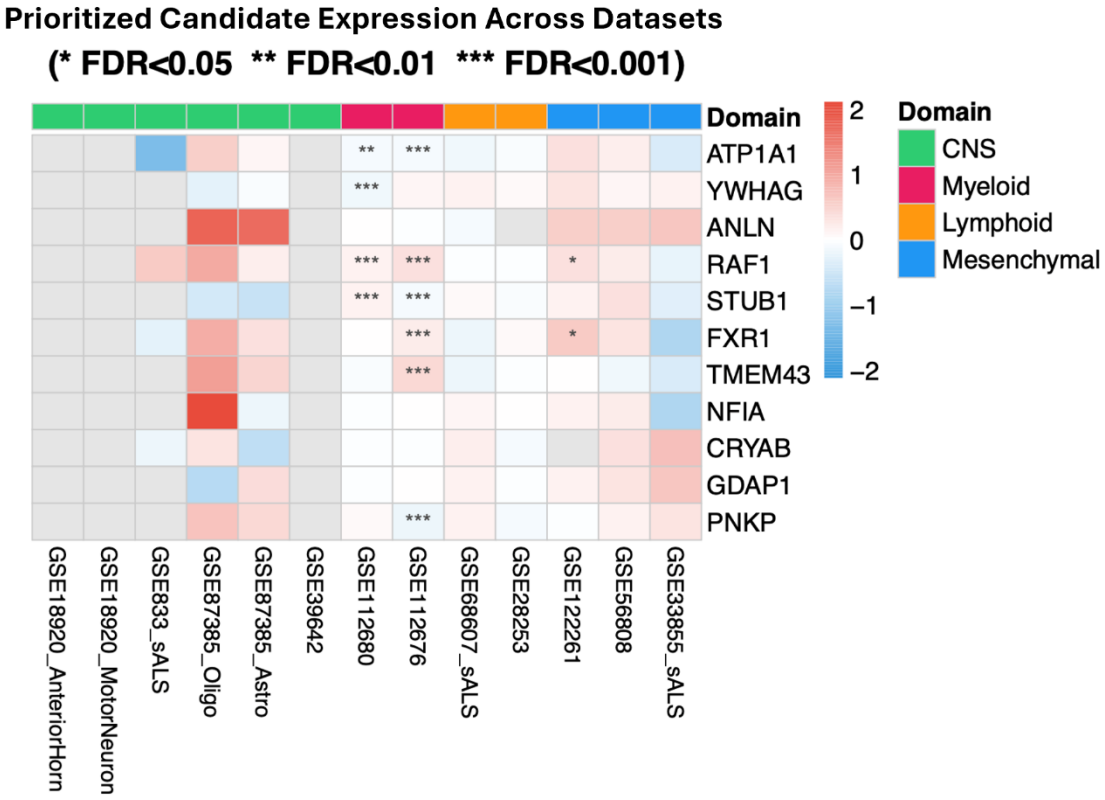


Supplementary Figure S10: Multi-Evidence Scoring Heatmap for Top Prioritized Candidates. The heatmap displays normalized evidence scores (0.00–1.00) for the top 11 Prioritized Candidates (rows) across nine evidence components (columns), where darker red indicates stronger evidence support and lighter/white indicates weaker support.

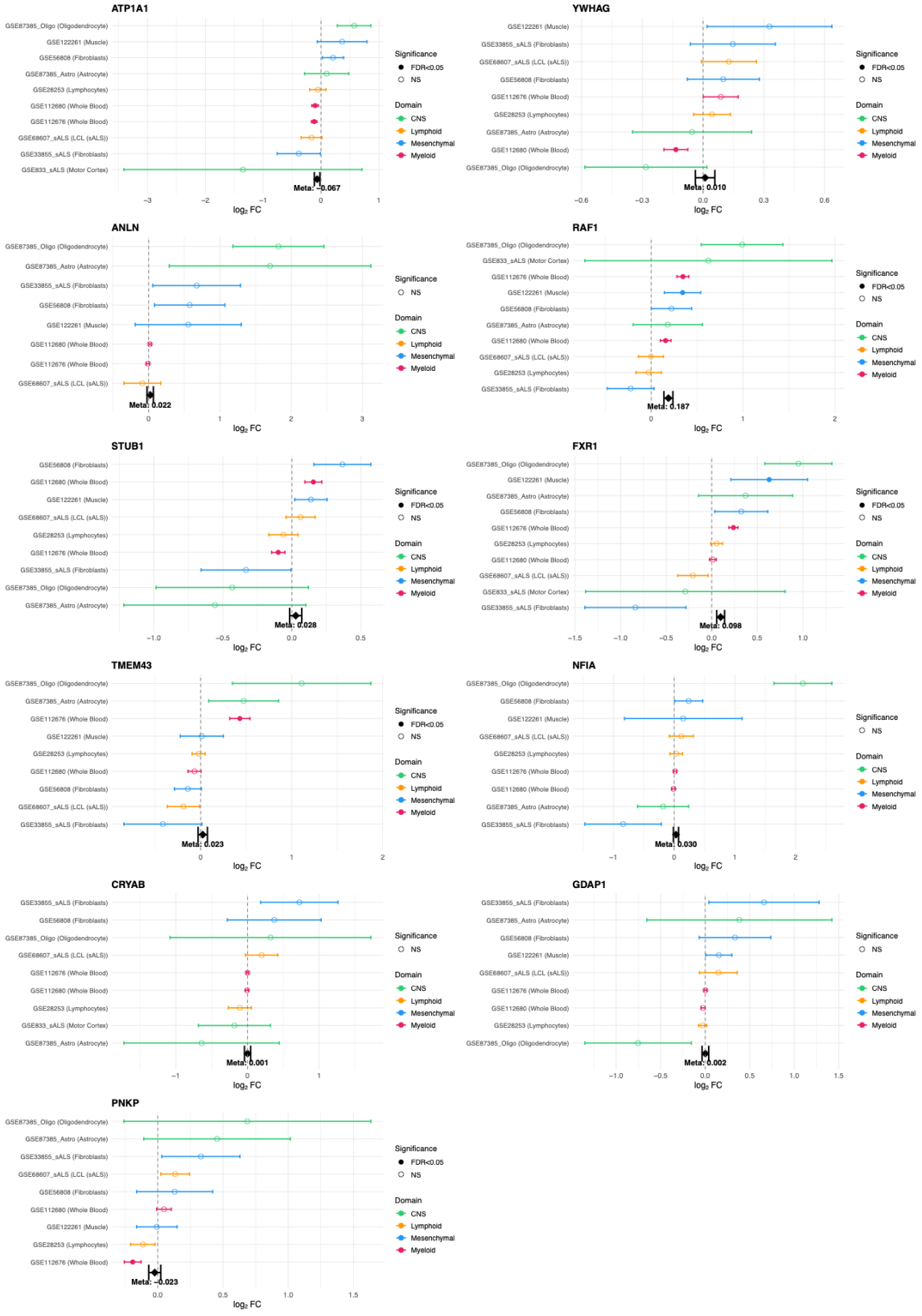


Supplementary Figure S11. Local SHAP (Shapley Additive exPlanations) waterfall plots isolating the top 8 individual feature contributions to the algorithmic prediction for the top 5 positive-class benchmarks (left column, green) and the top 5 computationally prioritized discoveries (right column, red). Baseline probability (intercept) indicates the background risk

of the dataset, while sequential bars demonstrate how specific multi-omic feature values push the algorithm's final probability toward 1.0 (positive-class classification).

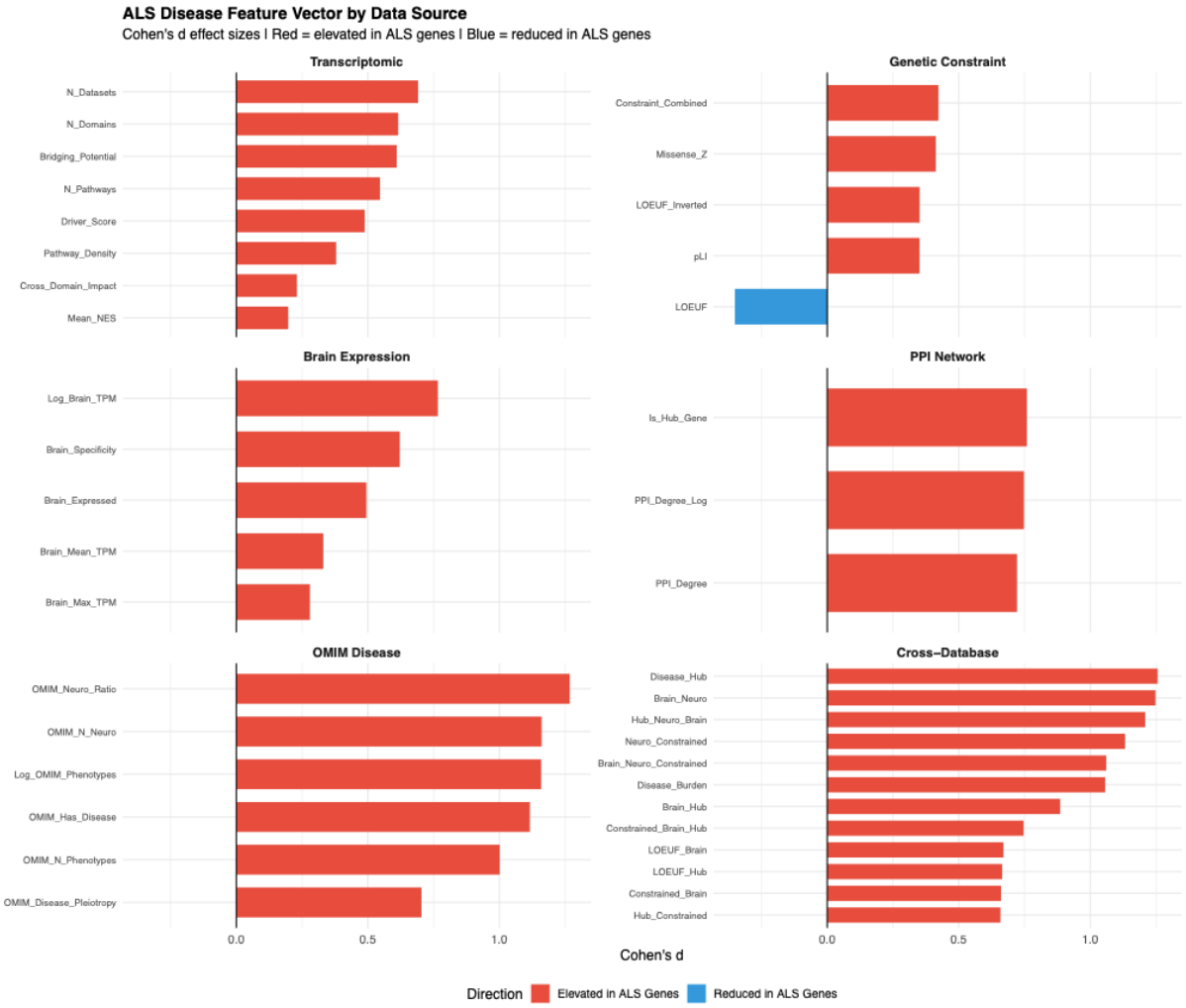


Supplementary Figure S12: Biological domain expression heatmap. Expression (log2 fold change) patterns of prioritized candidates across datasets.

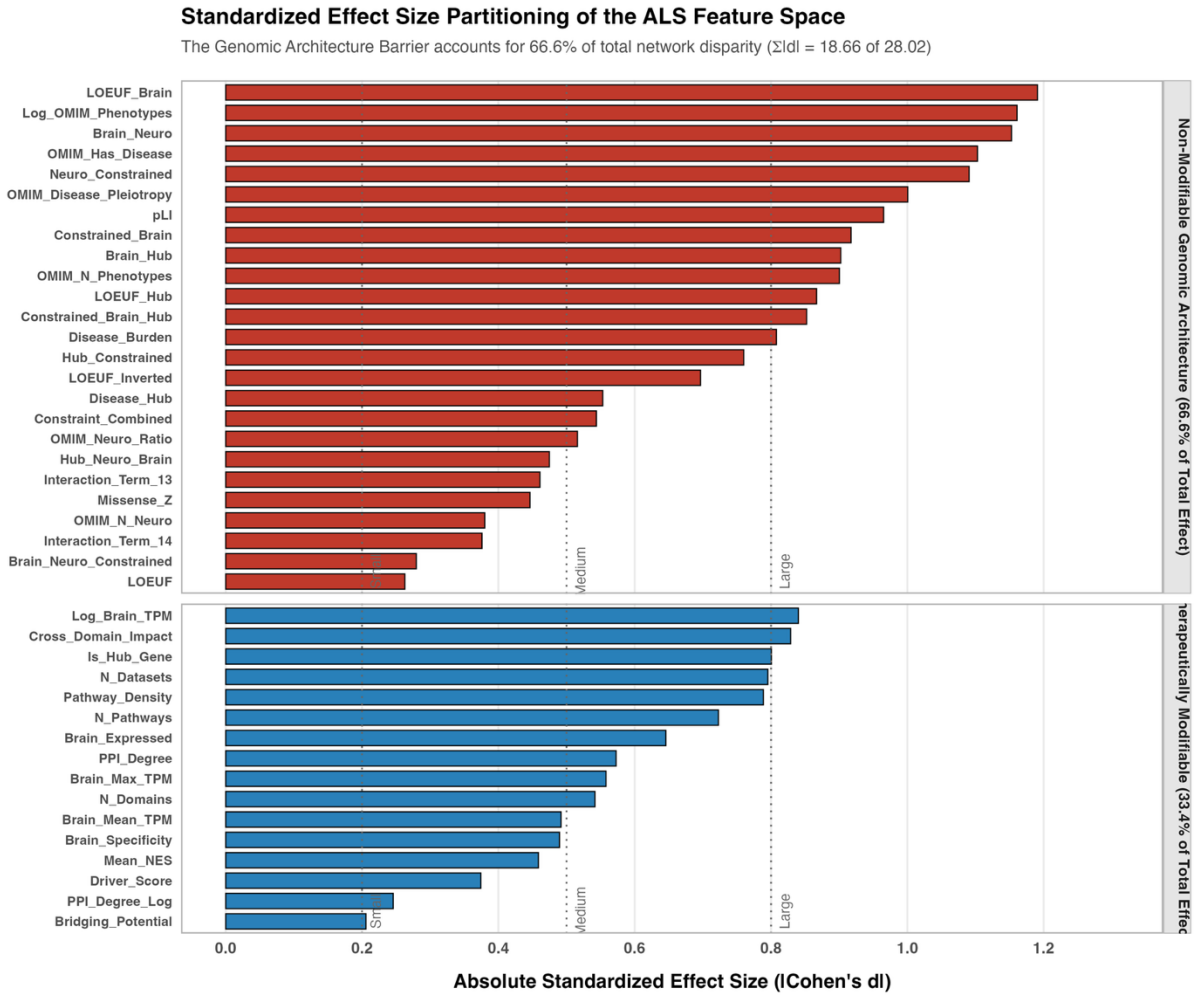


Supplementary Figure S13: Forest plots for prioritized candidates. Individual study effect sizes (log2 fold change ± 95% CI) for each of the 11 prioritized candidates across all datasets in which they were detected. Filled circles indicate FDR < 0.05; open circles indicate non-

significant. Black diamond represents the inverse-variance weighted meta-effect. Colors indicate biological domain.



Supplementary Figure S14. ALS disease feature vector characterizing known ALS genes relative to background. Feature effects stratified by data source category (Transcriptomic, Brain Expression, OMIM Disease, Genetic Constraint, PPI Network, and Cross-Database). Red bars indicate features elevated in ALS genes; blue bars indicate features reduced in ALS genes.



Supplementary Figure S15. **Standardized effect size partitioning of the ALS feature landscape.** Absolute Cohen's *d* values (|*d*|) quantifying the standardized mean difference between Positive class and Background genes for all 40 predictive features. Features are grouped into two prespecified categories used in the perturbation simulation. Red bars indicate non-modifiable genomic architecture variables (e.g., evolutionary constraint, historical pleiotropy), which collectively account for 66.6% of the total systemic disparity ($\sum|\mathrm{d}| = 18.66$). Blue bars indicate therapeutically modifiable variables (e.g., transcriptomic expression, pathway enrichment), which account for only 33.4% of the total disparity ($\sum|\mathrm{d}| = 9.36$). Vertical dotted lines denote standard statistical thresholds for small (0.2), medium (0.5), and large (0.8) effect sizes. The disproportionate magnitude of non-modifiable features illustrates the larger standardized separation assigned to the non-modifiable category within this prespecified model-based grouping.

Supplementary Note 3: Additional analyses added in revision: model evaluation under class imbalance, feature ablation, baseline benchmarking, bias controls, external validation, architecture generalization, and gene-set definition

Molecular pathway assignment of the eleven prioritized candidates

The eleven candidates map onto four molecular axes with established relevance to ALS pathophysiology: ion homeostasis and membrane excitability, protein quality control and proteostasis, RNA processing and transport, and signal transduction together with additional transcriptional-regulation, nuclear-envelope, mitochondrial and DNA-repair functions. All but one physically interact with established ALS genes in BioGRID.

Table S4. Functional axis, priority tier and established-ALS-gene interactors of the eleven candidates.

Gene	Tier	ALS-relevant functional axis	ALS-gene interactors (n)	Representative ALS-gene interactors
ATP1A1	Tier 1	Ion homeostasis / membrane excitability (druggable ion-channel family)	8	SQSTM1, VCP, C9orf72, SOD1, UBQLN2
RAF1	Tier 1	Kinase signal transduction	20	TBK1, SOD1, TARDBP, VCP, C9orf72
ANLN	Tier 1	Cytoskeleton / RNA-processing network	17	TARDBP, FUS, HNRNPA1, MATR3, ATXN2
YWHAG	Tier 1	Signal transduction (14-3-3 scaffolding)	13	NEFL, KIF5A, SOD1, OPTN, C9orf72
FXR1	Tier 1	RNA processing and transport	12	ATXN2, EWSR1, SMN1, GLE1, C9orf72
STUB1	Tier 1	Protein quality control / proteostasis	11	VCP, SQSTM1, C9orf72, DNAJC7, HNRNPA1
CRYAB	Tier 1	Proteostasis (small heat-shock chaperone)	4	GFAP, NEFL, SQSTM1, SOD1

NFIA	Tier 1	Transcriptional regulation	4	ATXN1, MATR3, ATXN2, SS18L1
PNKP	Tier 1	DNA-damage repair	2	TNIP1, SQSTM1
TMEM43	Tier 2	Nuclear envelope / membrane	5	UBQLN2, VAPB, CCNF, TNIP1
GDAP1	Tier 2	Mitochondrial dynamics	0	—

Independent Validation

We assessed the eleven prioritized candidates in two ALS RNA-seq cohorts that were not used for model training or feature construction: a post-mortem prefrontal-cortex cohort (GSE314526; affected CNS tissue; primary) and a peripheral-blood cohort (GSE234297; accessible tissue; secondary). Both use RNA-seq, providing a cross-platform check relative to the microarray meta-analysis. Differential expression was assessed with limma-voom and empirical-Bayes moderation (TMM normalisation; genes filtered to CPM ≥ 1 in at least the control-group size).

1. Primary validation: prefrontal cortex (GSE314526)

GSE314526 (NY Genome Center) comprises bulk RNA-seq of 72 prefrontal-cortex samples (Brodmann areas BA44/45 and BA46) from 68 donors: ALS (n=28), ALS with cognitive impairment (ALSci, n=23) and non-neuropathological sudden-death controls (n=21). ALS+ALSci (n=51) were compared with controls (n=21) using limma-voom with empirical-Bayes moderation, adjusting for brain region, sex, age and post-mortem interval (19,303 expressed genes).

All eleven candidates were expressed in cortex. One replicated at FDR < 0.05, TMEM43 (log2FC = +0.24; FDR = 1.6×10^{-2}) was also the only candidate significant at nominal $p < 0.05$. Signed-direction concordance with the microarray meta-analysis was 6/11. ALS prefrontal cortex is broadly and diffusely dysregulated (723 genes at FDR < 0.05; ~22% of genes nominally DE), and bulk differential expression has limited per-gene power: the 66 established ALS genes themselves reach FDR < 0.05 for only 2/65, and neither the candidates nor the known ALS genes form a statistically enriched set (Fisher $p \approx 0.3$). The candidates thus replicate at a rate comparable to bona fide ALS genes.

Table S5. The eleven candidates in GSE314526 prefrontal cortex (ALS+ALSci vs control; limma-voom + eBayes, covariate-adjusted).

Gene	log2FC	p	FDR	Direction	DE
TMEM43	0.236	2.0e-4	0.016	up	FDR<0.05
ANLN	-0.477	0.076	0.281	down	ns
NFIA	0.184	0.101	0.319	up	ns
FXR1	-0.107	0.159	0.394	down	ns
RAF1	0.063	0.166	0.402	up	ns
GDAP1	-0.150	0.198	0.440	down	ns
STUB1	0.198	0.273	0.517	up	ns
PNKP	0.106	0.336	0.573	up	ns
YWHAG	-0.111	0.511	0.711	down	ns
CRYAB	-0.117	0.631	0.796	down	ns
ATP1A1	-0.083	0.668	0.820	down	ns

2. Secondary validation: peripheral blood (GSE234297)

GSE234297[13] is whole-blood RNA-seq of 96 sporadic ALS and 48 controls. Eight of eleven candidates were expressed above filter (ANLN, CRYAB and GDAP1 were not). No candidate reached FDR < 0.05; ATP1A1 (log2FC = −0.11; p = 0.014) and NFIA (p = 0.045) were nominally significant. Consistent with cortex, the 66 established ALS genes are likewise weakly replicated in blood (1/53 at FDR < 0.05), so the candidates are statistically indistinguishable from bona fide ALS genes in this accessible tissue.

Table S6. The eleven candidates in GSE234297 whole blood (sALS vs control; limma-voom + eBayes).

Gene	log2FC	p	FDR	Direction	DE
ATP1A1	-0.108	0.014	0.118	down	p<0.05
NFIA	0.264	0.045	0.197	up	p<0.05
STUB1	-0.149	0.065	0.235	down	ns
TMEM43	0.119	0.085	0.269	up	ns
RAF1	0.111	0.106	0.301	up	ns
PNKP	-0.101	0.219	0.451	down	ns
FXR1	0.073	0.458	0.674	up	ns
YWHAG	0.002	0.977	0.989	up	ns

ANLN / CRYAB / GDAP1	not detected	—	—	—	—

Architecture predicts ALS relevance in independent gene sets, beyond the classifier

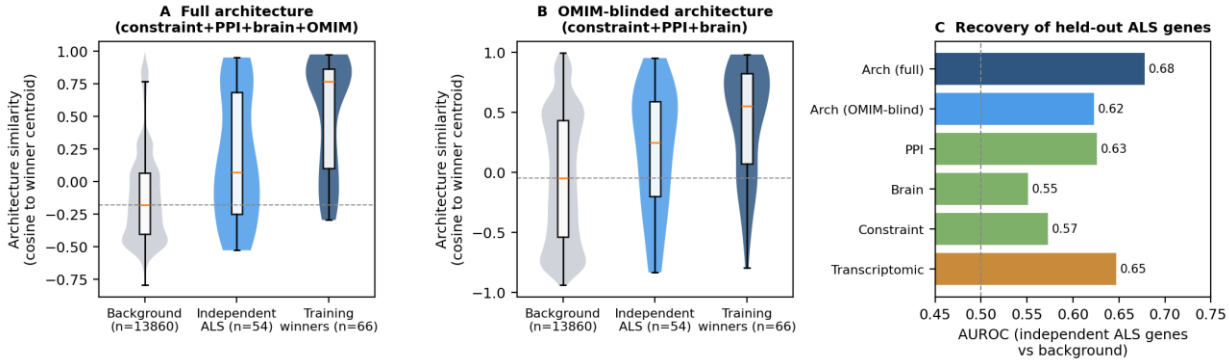
To test whether the genomic-architecture axis underlying the "Genomic Architecture Barrier" reflects a transferable property of ALS genes rather than the geometry of the trained classifier (Reviewer 1.2), we scored every gene by a *model-free* architecture score, the cosine similarity of its standardized non-transcriptomic profile (evolutionary constraint, protein-interaction topology, brain-enriched expression, neurological-disease annotation) to the mean profile (centroid) of the 66 Positive Class. No machine-learning classifier is used. We then tested recovery of an independent set of 54 ALS-associated genes obtained from a DisGeNET curation (UMLS C0002736) that were NOT among the 66 training positives.

These held-out ALS genes are recovered well above chance (AUROC = 0.68; median 75th percentile; Mann-Whitney $p = 3\times10^{-6}$). Crucially, recovery persists when all OMIM disease-annotation features are removed (constraint + network + brain only: AUROC = 0.62; $p = 9\times10^{-4}$), showing the signal is not merely re-reading disease databases. Recently reported associations (year ≥ 2015) are recovered more strongly still (AUROC ≈ 0.70–0.74). Architecture therefore encodes a genuine, transferable signature of ALS relevance that generalises to genes the model never saw.

Table S7. Recovery of 54 independent (held-out) ALS genes by a model-free architecture score.

Feature set	**AUROC (independent)**	**Median percentile**	**AUROC (recent, year≥2015)**
Architecture, full (constraint+PPI+brain+OMIM)	0.68	0.75	0.74
Architecture, OMIM-blinded (constraint+PPI+brain)	0.62	0.64	0.70
PPI topology only	0.63	0.63	0.69
Evolutionary constraint only	0.57	0.61	0.61
Brain expression only	0.55	0.53	0.62

Transcriptomic (differential expression) only	0.65	0.68	0.71



Supplementary Figure S16. (A) Model-free architecture-similarity score (cosine to the Positive Class centroid) for background genes, 54 independent held-out ALS genes, and the 66 training positive-class genes: independent curated ALS genes are shifted toward the positive class genes. (B) The same with OMIM annotation removed (constraint + network + brain only). (C) AUROC for recovery of the independent ALS genes across feature sets; dashed line = chance.

Under a properly moderated model the external transcriptomic signal is modest: one candidate (TMEM43) replicates at FDR < 0.05 in affected cortex, and the others are not individually significant. The decisive context is that the 66 established ALS genes replicate just as weakly in these cohorts (2/65 at FDR in cortex; 1/53 in blood), because ALS cortex is broadly dysregulated and bulk DE has limited per-gene power. The candidates therefore behave like bona fide ALS genes, and neither group is enriched above background. This pattern is consistent with indeed predicted by the central premise of the study: ALS-relevant genes are not distinguished by transcriptomic fold-change magnitude, which is why differential-expression screens fail to recover them; their disease relevance is encoded in genomic architecture (evolutionary constraint, neurological-disease annotation, network centrality, brain-enriched expression), features derived from cohort-independent reference databases. We present this as an honest transcriptomic concordance check in independent cohorts, with one robust positive (TMEM43); it is not independent proof of causal involvement, and definitive confirmation will require functional studies in iPSC-derived motor neurons and CNS proteomic data.

Definition and sources of the 66 Positive Class (established ALS-associated) genes

Inclusion criterion: genes catalogued in ALSoD *or* annotated to amyotrophic lateral sclerosis (UMLS C0002736) in the DisGeNET curated gene–disease database, that are present in the transcriptomic universe (13,980 protein-coding genes). Seventy genes were catalogued; four (CFAP410, GLT8D1, ARHGEF28, CSNK1G3) were absent from the universe, leaving n = 66 positives (the "Positive Class"). Of the 66, 31 additionally carry an "amyotrophic lateral sclerosis" phenotype in OMIM and 56 have a scored DisGeNET gene–disease association (score 0–1); the primary evidence linking these genes to ALS spans genetic linkage and GWAS, rare-variant/familial studies, and, for the canonical associated genes (e.g. SOD1, C9orf72, TARDBP, FUS, TBK1), extensive model-organism and cell-based support.

Table S8. The 66 Positive Class genes, their source databases and evidence indicators.

Gene	Source	OMIM ALS phenotype	DisGeNET score	DisGeNET first report
ALS2	ALSoD + DisGeNET	Yes	0.90	1993
ANG	ALSoD + DisGeNET	Yes	1.00	1993
ANXA11	ALSoD + DisGeNET	Yes	0.90	1993
ATXN1	ALSoD + DisGeNET	-	-	-
ATXN2	ALSoD + DisGeNET	Yes	1.00	1993
C9orf72	ALSoD + DisGeNET	Yes	1.00	2013
CCNF	ALSoD + DisGeNET	Yes	0.90	1993
CDH22	ALSoD + DisGeNET	-	-	-
CHCHD10	ALSoD + DisGeNET	Yes	0.90	1993
CHMP2B	ALSoD + DisGeNET	Yes	1.00	1993
CNTN6	ALSoD + DisGeNET	-	-	-
CRYM	ALSoD + DisGeNET	-	-	-
CX3CR1	ALSoD + DisGeNET	-	-	-
DAO	ALSoD + DisGeNET	-	1.00	1993
DCTN1	ALSoD + DisGeNET	Yes	1.00	1993
DNAJC7	ALSoD + DisGeNET	-	0.80	1993
DNMT3A	ALSoD + DisGeNET	-	-	-
DPP6	ALSoD + DisGeNET	-	0.85	2008
EPHA4	ALSoD + DisGeNET	-	-	-

ERBB4	ALSoD + DisGeNET	Yes	0.95	1993
EWSR1	ALSoD + DisGeNET	-	0.85	1993
FIG4	ALSoD + DisGeNET	Yes	0.90	1993
FUS	ALSoD + DisGeNET	Yes	1.00	1993
GFAP	ALSoD + DisGeNET	-	0.85	2001
GLE1	ALSoD + DisGeNET	-	0.85	1993
GPX3	ALSoD + DisGeNET	-	-	-
GRN	ALSoD + DisGeNET	-	0.85	-
GSK3B	ALSoD + DisGeNET	-	0.80	-
HNRNPA1	ALSoD + DisGeNET	Yes	0.85	2013
HNRNPA2B1	ALSoD + DisGeNET	-	0.80	-
KIF5A	ALSoD + DisGeNET	Yes	0.90	1993
LMNB1	ALSoD + DisGeNET	-	-	-
MATR3	ALSoD + DisGeNET	Yes	0.90	1993
NEFH	ALSoD + DisGeNET	Yes	1.00	1993
NEFL	ALSoD + DisGeNET	-	0.95	-
NEK1	ALSoD + DisGeNET	Yes	0.90	1993
NFE2L2	ALSoD + DisGeNET	-	0.90	2016
NIPA1	ALSoD + DisGeNET	-	0.60	1993
OPTN	ALSoD + DisGeNET	Yes	1.00	1993
PFN1	ALSoD + DisGeNET	Yes	0.90	1993
PON1	ALSoD + DisGeNET	-	1.00	1993
PON2	ALSoD + DisGeNET	-	0.80	2010
PPARGC1A	ALSoD + DisGeNET	-	1.00	2013
PRPH	ALSoD + DisGeNET	Yes	1.00	1975
PTGS2	ALSoD + DisGeNET	-	0.80	2001
SARM1	ALSoD + DisGeNET	-	0.80	1993
SCFD1	ALSoD + DisGeNET	-	0.85	1993
SETX	ALSoD + DisGeNET	Yes	0.90	1993
SIGMAR1	ALSoD + DisGeNET	Yes	0.80	2014
SLC1A2	ALSoD + DisGeNET	-	0.80	2001
SMN1	ALSoD + DisGeNET	-	-	-

SOD1	ALSoD + DisGeNET	Yes	1.00	1993
SPG11	ALSoD + DisGeNET	Yes	0.90	1993
SQSTM1	ALSoD + DisGeNET	Yes	1.00	1993
SS18L1	ALSoD + DisGeNET	-	0.85	1993
TAF15	ALSoD + DisGeNET	-	0.85	2011
TARDBP	ALSoD + DisGeNET	Yes	1.00	2008
TBK1	ALSoD + DisGeNET	Yes	0.95	1993
TNIP1	ALSoD + DisGeNET	-	0.70	1993
TREM2	ALSoD + DisGeNET	-	0.80	2014
TUBA4A	ALSoD + DisGeNET	Yes	0.80	1993
UBQLN2	ALSoD + DisGeNET	Yes	1.00	2013
UNC13A	ALSoD + DisGeNET	-	1.00	1993
VAPB	ALSoD + DisGeNET	Yes	1.00	1993
VCP	ALSoD + DisGeNET	Yes	1.00	1993
VDR	ALSoD + DisGeNET	-	0.80	-

Model evaluation under class imbalance

Out-of-fold predictions (66 positives, 13,914 background; class prevalence 0.0047) were evaluated with precision-recall AUC (PR-AUC), precision@top-k and calibration, in addition to ROC-AUC. The SVM is a weak learner; ensemble performance is carried by Random Forest and XGBoost. The model is well calibrated (Brier score = 0.0047).

Table S9. Imbalance-aware discrimination of Positive Class from background (out-of-fold).

Model	**ROC-AUC**	**PR-AUC**	**× baseline**	**P@10**	**P@20**	**P@50**
Ensemble (RF+SVM+XGB)	0.846	0.101	21.4	0.40	0.25	0.18
XGBoost	0.813	0.093	19.6	0.40	0.25	0.18
Random Forest	0.846	0.064	13.6	0.00	0.00	0.14

SVM	0.597	0.006	1.2	0.00	0.00	0.00

Feature ablation

The RF+XGB ensemble was re-trained after removing feature blocks. Neurological-disease (OMIM) annotation is the dominant signal; network-hub (PPI) features and the engineered multiplicative interaction terms are not required for model performance (performance is unchanged or higher without them).

Table S10. Feature-ablation of the ensemble (out-of-fold).

Feature set	**# features**	**ROC-AUC**	**PR-AUC**
Full model	40	0.818	0.077
− OMIM (and disease interaction terms)	27	0.721	0.011
− PPI (degree, hub)	37	0.815	0.080
− OMIM and PPI	24	0.696	0.011
− disease cross-database interactions	34	0.824	0.075
− all engineered interaction terms	28	0.840	0.070

Benchmarking against simpler prioritization methods

The framework was compared with single-axis rankings and simpler learners on recovery of known ALS genes. Differential-expression ranking alone is the weakest prioritizer, and a regularised logistic regression on the same features is competitive with the ensemble, the gain derives from multi-omic feature integration, not the specific nonlinear algorithm.

Table S11. Baseline benchmarking (out-of-fold).

Method	**ROC-AUC**	**PR-AUC**	**P@20**
Differential-expression consensus (Driver score)	0.621	0.011	0.00
PPI degree	0.716	0.023	0.05
Brain expression (TPM)	0.706	0.020	0.10
OMIM neurological-phenotype count	0.769	0.044	0.05
Integrated z-score (5 axes)	0.827	0.067	0.25
Logistic regression (all features)	0.801	0.098	0.25

Elastic net (all features)	0.772	0.069	0.15
Full multi-omic ensemble	0.846	0.101	0.25

Bias controls and robustness

To address circularity and literature/annotation bias, known ALS genes were compared with matched background controls, an alternative background of other disease genes, and a positive-unlabeled (spy) test; model stability was assessed across random seeds.

Table S12. Bias-control and robustness analyses.

Analysis	**Result**
Degree-matched negative controls (ALS gene ranked above its PPI-degree-matched background control)	75.8% of 66 pairs
Annotation-density-matched negative controls (matched on OMIM phenotype count)	75.8% of 66 pairs
Alternative background = genes that themselves carry an OMIM disease entry (n = 4,258; 66 positives)	ROC-AUC 0.851; PR-AUC 0.354
Positive-unlabeled (spy): 20% of positives relabeled as background, then recovered	median 96.4th percentile
Repeated 5-fold out-of-fold across 5 random seeds	ROC-AUC 0.839 ± 0.014; PR-AUC 0.082 ± 0.010
Calibration (Brier score)	0.0047

Feature correlation and redundancy

Of 780 feature pairs, 61 have |Spearman r| > 0.8 (mean |r| across all pairs = 0.284). The strongest correlations are between a feature and its own monotone transform (e.g. LOEUF and pLI, raw and log-transformed counts) or between an interaction term and its parent features. Because the design matrix is therefore rank-deficient, we report tree-based importance and SHAP (robust to collinearity) rather than coefficient-based inference, and show (Table S10) that removing the correlated interaction block leaves performance unchanged.

Table S13. Most strongly correlated feature pairs (|Spearman r|).

Feature pair	|r|
LOEUF – pLI (synthetic, derived from LOEUF)	1.00
LOEUF – LOEUF_Inverted	1.00
Brain_Max_TPM – Log_Brain_TPM	1.00
PPI_Degree – PPI_Degree_Log	1.00
OMIM_N_Phenotypes – Log_OMIM_Phenotypes	1.00
OMIM_N_Neuro – OMIM_Neuro_Ratio	0.999
Hub_Constrained – LOEUF_Hub (interaction terms)	1.00

Per-dataset metadata

The transcriptomic meta-analysis comprises 13 datasets (1,227 samples) across four biological domains. Domain sample sizes are highly imbalanced: the myeloid compartment contributes 1,062 of 1,227 samples (86.5%), dominated by two whole-blood microarrays (GSE112676, GSE112680).

Table S14. The 13 transcriptomic datasets included in the meta-analysis.

Dataset (GEO)	Domain	Tissue	N samples	Platform
GSE33855	Mesenchymal	Fibroblasts	12	Microarray
GSE112676	Myeloid	Blood	741	Microarray
GSE68607_Spor_vs_Ctrl	Lymphoid	LCL	38	Microarray
GSE28253	Lymphoid	Lymphocytes	22	Microarray
GSE122261	Mesenchymal	Muscle_Cells	12	Microarray
GSE18920_AntHorn	CNS	Anterior_Horn	22	Microarray
GSE18920_MotNeuro	CNS	Motor_Neuron	22	Microarray
GSE833	CNS	Motor_Cortex	9	Microarray
GSE87385_Oligo	CNS	Oligodendrocyte	4	Microarray
GSE87385_Astro	CNS	Astrocyte	4	Microarray
GSE112680	Myeloid	Blood	301	Microarray
GSE39642	Myeloid	Monocytes	20	NanoString
GSE56808	Mesenchymal	Fibroblasts	20	Microarray

Leave-one-dataset-out and leave-one-domain-out sensitivity, and influence of the myeloid sample-size imbalance

To test how sensitive the transcriptomic consensus and the downstream prioritization are to individual studies and to the domain imbalance, we recomputed the sqrt(n)-weighted Stouffer consensus signature after removing (i) each dataset in turn (leave-one-dataset-out, LODO) and (ii) each biological domain in turn (leave-one-domain-out, LODomO), and compared each reduced consensus to the full consensus by Spearman rank correlation (genome-wide) and by the Jaccard overlap of the top-100 genes.

The genome-wide gene ranking is highly stable to dropping any single dataset (Spearman ρ = 0.91–1.00, median 0.99; Table S15). However, the *identity of the top-ranked consensus genes* is driven by the large myeloid studies: removing the myeloid domain changes the top-100 set almost entirely (Jaccard = 0.005), and removing the single largest study alone (GSE112676, n=741) reduces the top-100 overlap to 0.09, whereas removing the CNS, lymphoid or mesenchymal domains preserves it (Jaccard 0.77–0.92; Table S16). This confirms the reviewers' concern quantitatively and is the basis for reporting the per-domain up/down gene counts (Supplementary Figure S2) as descriptive material only, with batch/sample-size caveats, rather than as biological domain differences.

Table S15. Leave-one-dataset-out stability of the transcriptomic consensus (each dataset removed; compared to the full 13-dataset consensus).

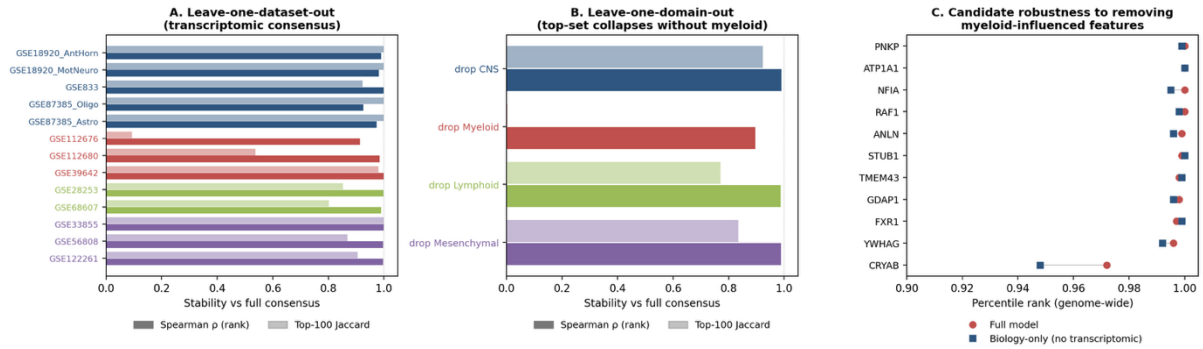
Dataset removed	Spearman ρ (genome-wide)	Top-100 Jaccard
GSE18920_AntHorn	0.99	1.0
GSE18920_MotNeuro	0.982	1.0
GSE833	1.0	0.923
GSE87385_Oligo	0.926	1.0
GSE87385_Astro	0.973	1.0
GSE112676	0.914	0.093
GSE112680	0.985	0.538
GSE39642	1.0	0.98
GSE28253	0.998	0.852
GSE68607	0.99	0.802
GSE33855	0.999	1.0
GSE56808	0.997	0.869

GSE122261	0.997	0.905

Table S16. Leave-one-domain-out stability of the transcriptomic consensus (each domain removed).

Domain removed	Datasets	Spearman ρ	Top-100 Jaccard	Candidates in top 10%
drop CNS	5 datasets	0.99	0.923	5/11
drop Myeloid	3 datasets	0.896	0.005	3/11
drop Lymphoid	2 datasets	0.987	0.77	5/11
drop Mesenchymal	3 datasets	0.989	0.835	5/11

Critically, the multi-omic target identification, the study's primary contribution does *not* depend on this myeloid-driven transcriptomic signal. The meta-analysis contributes only 9 of the 40 model features. Retraining the RF+XGBoost ensemble with all 9 transcriptomic features removed ("biology-only": constraint, network, brain-expression and disease-annotation features, none of which are affected by the transcriptomic sample sizes) leaves discrimination essentially unchanged (out-of-fold ROC 0.83 vs 0.82 for the full model) and keeps all eleven prioritized candidates in the top ~1–5% genome-wide (Table S17; Supplementary Figure S17C). The prioritized targets are therefore not artifacts of the large myeloid sample size.



Supplementary Figure S17. Sensitivity of the transcriptomic consensus and of target prioritization to individual studies and to the myeloid imbalance. (A) Leave-one-dataset-out and (B) leave-one-domain-out stability of the weighted-Stouffer consensus (Spearman ρ for the genome-wide ranking and Jaccard overlap of the top-100 genes); the top-100 set collapses only when the myeloid domain or its largest study is removed. (C) Genome-wide percentile

rank of the eleven candidates in the full model versus a biology-only model with all transcriptomic features removed; all candidates remain in the top ~1–5% either way.

Table S17. Genome-wide percentile rank of the eleven candidates in the full model versus a biology-only model with all transcriptomic (myeloid-influenced) features removed.

Candidate	Percentile (full model)	Percentile (biology-only)
RAF1	1.0	0.998
NFIA	1.0	0.995
ATP1A1	1.0	1.0
PNKP	1.0	0.999
STUB1	0.999	1.0
ANLN	0.999	0.996
GDAP1	0.998	0.996
TMEM43	0.998	0.999
FXR1	0.997	0.999
YWHAG	0.996	0.992
CRYAB	0.972	0.948

References

1. Béland, L.-C. *et al.* Immunity in amyotrophic lateral sclerosis: blurred lines between excessive inflammation and inefficient immune responses. *Brain Commun* **2**, fcaa124 (2020).
2. Neumann, M. *et al.* Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science* 314, 130–133 (2006).
3. DeJesus-Hernandez, M. *et al.* Expanded GGGGCC Hexanucleotide Repeat in Noncoding Region of C9ORF72 Causes Chromosome 9p-Linked FTD and ALS. *Neuron* 72, 245–256 (2011).
4. Huang, S., Dong, W., Lin, X. & Bian, J. Na+/K+-ATPase: ion pump, signal transducer, or cytoprotective protein, and novel biological functions. *Neural Regen Res* 19, 2684–2697 (2024).
5. Biondo, E. D., Spontarelli, K., Ababioh, G., Méndez, L. & Artigas, P. Diseases caused by mutations in the Na+/K+ pump α1 gene ATP1A1. *Am J Physiol Cell Physiol* 321, C394–C408 (2021).
6. Malaspina, A., Kaushik, N. & Belleroche, J. de. A 14-3-3 mRNA Is Up-Regulated in Amyotrophic Lateral Sclerosis Spinal Cord. *Journal of Neurochemistry* **75**, 2511–2520 (2000).
7. García-Criado, F. *et al.* Integrative Transcriptomic and Network-Based Analysis of Neuromuscular Diseases. *Int J Mol Sci* **26**, 9376 (2025).
8. Kubinski, S. & Claus, P. Protein Network Analysis Reveals a Functional Connectivity of Dysregulated Processes in ALS and SMA. *Neurosci Insights* **17**, 26331055221087740 (2022).

9. Chiu, H.-H. *et al.* Clinical and Genetic Characterization of Autosomal Recessive Spinocerebellar Ataxia Type 16 (SCAR16) in Taiwan. *Cerebellum* **19**, 544–549 (2020).
10. Hayer, S. N. *et al.* STUB1/CHIP mutations cause Gordon Holmes syndrome as part of a widespread multisystemic neurodegeneration: evidence from four novel mutations. *Orphanet J Rare Dis* **12**, 31 (2017).
11. Mueller, S., Decker, L., Menge, S., Ludolph, A. C. & Freischmidt, A. The Fragile X Protein Family in Amyotrophic Lateral Sclerosis. *Mol Neurobiol* **60**, 3898–3910 (2023).
12. Freischmidt, A. *et al.* A serum microRNA sequence reveals fragile X protein pathology in amyotrophic lateral sclerosis. *Brain* **144**, 1214–1229 (2021).
13. Grima, N. *et al.* RNA sequencing of peripheral blood in amyotrophic lateral sclerosis reveals distinct molecular subtypes: Considerations for biomarker discovery. *Neuropathology and Applied Neurobiology* **49**, e12943 (2023).