nature portfolio

Peer Review File

Gene Expression Signatures from Whole Blood Predict Amyotrophic Lateral Sclerosis Case Status and Survival

Corresponding Author: Professor Eva Feldman

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
The study : Gene Expression Signatures from Whole Blood Predict Amyotrophic Lateral Sclerosis Case Status and Survival" by Zhao et al. is an investigation on the utility of whole-blood transcriptomic data derived by RNA-seq for the machine-aided diagnosis of ALS and prognosis for survival. Integrating gene expression with clinical variables enhanced prognostic accuracy, distinguishing patients with different survival durations. Analysis identified ALS-relevant pathways, core genes shared with primary disease tissues, and potential therapeutic candidates.
The study is well-described, the methods are adequate and the translational potential high. However, the study in its current form does not represent a significant advancement in the field judging by the following criteria:
• Novelty – the study does not apply a novel concept, whole-blood transcriptomics for diagnostics has already been demonstrated in ALS (van Rheenen, Swindell, etc.). The study demonstrates incremental improvements, but these do not address the current limitations – differential diagnosis against clinically relevant conditions. The study does not apply a new computational/data analysis method that will promise to overcome current limitations and does not provide an application of a novel biotechnology that will improve the cost/accuracy/applicability of the method. The premise is also not completely addressed – the authors highlight the importance of definite ALS diagnosis before symptom onset of very fast progression, but the study does not utilize pre-symptomatic mutation carrier ALS blood samples. The study does not analyze the contribution of this method to expedite diagnosis in clinical settings, and does not demonstrate its value to overcome problems that slow down definite diagnosis (differential diagnosis, compliance, speed of method).
• Systematic analysis – if not offering novelty, the study could contribute as a systematic analysis of its own data and available resources. However, this is not rigorously performed either – the data is compared only to several of the available studies; it has not been analyzed how it compares to the large resource of other ALS transcriptomic studies – from different tissues – or from PBMCs instead of whole blood; the study does not analyze in depth why and how the differentially expressed genes and patterns compare to the current knowledge and re-iterates previous findings/pathways just as confirmation. The drug perturbation is a nice novel approach, but it lacks more rigor – the authors could study if any of the compounds has been tested in similar settings, or if they belong to drug families which have been tested in ALS or other neurodegenerative diseases. The authors could alsos check for epidemiological data on ALS incidence in cohorts which have been treated with drugs from these families.
• Resource – the data is rather large and will be a valuable resource for the community. However, the data is not being made publicly available, the source code/methods neither, and even source data for the figures is not provided. The data availability statement is somewhat subjective – it is not clear, by what criteria researchers will be classified as 'qualified' to access the data.

Furthermore the following specific points could be addressed to improve the study:

- Age, Sex is not perfectly matched – how was this accounted for?
- Table 1: 1/5 Possible, or Suspected, how does this relate to the classification?
- Relatively small Log2FCs in DEG, how will the authors explain this? Feature of the large sample, of the specific modelling, or something else? (The DEG modelling is really nice here, especially the SVA modification, but it's hard to judge without the code or supplementary information)
- No "random" sampling controls and statistics for the overlap with literature DEGs on p.5, l.151, are hard to interpret without that; also guessing 1928 DEGs to be matched was because only these had gene symbols, but this has to be clearly stated -> On a simple check, the overlap seems very impressive, maybe cite some controls/statistics.

- qPCR – primers' sequences should be provided
- For the overlap of the DEGs from other studies and this one, it would be useful to see a scatterplot of the log fold changes: the same data as in Fig 1 h, but with X axis the LogFoldChange from this study, the Y axis the LogFoldChange from the literature, and a scatter point for each DEG gene – this is missing in Suppl. Fig. S2
- Suppl. Fig S3: It seems a bit contradictory that in the VolcanoPlots in Fig. 1, there is only one or two genes with Log2FoldChange close to 1, but in this figure, 4 genes have a LogFoldChange ~ 1.2. Also, were all four of these genes up-regulated? If any of them were down-regulated in ALS, it would help to label the Y-axis here 'Absolute Fold Change (ALS vs. Control)'. It would also help if these 4 genes are labeled directly in the VolcanoPlots in Fig. 1 a
- What was the rationale behind selecting exactly these 4 genes for qPCR validation? They seem not to be in the top 10 Up- and Down-regulated DEGs? The statement 'performed whole blood qPCR of 5 genes, chosen based on altered expression and biological relevance.' is a bit vague. Also, 4 genes are shown, not 5 as stated.
- What are the characteristics of these controls? Clinical? Co-morbidities? In such a large cohort, this should be considered
In the second paragraph of the results, "A gene expression signature predicts ALS case status":
- It is not clear how exactly the subset of samples to train the first round of classifiers is selected: Why? Which samples? Cross-fold validation?
- Is the statement "generating panels of 27 (FDR<0.01; Fig. 2b)," missing the definition of the AvExpr threshold for this group? Otherwise it seems contradictory that filtering with fdr<0.01 gives 27 DEGs, but filtering with the same fdr<0.01 and an additional threshold AvExpr>2, gives more (i.e., 29) DEGs. It would be expected that more stringent filtering further reduces the list of the DEGs
- If the authors cite here that the XGBoost classifier based on the DEGs performs well on the Grima test cohort, they should also cite how it performs on the Swindell dataset. The point that the classifier is performing well on unseen data is the most critical in this paragraph, and crucial for the manuscript's main message, so it is important to make it convincing. I would even recommend comparing it to even more datasets, and comparing with several of the classifiers, to convincingly demonstrate that there is an underlying ground truth. It is not bad if other classifiers and other data (not whole blood, other platforms) don't perform that well, as long as a tendency is clearly visible.
- Fig2: I think there's a typo in the Grima dataset description in the figure, the ALS cases are n=86, not n=96, right?
- QPCR validation: Why are the RNA samples isolated with the miRNA kit of PAXGene, when the selected 5 genes to validate were not miRNAs? Also, I don't doubt on the quality of the qPCR data, but it is still important to control with at least two house-keeping genes.

Reviewer #2

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Reviewer #3

(Remarks to the Author)
Zhao et al., profiled whole blood by RNA-seq from ALS cases (n=422) versus controls (n=272) with the goal of developing an ALS diagnostic tool. This represents the largest sample set of its kind. The authors use the dataset to perform DEG analysis and identify a gene expression signature that differentiates ALS cases from controls; use machine learning algorithms to predict ALS case status and survival; perform pathway analysis to reveal disease associated signatures like immune pathways, RNA processing and energy production; and lastly perform an in silico drug perturbation analysis based on "core genes" that are perturbed in blood as well as iPSC models of TDP-43 KD and ALS spinal cord samples. The study provides a valuable resource for the community and makes some interesting observations. The main thrust is the potential diagnostic/prognostic tool based on gene expression signatures in the blood. This could prove to be very useful. However, as presented the findings seem preliminary and require additional validation.

Major comments
Fig 1
The DEGs between patients and controls are not particularly insightful and entirely descriptive. What is the interpretation of the 50% and 46% overlap between this dataset and the Swindell and Grima DEGs?
Fig 2
The use of the machine learning algorithms is intriguing, but the utility of these tools should be tested further in other datasets beyond Grima. Why not use the Swindell dataset to determine the accuracy of these tools to predict ALS case status? Additional testing against disease mimics such as AD, PD, MS etc. will assess the potential advantage of this approach and specificity relative to NfL levels as the authors discuss in the introduction.
Furthermore, transforming this to a PCR panel and use it in an independent dataset is a critical step towards validating the utility of it as a useful biomarker, which is the intend of the study.
The description and presentation of the algorithms is not intuitive and should be improved.
Figure 3:
Incorporating 8/18 genes on top of clinical features appears to result in only modest improvements to AUC. Authors are careful to specify their models generated "numerically larger AUC values" but p-values presented somewhere would be appreciated. It's unclear whether there is statistical significance. If not, then what is real value of adding genes? This section is not well written—in the end I was left wondering whether the signature accurately predicted ALS survival or not in the Grima dataset.
Figure 4:

A Venn diagram of adjusted/unadjusted DEGs would be helpful.
As this data is based on bulk RNA Seq some of the DEGs could be driven by different cell proportions.
Are there differences based on disease status or long/short classification?
Authors use the same RNA PCs in both their adjusted/unadjusted model. They should check whether any of the RNA PCs correlate with cell abundance (and thus would already be adjusted for in their model).
Figure 5:
Unclear what the contribution of the whole blood transcriptomic dataset is to the drug perturbation analysis relative to the iPSC neurons and spinal cord datasets. Using core genes from each would clarify this concern and increase the confidence in the potential utility of these so-called therapeutic candidates.
Assessing the effects of 1-2 of these drugs on an in vitro model of ALS pathology is critically important to substantiate the utility of the in-silico analysis.

Discussion
This section is well written and informative and the addition f a limitations and strengths section is great for the reader.

Minor comments
Figure 2a isn't an effective schematic. No idea what it means
Line 194: "Therefore, our classifier showed potential as a potential" is confusing

Reviewer #4

(Remarks to the Author)
Review for "Gene Expression Signatures from Whole Blood Predict Amyotrophic Lateral Sclerosis Case Status and Survival"

Summary
This manuscript presents a promising analysis of gene expression signatures in whole blood for predicting ALS status and patient survival. The authors have assembled a substantial cohort and integrated blood data with iPSC-neuron and postmortem spinal cord datasets to develop potentially valuable prediction models. While the topic is clinically important and the approach innovative, there are several methodological concerns that should be addressed to strengthen the reliability of the findings.

Strengths
The study demonstrates several notable strengths, including a large sample size (225 ALS cases, 119 controls) providing adequate statistical power. The integration of blood signatures with CNS-relevant tissues offers important biological context. The authors show thoughtful consideration of cell type proportions in their analysis. The comprehensive pathway enrichment analyses further enhance the study's value.

Areas for Improvement
1. Model development methodology: The gene panel derivation occurs on the same dataset used for model training, which may lead to inflated performance estimates. From a ML perspective identifying differential expression is feature selection. I suggest implementing proper nested cross-validation where feature selection occurs independently within each fold, or clearly separating discovery and validation cohorts before any feature selection (i.e. perform differential expression on each CV split)
2. Performance evaluation metrics: Reliance on AUC alone may not fully characterize model performance given the dataset imbalance. I suggest including PR-curves, precision, recall, F1 scores, and balanced accuracy metrics to provide a more comprehensive assessment of model performance, particularly when comparing to the external dataset.
3. Model comparison framework: Current comparisons between clinical-only and clinical-only+genes-combined models (e.g. Figure 3) don't isolate the independent contribution of gene signatures. I suggest adding explicit genes-only models to quantify the independent predictive value of the gene signatures beyond clinical variables alone.
4. Feature selection stability: The inconsistency in top features across different models (Fig 2b,c,d,e) raises questions about signature robustness. I suggest implementing bootstrap analysis to assess feature selection stability and provide confidence intervals for feature importance rankings.

Additional Recommendations
The authors should make the underlying Code available to enhance reproducibility and transparency.

Recommendation
This work addresses an important clinical need in ALS biomarker development and shows considerable promise. Addressing the methodological concerns outlined above would significantly strengthen the reliability of the findings and enhance their potential impact on clinical practice. I look forward to seeing a revised version that builds on the study's strengths while implementing these improvements.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
Zhao et al. have thoroughly revised the manuscript, putting substantial efforts into improving the underlying study, not only the manuscript. We sincerely appreciate the efforts made and the way that all concerns have been addressed extensively. Although some generalization concern may still remain, we think the revision of this work has rendered it significanly more valuable for publication. In particlar, the publishing of the data and the code increased its value as a resource. Similarly, the revisions of the drug perturbation data and the clarifications in the text as to how this study improves the field - and prepares for the next step - now highlight its value.

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)
The authors have addressed most of the critiques raised previously and the manuscript is substantially improved.

Rev 1 Comment 14 and Rev 3 Comment 2: As requested the authors have tested their classifier on the Swindell dataset but did not go beyond that although both reviewers requested testing against more datasets, which is disappointing.

Reviewer 3, comment 3: The authors claim that “building a classifier for ALS-ALS mimic classification, as well as a PCR format, are both future avenues that are outside the scope of the present study.” While their points are well taken, I do believe that without this crucial step their study is not novel or significant enough beyond the prior work that has been, which they cite. The main novelty here is that “this study has shown blood transcriptomics can accurately predict case-control status in external cohorts”.

(Remarks on code availability)

Reviewer #4

(Remarks to the Author)
The Discovery and Validation Split follows best-pracites. The independent validation on two external dataset with platform differences is strong. Great addition with the added metrics.

The test of the clinical variable only model gives a good idea on the contribution of the gene model, which shows a high predictive power of the clinical variables or genes alone.

The bootstrap analysis is great and shows the variability of the top features.

The added analysis add a lot of value and show a diligent characterization of the data and gives a good perspective of the variability.

I could access the GitHub code - I would recommend to add a Readme document to allow exploration.

The authors addressed my concerns and I can recommend publication in Nature Communications.

(Remarks on code availability)
Code is available but would benefit from a small Readme file.

Reviewer 1

The study: Gene Expression Signatures from Whole Blood Predict Amyotrophic Lateral Sclerosis Case Status and Survival" by Zhao et al. is an investigation on the utility of whole-blood transcriptomic data derived by RNA-seq for the machine-aided diagnosis of ALS and prognosis for survival. Integrating gene expression with clinical variables enhanced prognostic accuracy, distinguishing patients with different survival durations. Analysis identified ALS-relevant pathways, core genes shared with primary disease tissues, and potential therapeutic candidates.

The study is well-described, the methods are adequate and the translational potential high. However, the study in its current form does not represent a significant advancement in the field judging by the following criteria:

Our response: We thank Reviewer 1 for their expert evaluation of our manuscript and their insightful critiques, which we have addressed point-by-point below. The Reviewer's suggestions have helped us substantially improve the manuscript and deliver a better perspective for the *Nature Communications* readership.

Reviewer 1, comment 1: Novelty – the study does not apply a novel concept, whole-blood transcriptomics for diagnostics has already been demonstrated in ALS (van Rheenen, Swindell, etc.). The study demonstrates incremental improvements, but these do not address the current limitations – differential diagnosis against clinically relevant conditions. The study does not apply a new computational/data analysis method that will promise to overcome current limitations and does not provide an application of a novel biotechnology that will improve the cost/accuracy/applicability of the method. The premise is also not completely addressed – the authors highlight the importance of definite ALS diagnosis before symptom onset of very fast progression, but the study does not utilize pre-symptomatic mutation carrier ALS blood samples. The study does not analyze the contribution of this method to expedite diagnosis in clinical settings, and does not demonstrate its value to overcome problems that slow down definite diagnosis (differential diagnosis, compliance, speed of method).

Our response: We acknowledge and fully appreciate the Reviewer's critiques on novelty. We agree that our work is not *conceptually* new, but our work presents *novel advances in methodology required to advance the concept of whole-blood transcriptomics as an ALS diagnostic tool.* Our advances are made possible by our large sample size and granular RNA-seq analysis, which are crucial for moving blood transcriptomics into the diagnostic realm in ALS. A major reason we need these methodological advances as a step towards eventual diagnostic capability is ALS heterogeneity, which is reflected in diverse ALS phenotypes and clinical presentations. Although prior research from van Rheenen et al.[1]/Swindell et al.[2] and Grima et al.[3] leveraged blood transcriptomics for case-control classification in their respective cohorts, they could not successfully apply classification in external cohorts, likely due to the inherent disease heterogeneity. Before we can begin to investigate the potential of transcriptomics for differential diagnosis against conditions clinically mimicking ALS, we need to ascertain that blood transcriptomics can successfully perform case-control classification in independent cohorts and that the methodology is robust and encompasses ALS heterogeneity. Our work accomplishes this objective, and we are now poised to make further advances by integrating blood transcriptomics of ALS mimics to build a classifier that will be helpful for differential diagnosis. We are also now able to move forward on determining whether blood transcriptomics may have utility in presymptomatic ALS.

We had previously acknowledged the lack of ALS mimics as a study weakness in our limitations paragraph. We now also acknowledge the limitation that our study lacks presymptomatic ALS mutation carriers. The revised manuscript now reads, lines 456-460:

"First, we did not include ALS mimics or presymptomatic ALS mutation carriers. In a clinical scenario, a diagnostic tool would need to differentiate patients with ALS from patients without ALS that manifest similar symptoms, i.e., ALS mimics. Moreover, a diagnostic tool capable of identifying imminent phenoconversion to ALS in mutation carriers is lacking; however, our study did not include presymptomatic mutation carriers."

Moreover, we have also better clarified the crucial need for this technical advancement before blood transcriptomics can be leveraged for diagnostic purposes in the conclusions, as follows, lines 492-504:

"Although no such gene expression panel is available for ALS diagnosis and/or prognosis, our findings strongly support the ability of gene expression profiles from whole blood to differentiate ALS cases from controls, to potentially predict survival, and to infer, by pathway enrichment, relevant disease pathways and identify potential drug candidates. Towards this goal, we developed an accurate, externally validated blood-based gene expression signature panel for ALS classification, underscoring potential clinical diagnostic use as a

biomarker. We also combined gene expression features with clinical variables for survival prediction, highlighting potential prognostic utility. *Now that our study has shown blood transcriptomics can accurately predict case-control status in external cohorts, we can begin to assess potential utility for classification versus ALS mimics and in presymptomatic ALS mutation carriers prior to phenoconversion*. We envision possible translation to a clinically viable platform that could accelerate ALS diagnosis and provide patients, based on their personalized gene expression profiles, anticipated survival times."

Reviewer 1, comment 2: Systematic analysis – if not offering novelty, the study could contribute as a systematic analysis of its own data and available resources. However, this is not rigorously performed either – the data is compared only to several of the available studies; it has not been analyzed how it compares to the large resource of other ALS transcriptomic studies – from different tissues – or from PBMCs instead of whole blood; the study does not analyze in depth why and how the differentially expressed genes and patterns compare to the current knowledge and re-iterates previous findings/pathways just as confirmation.

Our response: We are very grateful to Reviewer 1 for these insightful comments and realize we had not sufficiently communicated the purpose of our analyses. The first half of the manuscript is dedicated to case-control classification (potential diagnostic utility) and survival prediction among cases (potential prognostic utility). The second half of the manuscript is dedicated to demonstrating the presence of an ALS disease signature in blood that can also be leveraged in drug perturbation analysis.

Our pathway analysis of differentially expressed genes (DEGs; **Fig. 5** new numbering [**Fig. 4** old numbering]) was not solely a confirmation of previously known pathways but a demonstration that directly adjusting for immune cell proportions (i.e. the relative abundance of different types of immune cells within a sample) further enhances representation of known ALS pathways in the blood. Adjusting DEGs and ensuing pathways for cell proportions accounts for differences in immune cells between ALS cases versus controls and thus enhances pathways intrinsic to the disease process.

To better convey our motivation to readers, we have made edits to the second portion of the paper. We have changed the prior section title from "Pathway analysis of the blood transcriptome reveals ALS-related pathways" to "Pathway analysis of the blood transcriptome reveals an ALS disease signature" (line 257).

We edited the opening sentences of this section to read as follows, lines 258-260:

"After demonstrating the diagnostic and prognostic potential of our ALS blood transcriptomics dataset, *we next* performed pathway enrichment to identify disease-related pathways with the goal of creating an ALS disease signature observable in blood."

We edited the closing sentences of this section to read as follows, lines 303-307:

"Overall, enrichment analysis of the whole blood transcriptome revealed ALS-relevant pathways beyond immune-related pathways. Adjusting for cell proportions, in blood further enriched these ALS pathways and *confirmed the presence of an ALS disease signature in blood*. Thus, we next sought to leverage the blood transcriptomic dataset to identify potential drug candidates by drug perturbation analysis."

As recommended by the Reviewer (but only initially presented in the supplementary information in our initial submission), we completed an assessment of the overlap of our reported DEGs with other tissue types. Specifically, we analyzed the overlap with postmortem spinal cord and TDP-43 knockdown iPSC DEGs. These findings are now presented as a main text figure (***new* Fig. 6a-e**).

Regarding PBMCs, there is no available dataset that is sufficiently large. Current datasets that we could find include Zhang et al. 2011 (PMID 20884065, ALS n=20, controls n=22), Lam et al. 2015 (PMID 26807342, ALS n=9, controls n=1), Gagliardi et al. 2018 (PMID 29402919, ALS n=7, controls n=4), and Zucca et al. 2019 (PMID 29402919, ALS n=24, controls n=7). Therefore, we were unable to compare our dataset to a PBMCs dataset.

Reviewer 1, comment 2, continued: The drug perturbation is a nice novel approach, but it lacks more rigor – the authors could study if any of the compounds has been tested in similar settings, or if they belong to drug families which have been tested in ALS or other neurodegenerative diseases. The authors could also check for epidemiological data on ALS incidence in cohorts which have been treated with drugs from these families.

Our response: Regarding the drug perturbation analysis, we thank the Reviewer for their favorable overall appraisal. We had performed a literature search of the identified candidates for added rigor and relevance to ALS. Of the eight identified candidates from the primary analysis (i.e., overlapping between adjusted overall "core genes" and iPSC-neurons and spinal cord candidates), we found pertinent literature on three. Briefly, FDA-approved trifluoperazine was identified by another drug repurposing effort in ALS,[4] and is a potent allosteric modulator of the purinergic P2X7 receptor,[5] whose inhibition has been proposed as an ALS therapeutic by modulating neuroinflammation.[6] FDA-approved ibrutinib, an irreversible BTK inhibitor that blocks B-cell proliferation, exerted salutary effects in *SOD1^{G93A}* mice.[7] Thirdly, BX-795 is an inhibitor of the ALS gene *TBK1*, linked to innate immunity, autophagy, and cell cycle,[8] proposed as a drug repurposing target based on Mendelian randomization.[9]

Based on the Reviewer's excellent suggestions, we have expanded our discussion of the literature in support of our identified candidates by considering compounds tested in similar settings, belonging to a broader drug family tested in ALS or other neurodegenerative diseases, and from epidemiological ALS data. This portion of the discussion now reads, lines 431-455:

"Among the candidates was an FDA-approved drug, the phenothiazine trifluoperazine, previously identified by a drug repurposing effort in ALS.[4] *Trifluoperazine is a typical antipsychotic primarily used to treat schizophrenia, a neuropsychiatric disorder that may overlap with ALS.[10] While there are reports that use of antipsychotics may lower the risk of ALS[11] and other age-related neurodegenerative diseases,[12] some studies report no association*.[13] Trifluoperazine is also a potent allosteric modulator of the human purinergic P2X7 receptor.[5] P2X7 receptor antagonism has been proposed as an ALS therapeutic strategy by modulating neuroinflammation.[6]

A second identified candidate was an FDA-approved irreversible BTK inhibitor, ibrutinib, that blocks B-cell proliferation and survival and is reported to delay symptom onset, reduce muscular atrophy, and decrease pro-inflammatory brain cytokine production in *SOD1^{G93A}* mice.[7] *Bosutinib, another tyrosine kinase inhibitor approved for chronic myelogenous leukemia, was repurposed for ALS based on an iPSC screen.[14] A small open-label, dose-escalation phase I trial found the drug was safe and potentially effective in an ALS participant subset.[15] Overall, several kinase inhibitors have or are being considered for ALS*.[16] Our drug perturbation analysis also selected BX-795, an inhibitor of TBK1 kinase, a known ALS gene linked to innate immunity, autophagy, and cell cycle.[8] A recent Mendelian randomization of ALS genome-wide association studies proposed TBK1 as a drug repurposing target.[9]

Some candidates had no known literature linked to ALS, including AZD-9291, BIX-02189, fludarabine, valrubicin, and XMD−1150,[17] suggesting future possible research avenues. Indeed, XMD-1150 may potentially stimulate autophagy,[17] impaired in ALS,[8] and inhibit LRRK2 (leucine rich repeat kinase 2),[18] also linked to autophagy in neurodegeneration, especially Parkinson's disease.[19] Overall, our drug perturbation analysis promotes a strategy for identifying therapeutic candidates and supports further investigation into three drugs previously suggested as potential ALS therapies.[20,21]"

Finally, we also refer the Reviewer to our response to **Reviewer 3, comment 10**.

Reviewer 1, comment 3: Resource – the data is rather large and will be a valuable resource for the community. However, the data is not being made publicly available, the source code/methods neither, and even source data for the figures is not provided. The data availability statement is somewhat subjective – it is not clear, by what criteria researchers will be classified as 'qualified' to access the data.

Our response: We apologize to the Reviewer for our earlier omissions. We have deposited the data into a public, controlled database, according to the terms of our data-sharing agreement with the NIH, which funded the work.

This is now reflected in our new data availability statement, added after the Methods, now reads: "Data have been deposited into the Database of Genotypes and Phenotypes (dbGAP) and the European Genome-Phenome Archive (EGA) with accession number EGAS50000001019 at https://ega-archive.org/studies/EGAS50000001019."

We also made the code available on our GitHub site for this manuscript, reflected in our code availability statement, added after the data availability statement, which now reads: "Code has been made available at

https://github.com/yzhao80/ALS_RNA." Additionally, the source data to reproduce figures has been made available.

These data and source code resources will be extremely valuable for future studies of blood transcriptomics in ALS cases and age/sex-matched controls, and we thank the Reviewer for this important suggestion.

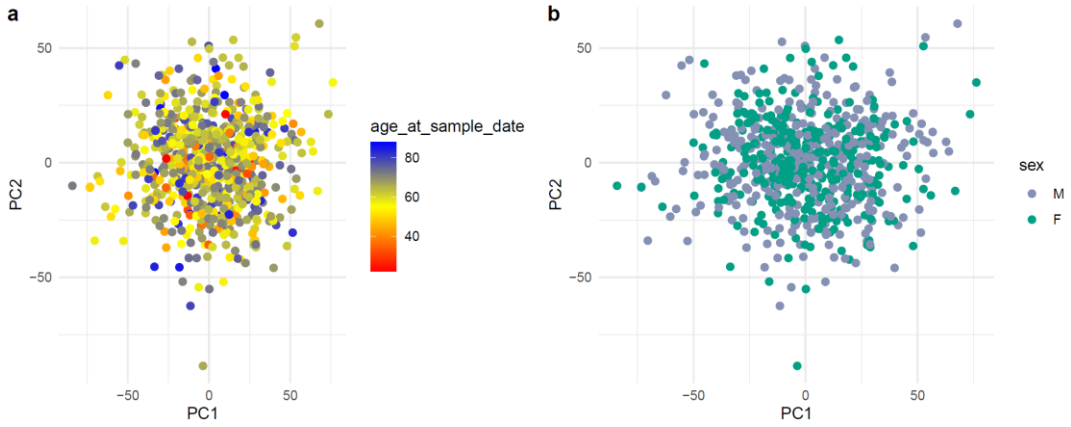
Reviewer 1, comment 4: Age, Sex is not perfectly matched – how was this accounted for?

Our response: We thank Reviewer 1 for raising this issue and the opportunity to clarify this important point.

In the ALS case versus control DEG analysis, we accounted for age and sex by including them as covariates in the linear model, y ~ group + age + sex + genetic_PCs$_{1-4}$ + RNA_PCs$_{2-10}$ + batch + library_type in the limma R package (PC, principal components; v3.48.3; functions: voom, lmFit, eBayes). Principal component analysis (PCA) of gene expression residuals from the linear model y ~ age + sex + genetic_PCs$_{1-4}$ + RNA_PCs$_{2-10}$ + batch + library type is shown in the **Response Figure R1** below**.** In the PCA plots, there were no differences in age (continuous color spectrum) and sex (binary colors) among cases and controls after accounting for them in the linear model.

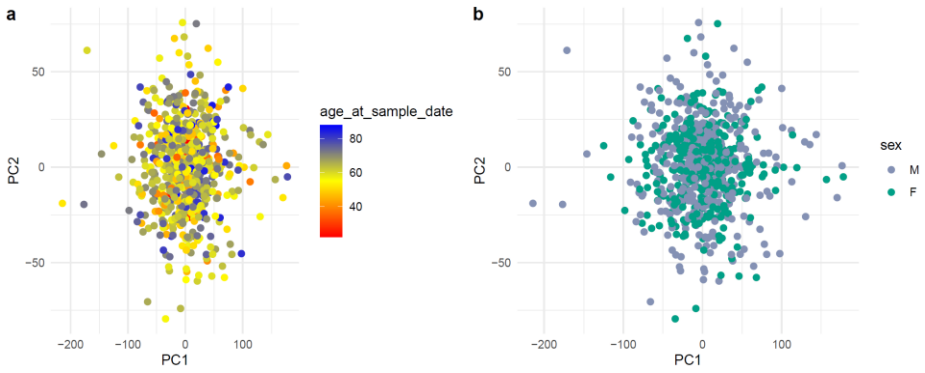
Our procedure for accounting for age and sex is outlined in the manuscript as follows, lines 629-631:

"This model controlled for variation in demographics (age, sex) as well as technical variations due to library preparation, genetic heterogeneity (the first four genetic PCs), and unknown variance (selected RNA PCs) (**Supplementary Fig. S14d**)."



Response Figure R1. PCA plots for primary DEGs analysis. The samples were color-coded by (**a**) age at sample collection, i.e., age_at_sample_date; and (**b**) sex. ALS=422, controls=272, total n=694. Input for PCA were the residuals, which were trimmed mean of M-values normalized $\log_2$ count per million values from linear model y ~ age + sex + genetic_PCs$_{1-4}$ + RNA_PCs$_{2-10}$ + batch + library_type.

Similarly, DEGs between ALS cases versus controls adjusted for cell proportions were identified using a similar model including age and sex. Once again, there were no differences in age or sex among cases and controls after accounting for them in the linear model as presented in the **Response Figure R2**.



Response Figure R2. PCA plots for DEGs analysis adjusted for cell proportions. The samples were color-coded by (**a**) age at sample collection, i.e., age_at_sample_date; (**b**) sex. ALS=422, controls=272, total n=694. Input for PCA were the residuals, which were trimmed mean of M-values normalized $\log_2$ count per million values from linear model y ~ age + sex + CD8T + Mono + NK + BCell + Neu + genetic_PCs$_{1\text{-}4}$ + RNA_PCs$_{2,4\text{-}10}$ + batch + library_type.

Finally, in our case-control classification models, feature selection was performed on DEGs adjusted for age and sex using a linear model. This is specified in the manuscript as follows, lines 699-700:

“In our case-control classification models, feature selection was performed on DEGs adjusted for age and sex and filtered based on average expression.”

Reviewer 1, comment 5: Table 1: 1/5 Possible, or Suspected, how does this relate to the classification?

Our response: We thank the Reviewer for their careful assessment of our manuscript, which is deeply appreciated and provides us the opportunity to improve clarity. In this study, a participant was considered to have ALS if they met the newer Gold Coast criteria,[22] namely (i) progressive motor impairment, documented by history or repeated clinical assessment, preceded by normal motor function; (ii) upper and lower motor neuron dysfunction in at least one body region, or lower motor neuron dysfunction in at least two body regions; and (iii) investigative findings that exclude alternative diseases. The Gold Coast criteria facilitates early and definitive diagnosis, with similar or better diagnostic sensitivity than the revised El Escorial and Awaji criteria.[23,24]

This was specified in the manuscript as follows, lines 508-510:

“The study recruited participants with ALS (n=422) meeting the Gold Coast definition of ALS during their clinical visit at the University of Michigan Pranger ALS Clinic between 2011 and 2021.”

We have now additionally included the following, lines 516-520:

“Gold Coast criteria were used to group participants as cases for classification purposes, while the revised El Escorial criteria were collected as part of their diagnostic workup. Possible or suspected El Escorial criteria at diagnosis did not affect case status prediction relative to definite, probable, or probable, laboratory supported criteria (***new* Supplementary Fig. S13**).”

Since the external datasets did not provide El Escorial criteria at diagnosis, we tested how the 18% of possible and suspected cases affected the performance of our case-control classifier (***new* Supplementary Fig. S13**). We assigned “definite”, “probable”, and “probable, lab supported” cases as Group 1, and “possible” and “suspected” cases as Group 2. Fisher’s exact test results show that incorrect classifications were not enriched in Group 2 versus Group 1 for any of the three XGBoost classifiers.

***new* Supplementary Figure S13**. Classification performance on internal training data by 10-fold cross validation.

		27-gene classifier		30-gene classifier		29-gene classifier		
	El Escorial criteria at diagnosis	**Incorrect**	**Correct**	**Incorrect**	**Correct**	**Incorrect**	**Correct**	**Total**
Group1	**Definite**	5	105	6	104	5	105	110
	Probable	2	124	8	118	7	119	126
	Probable, lab supported	8	96	9	95	9	95	104
Group2	**Possible**	2	52	2	52	1	53	54
	Suspected	2	20	1	21	2	20	22
	All	19	397	26	390	24	392	416
	Fisher's exact test p-value	0.47		0.89		0.85		
	Fisher's exact test odds ratio	0.83 (95%CI 0-2.80)		1.76 (95%CI 0- 7.09)		1.60 (95%CI 0-6.47)		

Reviewer 1, comment 6: Relatively small Log2FCs in DEG, how will the authors explain this? Feature of the large sample, of the specific modelling, or something else? (The DEG modelling is really nice here, especially the SVA modification, but it's hard to judge without the code or supplementary information)

Our response: This is an excellent Reviewer observation. The relatively small log_2FC values in DEGs are primarily due to analysis of blood versus single cell type. Blood is a heterogeneous mixture of cell types, which dilutes gene expression differences in disease versus control, decreasing log_2FC. For example, if a gene changes two-fold in a cell type that represents 5% of blood, this would be observed as only a 1.05-fold change in our whole blood analysis. Additionally, it is also feasible that since blood is not the primary affected tissue in ALS, log_2FC values may be relatively small. In our analysis of blood DEGs that overlap with postmortem spinal cord DEGs (**Supplementary Fig. S12**), all shared DEGs were expressed with smaller Log_2FC values in blood than in spinal cord. Moreover, our log_2FC values are aligned with literature values. Swindell et al.[2] reported 1.11- to 1.79-fold-changes (0.154 to 0.841 log_2FC values) for top upregulated DEGs and 0.91- to 0.71-fold-changes (−0.137 to −0.493 log_2FC values) for top downregulated DEGs. Grima et al.[3] reported 0.195 to 0.900 log_2FC values for all upregulated DEGs (FDR<0.05) and −0.107 to −0.649 log_2FC values for top downregulated DEGs (FDR<0.05). Grima et al.[3] noted these low log_2FC values, stating "Large fold changes in gene expression were not observed between sALS cases and controls, with no significant differentially expressed genes demonstrating a fold change >2."

Regarding the code for DEG modelling, which incorporates the modified SVA, we have made it available to support reproducibility (please refer to our response to **Editorial comment 4**).

Reviewer 1, comment 7: No "random" sampling controls and statistics for the overlap with literature DEGs on p.5, l.151, are hard to interpret without that; also guessing 1928 DEGs to be matched was because only these had gene symbols, but this has to be clearly stated -> On a simple check, the overlap seems very impressive, maybe cite some controls/statistics.

Our response: We thank Reviewer 1 for this excellent suggestion, also aligned with **Reviewer 3, comment 1**. First, just as the Reviewer had surmised, the 1,928 DEGs were because we could specifically match the gene symbols. This statement can be found in the results, lines 156-157, as:

"We matched 1,928 of our DEGs based on gene symbol to literature DEGs."

Second, to improve statistical rigor, we performed Fisher's exact tests to assess the overlap of our DEGs with Swindell et al.[2] and Grima et al.[3] DEGs. For the Swindell analysis, we defined the universe as the 7,509 genes common to both studies, i.e., the intersection of 9,822 genes assayed by Swindell (detectable in ≥20% of GSE112676 [>148/740] and GSE112680 [>61/301] samples) and our background set of 22,332 genes (***new* Supplementary Figure S1**, please see below). Upregulated Swindell DEGs were significantly enriched among our upregulated DEGs (odds ratio 3.52, $p<2.2\times10^{-16}$), and downregulated Swindell DEGs were likewise enriched among our downregulated DEGs (odds ratio 2.22, $p<2.2 \times 10^{-16}$). For the Grima analysis, we defined the universe as the 11,867 genes common to both studies, i.e., the intersection of 12,569 genes assayed by

Grima (lowly expressed genes were filtered by edgeR v3.38.4 filterByExpr function) and our background set of 22,332 genes. Upregulated Grima DEGs were enriched among our upregulated DEGs (odds ratio 4.44, $p=1.26\times10^{-10}$), while downregulated Grima DEGs enriched among our downregulated DEGs (odds ratio 3.48, $p=5.71\times10^{-12}$).

We have now added these results as follows, lines 157-165:

"Of our upregulated ALS DEGs, 381 and 39 overlapped with approximately 50% of Swindell and 46% of Grima DEGs, respectively (**Fig. 2d**). The extent of overlap was lower among downregulated DEGs, with only 267 shared with Swindell (35%) and 56 with Grima (35%) (**Fig. 2e**). Nevertheless, both upregulated and downregulated Swindell DEGs were significantly enriched among our upregulated (odds ratio 3.46, $p=1.31\times10^{-54}$) and downregulated (odds ratio 2.13, $p=2.22\times10^{-19}$) DEGs, respectively (**Supplementary Fig. S1a**). Additionally, upregulated and downregulated Grima DEGs were significantly enriched among our upregulated (odds ratio 4.44, $p=1.26\times10^{-10}$) and downregulated (odds ratio 3.48, $p=5.71\times10^{-12}$) DEGs, respectively (**Supplementary Fig. S1b**)."

***new* Supplementary Figure S1a-b**. 2×2 contingency tables for up- and downregulated DEGs from our dataset with Swindell et al.[2] (top) and Grima et al.[3] (bottom) datasets.

7,509 universe genes	**Upregulated in Swindell et al.[2]**	**NOT upregulated in Swindell et al.[2]**
Upregulated in our dataset	381	1,547
NOT upregulated in our dataset	371	5,210
Upregulated Swindell DEGs significantly enriched among our upregulated DEGs (odds ratio 3.46, $p=1.31\times10^{-54}$, Fisher's exact test)		
7,509 universe genes	**Downregulated in Swindell et al.[2]**	**NOT downregulated in Swindell et al.[2]**
Downregulated in our dataset	267	1,357
NOT downregulated in our dataset	497	5,388
Downregulated Swindell DEGs significantly enriched among our downregulated DEGs (odds ratio 2.13, $p=2.22\times10^{-19}$, Fisher's exact test)		

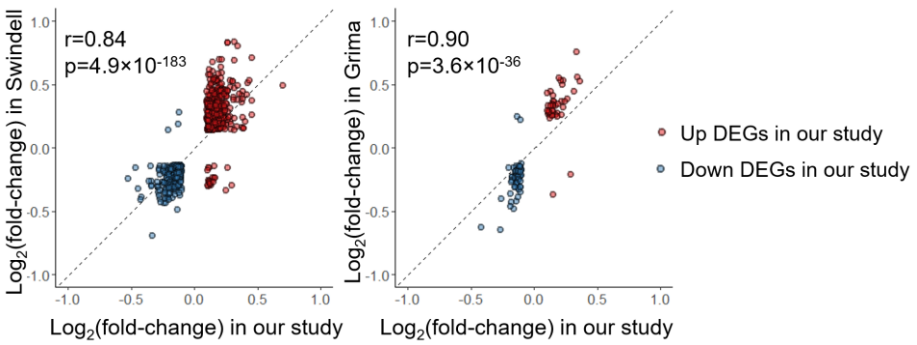
11,867 universe genes	**Upregulated in Grima et al.[3]**	**NOT upregulated in Grima et al.[3]**
Upregulated in our dataset	39	1,889
NOT upregulated in our dataset	46	9,893
Upregulated Grima DEGs significantly enriched among our upregulated DEGs (odds ratio 4.44, $p=1.26\times10^{-10}$, Fisher's exact test)		
11,867 universe genes	**Downregulated in Grima et al.[3]**	**NOT downregulated in Grima et al.[3]**
Downregulated in our dataset	56	1,568
NOT downregulated in our dataset	104	10,139
Downregulated Grima DEGs significantly enriched among our downregulated DEGs (odds ratio 3.48, $p=5.71\times10^{-12}$, Fisher's exact test)		

Reviewer 1, comment 8: For the overlap of the DEGs from other studies and this one, it would be useful to see a scatterplot of the log fold changes: the same data as in Fig 1 h, but with X axis the LogFoldChange from this study, the Y axis the LogFoldChange from the literature, and a scatter point for each DEG gene – this is missing in Suppl. Fig. S2

Our response: We again express our gratitude for another excellent recommendation. We observed a strong linear correlation between $\log_2$(fold-changes) in our DEGs to $\log_2$(fold-changes) of Swindell et al.[2] DEGs (Pearson $r=0.84$, $p=4.9\times10^{-183}$), and an even higher correlation to $\log_2$(fold-changes) of Grima et al.[3] DEGs ($r=0.90$, $p=3.6\times10^{-36}$). We have now added the scatter plots as ***new* Supplementary Fig. S2a**.

We have also added these results as follows, lines 171-173:

"Although shared DEGs were relatively independent of extent of fold-change (**Fig. 2h**) there was a strong correlation of our DEGs to published DEGs (**Supplementary Fig. S2a**)."



***new* Supplementary Figure S2a**. Scatter plot of log_2(fold-change) in Swindell et al.[2] (left) and Grima et al.[3] (right) DEGs versus log_2(fold-change) in DEGs that overlap in our dataset.

Reviewer 1, comment 9: Suppl. Fig S3: It seems a bit contradictory that in the VolcanoPlots in Fig. 1, there is only one or two genes with Log2FoldChange close to 1, but in this figure, 4 genes have a LogFoldChange ~ 1.2. Also, were all four of these genes up-regulated? If any of them were down-regulated in ALS, it would help to label the Y-axis here 'Absolute Fold Change (ALS vs. Control)'. It would also help if these 4 genes are labeled directly in the VolcanoPlots in Fig. 1a.

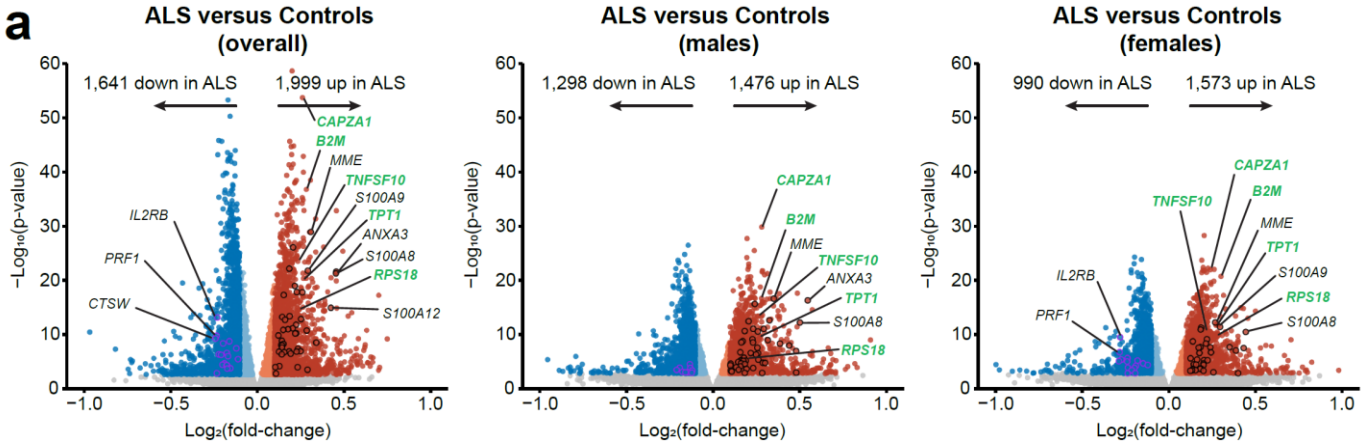
Our response: We again acknowledge the Reviewer's thorough evaluation of our manuscript and express our deepest appreciation. The reason for the discrepancy is the difference in scale between the plots. **Fig. 2** (new numbering, **Fig. 1** old numbering) volcano plots are on a log_2FC scale whereas **Supplementary Fig. S3** is on a fold-change scale (i.e., it is not log-transformed). All five genes were upregulated, which we have now indicated in the figure legend as follows:

"qPCR validation of five selected *upregulated* genes in an independent case-control cohort."

We also reemphasize this in the results, which now states, lines 175-178:

"Finally, to confirm our RNA-seq results, we selected a new, independent cohort of 29 ALS and 27 control participants (**Table 2**) and performed whole blood qPCR of *5 genes upregulated in ALS*, chosen based on altered expression and biological relevance."

We also completely agree on adding the five genes to the **Fig. 2a** (new numbering, **Fig. 1a** old numbering) volcano plots, and updated versions have been resubmitted with our revised manuscript. We thank the Reviewer for this excellent point.



***Revised* Figure 2a**. Volcano plots of DEGs in ALS cases versus controls in all participants (left), males (middle), and females (right). Circles represent upregulated (deep red) and downregulated (deep blue) DEGs in ALS meeting the criteria of absolute value of log_2(fold-change)>0.1 (absLFC) and Benjamini-Hochberg (BH)-adjusted p-value for false discovery rate (FDR)<0.01. Some immune-related DEGs annotated; neutrophil

markers (black outlined circles); natural killer cell markers (purple outlined circles). Grey, light red, and light blue circles denote DEGs not meeting the criteria. *The five select upregulated DEGs validated by qPCR are indicated in green font*.

Reviewer 1, comment 10: QPCR validation: What was the rationale behind selecting exactly these 4 genes for qPCR validation? They seem not to be in the top 10 Up- and Down-regulated DEGs? The statement 'performed whole blood qPCR of 5 genes, chosen based on altered expression and biological relevance.' is a bit vague. Also, 4 genes are shown, not 5 as stated. Why are the RNA samples isolated with the miRNA kit of PAXGene, when the selected 5 genes to validate were not miRNAs? Also, I don't doubt on the quality of the qPCR data, but it is still important to control with at least two house-keeping genes. qPCR – primers' sequences should be provided

Our response: We thank Reviewer and agree we must clarify the presentation of our results in this section. In redoing the qPCR to address the Reviewer's suggestion to compare to two house-keeping genes, we now have data on 5 genes, *B2M*, *CAPZA1*, *RPS18*, *TNFSF10*, and *TPT1*. We could not include the qPCR primer sequences in the methods since these sequences were not provided by Thermo Fisher Scientific; however, we have added the catalog numbers, as follows, lines 798-805:

"qPCR was performed using TaqMan Gene Expression Master Mix (catalog no. 4369016, Thermo Fisher Scientific) with TaqMan Gene Expression Assay primers for *B2M* (catalog no. 4331182, Assay ID Hs00187842_m1), *CAPZA1* (catalog no. 4331182, Assay ID Hs04187789_g1), *RPS18* (catalog no. 4331182, Assay ID Hs01375212_g1), *TPT1* (catalog no. 4331182, Assay ID Hs00372008_m1), and *TNFSF10* (catalog no. 4331182 Assay ID Hs00921974_m1) using both *GAPDH* (catalog no. 4331182 Assay ID Hs02786624_g1) and *ACTB* (catalog no. 4331182 Assay ID Hs01060665_g1) as endogenous controls, and data were analyzed by the delta-delta C_T method."

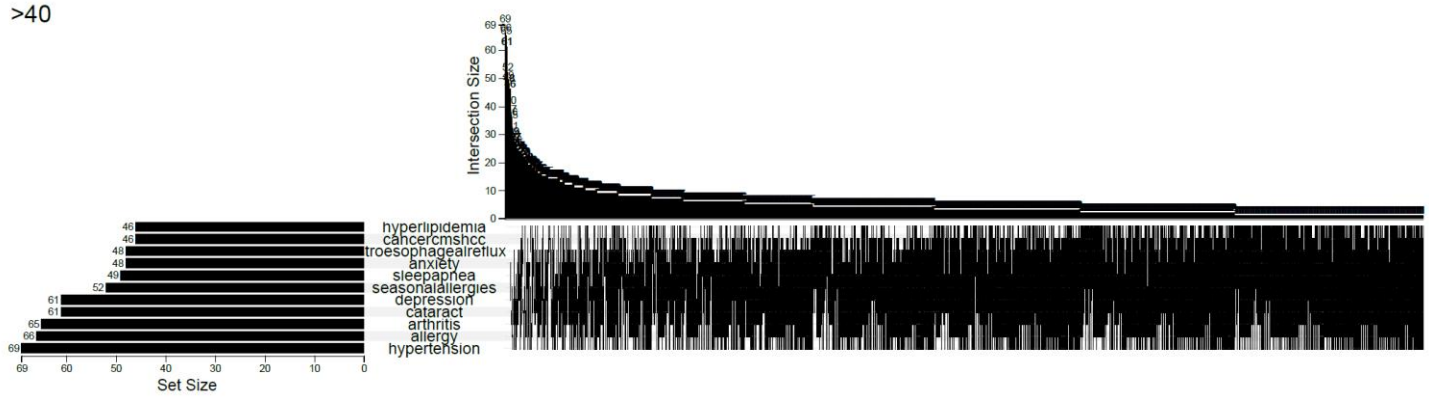
With respect to the RNA isolation from blood using the PAXGene miRNA kit (catalog no. 763134, Qiagen), despite the name, this kit yields "high-quality total RNA, including miRNA, for molecular analysis."

Lastly, regarding selection of targets, the top up- and downregulated DEGs of **Fig. 2c** (new numbering, **Fig. 1c** old numbering) were based solely on FDR. However, to select the qPCR targets, we considered multiple characteristics of biological relevance, such as average expression (AveExpr>0 in the unit of log count per million) and fold-change (absolute value of $\log_2$FC>0.2), as well as statistical significance (FDR<0.01) in DEGs from overall ALS cases versus controls and stratified by sex, yielding 639 filtered DEGs. These 639 DEGs were further filtered according to (i) AveExpr>5, (ii) FDR<0.05 stratified by sex, (iii) FDR<0.05 among DEGs also adjusted for cell proportions, overall and stratified by sex, and (iv) differentially expressed by Wilcoxon test (adjusted for batch effects, FDR < 0.05). We selected five targets from this list of 72 DEGs to perform qPCR, which were among the top 20 differentiating ALS cases from controls or with some potential literature relevance to ALS (beta-2-microglobulin [*B2M*],[25] capping actin protein of muscle Z-line subunit alpha 1 [*CAPZA1*], ribosomal protein S18 [*RPS18*],[26] tumor protein, translationally-controlled 1 [*TPT1*] a microtubule-binding protein with roles in autophagy regulation,[27-29] and TNF superfamily member 10 [*TNFSF10*]).[30]

Reviewer 1, comment 11: What are the characteristics of these controls? Clinical? Co-morbidities? In such a large cohort, this should be considered

Our response: We completely agree with the Reviewer, and we have comprehensively collected comorbidities in control participants. Among controls, the top 11 conditions (out of 881, present in 40 or more control participants) were hypertension (n=69, 25.4%), allergy (n=66, 24.3%), arthritis (n=65, 23.9%), depression (n=61, 22.4%), cataract (n=59, 21.7%), seasonal allergies (n=52, 19.1%), anxiety (n=48, 17.6%), GERD gastroesophageal reflux disease (n=48, 17.6%), sleep apnea (n=48, 17.6%), hyperlipidemia (n=46, 16.7%), and cancer CMS-HCC (n=44, 16.2%). All these conditions were present simultaneously in several individuals (see **Response Figure R3**). Overall, our control cohort was representative of an older population of adults.

Response Figure R3. Upset plot among the top 11 conditions present in 40 or more controls participants (n=272 total).



Reviewer 1, comment 12: In the second paragraph of the results, “A gene expression signature predicts ALS case status”: It is not clear how exactly the subset of samples to train the first round of classifiers is selected: Why? Which samples? Cross-fold validation?

Our response: We thank the Reviewer for the comment and the opportunity to clarify our approach. In the initial round to develop the classifier, we randomly split the full dataset into training (70%) and testing (30%) subsets, while preserving the proportion of ALS cases and controls in each. We implemented this random splitting using the createDataPartition function in the caret R package as follows:

```
set.seed(123)

trainIndex <- createDataPartition(data$ALS_status, p = 0.7, list = FALSE)

trainData <- data[trainIndex, ]

testingData <- data[-trainIndex, ]
```

We then used the same training and testing sets across all seven machine learning algorithms to ensure comparability. Importantly, we applied 10-fold cross-validation for hyperparameter tuning and model selection within the training set during which the 30% testing dataset remained untouched.

Our approach is described in the Methods section as follows, lines 659-662:

“To compare machine learning classifiers for this task, a random 70% (n=487 total with n=296 ALS, n=191 controls) of the total gene expression dataset (n=694) was used for 10-fold cross-validation training with the remaining 30% held out for testing (n=207 total with n=126 ALS, n=81 controls).”

We have now also further added as follows, lines 662-663:

“The same training and testing sets were used across all seven machine learning algorithms to ensure comparability.”

Reviewer 1, comment 13: Is the statement “generating panels of 27 (FDR<0.01; Fig. 2b),” missing the definition of the AvExpr threshold for this group? Otherwise it seems contradictory that filtering with fdr<0.01 gives 27 DEGs, but filtering with the same fdr<0.01 and an additional threshold AvExpr>2, gives more (i.e., 29) DEGs. It would be expected that more stringent filtering further reduces the list of the DEGs

Our response: We thank the Reviewer and fully agree that we should supply more information. The three gene panels in **Fig. 3b-d** (new numbering, **Fig. 2b-d** old numbering) are the resulting genes after XGBoost feature selection. The input gene features for XGBoost for the 27-gene panel were 3,640 DEGs in ALS cases versus controls after applying FDR<0.01 without any AvExpr filtering (1,999 upregulated, 1,641 downregulated). The input gene features for XGBoost for the 30-gene panel were 3,261 DEGs in ALS cases versus controls after applying FDR<0.01 and AvExpr>0 filtering (1,751 upregulated, 1,510 downregulated). Finally, the input gene features for XGBoost for the 29-gene panel were 2,621 DEGs in ALS cases versus

controls after applying FDR<0.01 and AvExpr>2 filtering (1,354 upregulated, 1,267 downregulated). Therefore, Reviewer 1 is correct, the number of input DEGs decreased as more stringent filtering criteria were applied.

We have now clarified this methodology in **Fig. 3a** (new numbering, **Fig. 2a** old numbering) and its corresponding figure legend, lines 1,130-1,140:

We have also reworded the results as follows, lines 197-200:

"We retrained XGBoost on three groups of DEGs from our entire cohort (ALS cases, n=422; controls, n=272) (**Fig. 3a**) filtered using the criteria FDR<0.01 (3,640 DEGs), FDR<0.01 with AvExpr>0 (3,261 DEGs), and FDR<0.01 with AvExpr>2 (2,621 DEGs), yielding 27- (**Fig. 3b**), 30- (**Fig. 3c**), and 29-gene (**Fig. 3d**) panels, respectively."

Reviewer 1, comment 14: If the authors cite here that the XGBoost classifier based on the DEGs performs well on the Grima test cohort, they should also cite how it performs on the Swindell dataset. The point that the classifier is performing well on unseen data is the most critical in this paragraph, and crucial for the manuscript's main message, so it is important to make it convincing. I would even recommend comparing it to even more datasets, and comparing with several of the classifiers, to convincingly demonstrate that there is an underlying ground truth. It is not bad if other classifiers and other data (not whole blood, other platforms) don't perform that well, as long as a tendency is clearly visible.

Our response: This is an excellent Reviewer recommendation to further validate the utility of our case-control classifier, and also aligns with **Reviewer 3, comment 2**. We initially applied our classifier to the Grima et al.[3] dataset since that study also leveraged RNA-seq, compared to the Swindell et al.[2] dataset, which relied on microarray. RNA-seq has better dynamic range and sensitivity for quantifying expression of low-abundance transcripts;[31] therefore, we felt a comparison of our classifier, built on RNA-seq gene expression, would be optimal against the Grima et al.[3] dataset. Additionally, 12 out of 46 gene features from our XGBoost models were missing from the van Rheenen et al.[1]/Swindell et al.[2] datasets, which would lower accuracy. However, we do completely agree with Reviewer 1 that, for the best demonstration of the universality of our classifier, applying it to the Swindell et al.[2] dataset would be informative.

Despite the above-stated limitations, our classifier still performed well, with AUCs ranging from 0.654 to 0.704 in the van Rheenen/Swindell GSE112676 dataset (233 ALS cases versus 508 controls, microarray) and 0.691 to 0.747 in the van Rheenen/Swindell GSE112680 (164 ALS cases versus 137 controls, microarray) (**Supplementary Fig. S5**). For the genes missing in the van Rheenen/Swindell datasets, we used their median value in the training dataset to replace the missing values.

We have added our analysis to the results as follows, lines 208-211:

"Additionally, we externally evaluated our classifier in the van Rheenan et al.[1]/Swindell et al.[2] cohort, attaining AUCs of 0.654 to 0.747 despite the microarray nature of the dataset, which also lacked 12 out of 46 gene features from our XGBoost models (**Supplementary Fig. S5**)."

We have also added the full results as ***new*** **Supplementary Figure S5**.

new **Supplementary Figure S5**. Performance metrics for predicting case status by the three XGBoost-selected gene panels plus the combined panel on the external Rheenan et al.[1]/Swindell et al.[2] dataset. AUC, area under the receiver operating characteristic curves.

	van Rheenen et al.[1]/Swindell et al.[2] (233 ALS cases versus 508 controls, microarray, GSE112676)				**van Rheenen et al.[1]/Swindell et al.[2] (164 ALS cases versus 137 controls, microarray, GSE112680)**			
	27-gene panel	**30-gene panel**	**29-gene panel**	**Combined 46-gene panel**	**27-gene panel**	**30-gene panel**	**29-gene panel**	**Combined 46-gene panel**
Accuracy	58.97%	64.10%	59.38%	60.73%	62.13%	67.44%	64.78%	64.78%
Balanced accuracy	60.32%	64.76%	61.09%	61.95%	61.99%	67.18%	64.62%	64.62%
Sensitivity	63.95%	66.52%	65.67%	65.24%	63.41%	70.12%	66.46%	66.46%

Specificity	56.69%	62.99%	56.50%	58.66%	60.58%	64.23%	62.77%	62.77%
AUC	65.42%	70.40%	66.54%	67.71%	69.05%	74.65%	70.78%	71.79%
Precision	40.38%	45.19%	40.91%	41.99%	65.82%	70.12%	68.12%	68.12%
F1 Score	49.50%	53.82%	50.41%	51.09%	64.60%	70.12%	67.28%	67.28%

Regarding the possibility to test our classifier in PBMCs, related to **Reviewer 1**, **comment 2**, the two datasets GSE106443 and GSE115259 combined contain only 27 ALS cases and 7 controls. As a distinct type of sample, the cell composition of PBMCs differs from whole blood, so our classifier may not necessarily predict accurately. Additionally, the PBMC dataset is small and imbalanced in case-control, making scaling of the dataset challenging.

We also tested the same three gene panels using other classifiers: LASSO, elastic net, and random forest. Interestingly, the alternative classifiers performed comparably well (**Response Table R1**), demonstrating the robustness of the selected gene features.

Response Table R1. Performance metrics of elastic net, LASSO, and random forest models on the three gene panels for predicting case status in the external Grima et al.[3] dataset. AUC, area under the receiver operating characteristic curves.

	Training data 10-fold cross-validation metrics (internal data, 422 ALS + 276 control)							Testing data metrics (external, 86 ALS + 46 control)						
Elastic Net (l1:l2=0.5)	Accuracy	AUC	sensitivity	Specificity	F1	Precision	Balanced Accuracy	Accuracy	AUC	sensitivity	Specificity	F1	Precision	Balanced Accuracy
Gene-27	92.94	98.19	94.98	89.84	94.16	93.49	92.41	82.58	89.43	84.88	78.26	86.39	87.95	81.57
Gene-29	93.52	97.84	95.42	90.40	94.64	93.98	92.91	83.33	88.30	86.05	78.26	87.06	88.10	82.15
Gene-30	93.81	97.96	95.22	91.81	94.83	94.52	93.52	85.61	89.84	91.86	73.91	89.27	86.81	82.89

	Training data 10-fold cross-validation metrics (internal data, 422 ALS + 276 control)							Testing data metrics (external, 86 ALS + 46 control)						
LASSO	Accuracy	AUC	sensitivity	Specificity	F1	Precision	Balanced Accuracy	Accuracy	AUC	sensitivity	Specificity	F1	Precision	Balanced Accuracy
Gene-27	93.22	98.25	94.22	91.62	94.32	94.52	92.92	83.33	89.33	84.88	80.43	86.9	89.02	82.66
Gene-29	93.66	97.64	95.19	91.07	94.75	94.43	93.13	81.82	88.47	83.72	78.26	85.71	87.8	80.99
Gene-30	93.80	97.92	95.00	91.94	94.86	94.76	93.47	83.33	90.75	88.37	73.91	87.36	86.36	81.14

	Testing data metrics (external, 86 ALS + 46 control)			
Random Forest	Accuracy	AUC	Sensitivity	Specificity
Gene-27	78.03%	87.65%	86.05%	63.04%
Gene-29	80.30%	87.97%	84.88%	71.74%
Gene-30	81.06%	89.12%	87.21%	69.57%

Reviewer 1, comment 15: Fig2: I think there's a typo in the Grima dataset description in the figure, the ALS cases are n=86, not n=96, right?

Our response: We thank the Reviewer for noting this typo through their careful review. Yes, the number of ALS cases should be 86. The Grima et al.[3] paper describe 96 sporadic ALS cases, but the deposited dataset GSE234297 includes data for only 86 cases. We have remade **Fig. 3a** (new numbering, **Fig. 2a** old numbering) in response to **Reviewer 3, comment 12**.

Finally, we would like to express our deepest gratitude to Reviewer 1 one last time. Addressing Reviewer 1's comments has significantly strengthened the presentation of our data and the new requested analyses expanded and markedly improved the scope of our manuscript.

Reviewer 2

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Reviewer 3

Zhao et al., profiled whole blood by RNA-seq from ALS cases (n=422) versus controls (n=272) with the goal of developing an ALS diagnostic tool. This represents the largest sample set of its kind. The authors use the dataset to perform DEG analysis and identify a gene expression signature that differentiates ALS cases from controls; use machine learning algorithms to predict ALS case status and survival; perform pathway analysis to reveal disease associated signatures like immune pathways, RNA processing and energy production; and lastly perform an in silico drug perturbation analysis based on "core genes" that are perturbed in blood as well as iPSC models of TDP-43 KD and ALS spinal cord samples. The study provides a valuable resource for the community and makes some interesting observations. The main thrust is the potential diagnostic/prognostic tool based on gene expression signatures in the blood. This could prove to be very useful. However, as presented the findings seem preliminary and require additional validation.

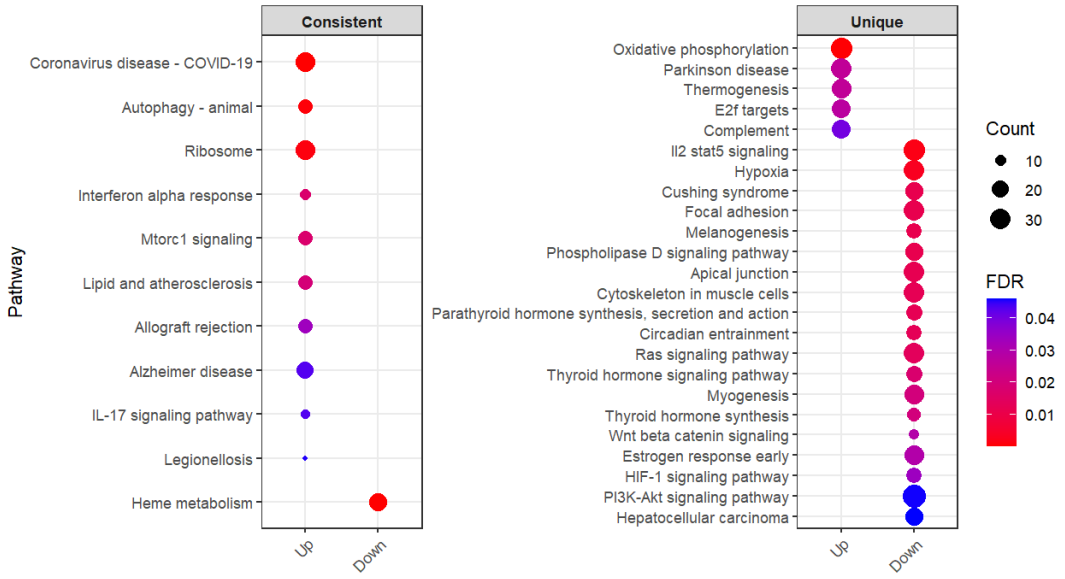
Our response: We are grateful to Reviewer 2 for their expert and insightful critiques of our manuscript. We have responded to all these helpful suggestions, as outlined in detail below, which, we believe, contributed to better depth and clarity in the manuscript.

Reviewer 3, comment 1: Fig 1. The DEGs between patients and controls are not particularly insightful and entirely descriptive. What is the interpretation of the 50% and 46% overlap between this dataset and the Swindell and Grima DEGs?

Our response: We agree with the Reviewer on the descriptive nature of **Fig. 2d-e** (new numbering, **Fig. 1d-e** old numbering). We also realize we did not sufficiently explain our motivation for examining the overlap, which was to assess the extent our dataset reflected previously published datasets, based on DEGs. We have reassessed the overlap following more stringent statistical analyses, also requested by **Reviewer 1, comment 7**, and we kindly direct Reviewer 3 to our response therein. We have also, besides the Venn diagram, generated volcano plots to better illustrate $\log_2$FC of the overlapping and unique DEGs, aligned with **Reviewer 1, comment 8**. Additionally, we have performed pathway enrichment of the overlap DEGs to identify the most recurrent, shared pathways, and of the unique DEGs to pinpoint the least recurrent pathways, outlined in our response to **Reviewer 1, comment 7**. The most significant neurodegenerative term was "Alzheimer's disease", which occurred among consistent DEGs, along with other relevant terms, including "IL-17 signaling pathway", "autophagy – animal", "mTORC1 signaling", and "interferon alpha response".

This is now stated in the results as follows, lines 165-170:

"The DEGs consistent across the three studies were enriched with "Alzheimer's disease", as well as other relevant terms, including "IL-17 signaling pathway", "autophagy – animal", "mTORC1 signaling", and "interferon alpha response". Among neurodegenerative relevant disease pathways, "Parkinson disease", "oxidative phosphorylation", and relevant immune pathways were uniquely enriched in our DEGs (**Supplementary Fig. S1c**)."



***new* Supplementary Figure S1c**. KEGG and MSigDB Hallmark pathway enrichment analysis of our DEGs that are consistent and unique with Swindell et al.[2] and Grima et al.[3] literature DEGs.

Reviewer 3, comment 2: Fig 2. The use of the machine learning algorithms is intriguing, but the utility of these tools should be tested further in other datasets beyond Grima. Why not use the Swindell dataset to determine the accuracy of these tools to predict ALS case status?

Our response: We thank the Reviewer for this excellent suggestion and kindly refer them to our response to **Reviewer 1, comment 14**, which is directly aligned with and fully addresses this comment.

Reviewer 3, comment 3: Additional testing against disease mimics such as AD, PD, MS etc. will assess the potential advantage of this approach and specificity relative to NfL levels as the authors discuss in the introduction. Furthermore, transforming this to a PCR panel and use it in an independent dataset is a critical step towards validating the utility of it as a useful biomarker, which is the intend of the study. The description and presentation of the algorithms is not intuitive and should be improved.

Our response: We completely share the Reviewer's perspective on the need to test against disease mimics and transforming the classifier into a PCR panel. Our current classifier was not trained incorporating transcriptomics data from ALS mimics and, therefore, cannot be applied to differentiating ALS cases from ALS mimics in its current form. Building a classifier for ALS-ALS mimic classification, as well as a PCR format, are both future avenues that are outside the scope of the present study. Prior to our work, it was uncertain whether blood transcriptomics could be leveraged to predict case-control status. Earlier papers, van Rheenen et al.[1]/Swindell et al.[2] and Grima et al.,[3] successfully applied blood transcriptomics for case-control classification in their respective cohorts but not to external cohorts. Therefore, our work achieved a crucial step by ascertaining that blood transcriptomics can successfully perform case-control classification in independent cohorts and that the methodology is robust even among heterogenous ALS cases and diverse cohorts.

Having accomplished this, we are now poised to make further advances by integrating blood transcriptomics of ALS mimics to build a classifier that will be helpful for differential diagnosis and, further, develop it into a PCR panel. However, this will require, per our power calculations, 75 cases of each of the most common ALS mimics, including inclusion body myositis, inflammatory demyelinating neuropathies (chronic inflammatory demyelinating neuropathy, multifocal motor neuropathy with conduction block), and cervical and lumbosacral radiculopathy, for a total of 225 participants. While out of the scope of this manuscript, our plan is to recruit these cohorts, as the ability to distinguish ALS from ALS mimics is an important future direction. Once we identify the gene panel for the best performing ALS-ALS mimic classifier, we will need to optimize for multiplex qPCR biomarker by testing technical concordance, another future direction for us.

We have better formulated the path forward, as follows, lines 499-504:

"Now that our study has shown blood transcriptomics can accurately predict case-control status in external cohorts, we can begin to assess potential utility for classification versus ALS mimics and in presymptomatic ALS mutation carriers prior to phenoconversion. We envision possible translation to a clinically viable platform that could accelerate ALS diagnosis and provide patients, based on their personalized gene expression profiles, anticipated survival times."

Lastly, we have edited the methods to clarify the description of the algorithms. The part for 10-fold cross-validation with XGBoost now reads as follows, lines 680-685:

"Using each set of filtered DEGs as input, three initial models were trained with a learning rate of 0.1, max depth 20, lambda 0.0003, alpha 0.0003, and learning task set to binary classification by logistic regression outputting probability. For feature selection for each of the three models, based on the relative importance of each feature, we repeatedly fit the models with the same hyperparameters by excluding the least important features in each iteration (i.e., recursive feature elimination)."

Reviewer 3, comment 4: Fig 3. Incorporating 8/18 genes on top of clinical features appears to result in only modest improvements to AUC. Authors are careful to specify their models generated "numerically larger AUC values" but p-values presented somewhere would be appreciated. It's unclear whether there is statistical significance. If not, then what is real value of adding genes? This section is not well written—in the end I was left wondering whether the signature accurately predicted ALS survival or not in the Grima dataset.

Our response: We deeply appreciate the Reviewer's critical feedback and fully agree that a statistical assessment of AUC improvements is necessary. To address this, first we performed paired comparisons of time-dependent AUCs across 30 random train/test splits for each year (1–8) in our internal dataset.[3] We applied the Wilcoxon signed-rank test at each time point to compare models, and adjusted the p-values by the Bonferroni method. The results show that adding 18 genes (selected by a stepwise model) or 8 genes (selected by the XGBoost model) to clinical variables **statistically significantly improved** AUCs over the baseline (clinical variables only) model in our internal dataset, particularly between years 4–8 (adjusted p-values<0.01 at most time points, ***new*** **Supplementary Fig. S6a**).

New **Supplementary Figure S6a**. **Comparison of stepwise, XGBoost, and clinical variables only models for predicting survival**. Wilcoxon signed-rank pairwise comparisons of time-dependent ROC AUCs at each year across 30 random train-test splits for models: stepwise (18-stepwise selected genes + clinical variables), XGBoost (8-XGBoost selected genes + clinical variables), and clinical variables only (clinical variables).

	Stepwise vs clinical variables only			**XGBoost vs clinical variables only**			**XGBoost vs stepwise**		
Year	**ΔAUC***	**95%CI**	**Adjusted p-value#**	**ΔAUC***	**95%CI**	**Adjusted p-value#**	**ΔAUC***	**95%CI**	**Adjusted p-value#**
1	-0.096	(-0.193, 0.036)	0.46	-0.03	(-0.078, 0.012)	0.33	0.059	(-0.066, 0.182)	0.87
2	0.041	(0.021, 0.063)	**0.00057**	0.019	(0.001, 0.037)	0.12	-0.022	(-0.045, 0)	0.15
3	0.031	(0.011, 0.051)	**0.015**	-0.01	(-0.027, 0.006)	0.59	-0.042	(-0.06, -0.024)	**0.00024**
4	0.067	(0.045, 0.088)	**1.50E-05**	0.025	(0.011, 0.038)	**0.006**	-0.044	(-0.062, -0.023)	**0.00027**
5	0.071	(0.049, 0.09)	**4.30E-06**	0.038	(0.022, 0.053)	**7.10E-05**	-0.032	(-0.052, -0.011)	**0.012**
6	0.071	(0.046, 0.097)	**9.70E-06**	0.053	(0.035, 0.071)	**1.50E-05**	-0.017	(-0.043, 0.01)	0.69
7	0.08	(0.059, 0.103)	**7.60E-07**	0.069	(0.049, 0.087)	**1.20E-06**	-0.011	(-0.032, 0.007)	0.74
8	0.081	(0.058, 0.105)	**7.60E-07**	0.08	(0.059, 0.101)	**2.40E-07**	0.001	(-0.021, 0.023)	1

* $\Delta AUC = \text{mean}(AUC_{Group1} - AUC_{Group2})$ for Group 1 versus Group 2;

Adjusted p-values were calculated by Bonferroni correction; statistically significant values (adjusted p-value<0.05) in bold.

We now state this in the results, as follows, lines 225-228:

"Next, we predicted survival in thirty random train-test splits of our internal dataset; compared to the clinical variables only model, the stepwise model had significantly larger AUC values for years 2 to 8, whereas XGBoost had higher AUCs for years 4 to 8 (**Fig. 4d**, **Supplementary Fig. S6a**, **Supplementary Table S4**)."

In the Grima et al. dataset,[3] we took a different approach since it was used for testing only. First, we classified the Grima et al. ALS patients as shorter-, intermediate-, or longer-surviving based on the risk scores predicted by our survival classifiers. In this analysis, including gene features in the XGBoost model improved overall survival classification as shown by maximizing median survival differences between shorter- versus intermediate- and intermediate- versus longer-surviving participants, p<0.0001 (**Fig. 4h**, new numbering), versus the clinical variables only model. Additionally, including gene features in the stepwise model improved survival classification earlier in the disease course because it maximized median survival differences between shorter- versus intermediate-surviving participants, p<0.0001 (**Fig. 4g**, new numbering), versus the clinical variables only model. In addition, for each model, we performed pairwise log-rank tests with Bonferroni-adjusted p-values (***new*** **Supplementary Fig. S6c**). The XGBoost model discriminated all pairwise comparisons while the clinical variables only models could not discriminate intermediate- from shorter-surviving participants, using an adjusted p-value cutoff of 0.05. Finally, we tested the difference between the clinical variables only model and each of the XGBoost and Stepwise models based on their concordance index (C-index) using the compareC R package. The XGBoost model (C-index=0.69) showed significantly higher C-index (p=0.05) than the clinical variables only model (C-index=0.66), however the stepwise model did not (C-index=0.65, p>0.05).

New **Supplementary Figure S6c**. Log rank p-values of pairwise comparisons among longer, intermediate, and shorter survival groups based on predictions from the three models using the external Grima et al. dataset.[3]

		Clinical variables only	Stepwise	XGBoost
Pairwise log-rank test p-value	Longer versus intermediate	0.036	0.177	0.00068
	Longer versus shorter	0.0003	0.033	1.30E-07
	Intermediate versus shorter	0.1019	0.261	0.0064
Compare C-index	C-index	0.66	0.65	0.69
	p-value (vs. Clinical variables only)		>0.05	0.05

We now state this in the results, as follows, lines 249-253:

“Overall, the XGBoost model outperformed the clinical variables only model for differentiating among survival groups in both the internal and external Grima et al. datasets (**Supplementary Fig. S6a,c**). This wider difference, reflecting a stronger overall distinction in survival classification by XGBoost, incorporated gene features, versus the clinical variables only model.”

We also changed the subtitle in lines 214:

“**A gene expression signature predicts ALS survival**”

Finally, we would also like to refer the Reviewer to our response to **Reviewer 4, comment 3**, where we demonstrate the performance of the genes only models, showing that genes only models independently contribute to survival prediction.

Reviewer 3, comment 5: Fig 4. A Venn diagram of adjusted/unadjusted DEGs would be helpful.

Our response: We fully agree with Reviewer 3 and have added the overlap between adjusted and unadjusted DEGs as ***new*** **Supplementary Fig. S16**. Please note that we have changed “unadjusted” to “primary” analysis in our revised version, per **Reviewer 3, comment 8**.

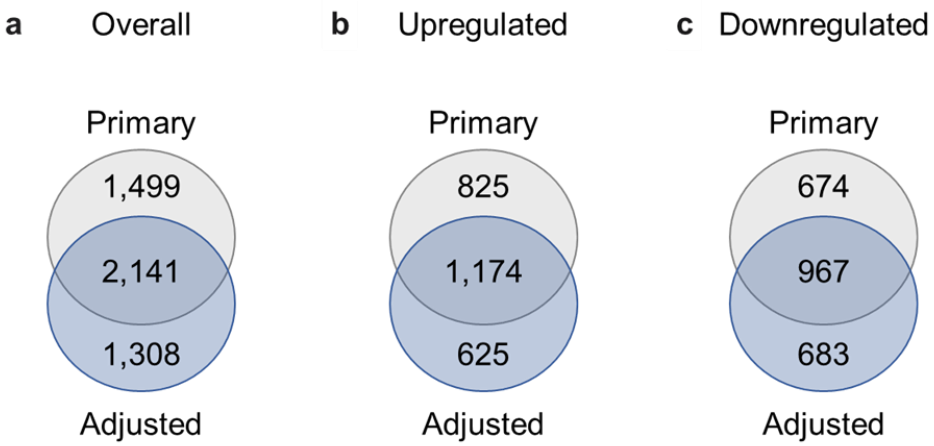
To quantitatively assess the overlap between these DEG sets, we performed Fisher’s exact tests and calculated overlap scores (Jaccard index) separately for upregulated, downregulated, and all DEGs. Each analytical model (primary and adjusted) identified hundreds of unique DEGs in both up- and downregulated directions, underscoring distinct biological insights inherent to each model. Despite these differences, we observed substantial overlap:

All DEGs (up + down): Fisher’s exact test showed a significant enrichment ($p<2.2x10^{-16}$, odds ratio = 19.0, 95%CI 17.39-20.73) with a Jaccard index of 0.43.

Upregulated DEGs: Fisher's exact test showed a significant enrichment ($p<2.2x10^{-16}$, odds ratio = 44.8, 95%CI 39.76-50.65) with a Jaccard index of 0.45.

Downregulated DEGs: Similarly, Fisher's exact test showed a significant enrichment ($p<2.2x10^{-16}$, odds ratio = 42.0, 95%CI 37.06-47.74) with a Jaccard index of 0.42.

Together, these results indicate that the DEG signatures from primary and adjusted models are statistically robust and directionally concordant.



***New* Supplementary Figure S16. Overlap of differentially expressed genes (DEGs) in ALS versus control blood samples in primary analyses and analyses adjusted for cell proportions.** Venn diagram of overlap between Ensembl IDs of primary DEGs (grey) and DEGs adjusted for cell proportions (blue) (**a**) overall, (**b**) upregulated, and (**c**) downregulated.

Additionally, we have alluded to the overlap in the results, as follows, lines 654-656:

"There was substantial overlap in DEGs for our primary analysis and adjusted for cell proportions, but the two DEGs analyses also identified hundreds of unique up- and down-DEGs (**Supplementary Fig. S16**)."

Reviewer 3, comment 6: Fig 4. As this data is based on bulk RNA Seq some of the DEGs could be driven by different cell proportions.

Our response: This is a very important point, which we completely agree with. In bulk RNA-seq, different cell proportions can indeed drive DEGs, and, consequently, pathways as well. This is why we performed two analyses in our submitted manuscript, one we called "unadjusted" and another we called "adjusted". The unadjusted (now called "primary") analysis did not include cell proportions in the linear model for identifying DEGs, whereas the adjusted analysis included proportions of CD4 T cells, CD8 T cells, monocytes, natural killers, B cells, and neutrophils, computationally estimated using the FlowSorted.Blood.EPIC R package[32] from DNAm data previously collected in this deeply phenotyped cohort. We determined that DNAm was a better method to compute cell proportions than RNA, at least in this dataset, by comparison to empirically determined cell proportions by flow cytometry. We also performed pathway enrichment of both the unadjusted (now called "primary") and adjusted DEGs, which in **Fig. 5** (new numbering, **Fig. 4** old numbering) are shown as "Enrichment of whole blood primary DEGs" and "Enrichment of whole blood DEGs adjusted for cell proportions".

Our methodology is outlined in our Materials and methods section, within the subsections "*Identification of the relative abundance of different cell types in blood (cell proportions)*", lines 577-606, and "*Differential gene expression analysis*", lines 608-656.

Please note that only pathway enrichment and drug perturbation employed both the primary and adjusted DEGs, since cell proportions influence the biological processes linked to these analyses. On the other hand, case-control and case survival prediction analyses used only the primary analysis DEGs, since prediction only needs to consider whether a biomarker can differentiate case-control or survival without considering the

underlying biology. We emphasize this in ***new* Fig. 1**, an overview of the study design, in response to **Reviewer 3, comment 12**.

Reviewer 3, comment 7: Fig 4. Are there differences based on disease status or long/short classification?

Our response: We thank the Reviewer for their insightful suggestion to evaluate pathway differences between ALS participants by shorter versus longer survival or by disease status. We first considered the longer versus shorter survivor comparison, but survival time is influenced by many factors beyond disease stage—such as age at onset, diagnostic delay, and individual variability in disease trajectory—making it an indirect and potentially heterogeneous marker of disease biology.

Therefore, to more directly capture the molecular changes associated with disease severity, we focused on ALSFRS-R scores, a clinically validated measure of functional impairment. We performed gene set enrichment analysis (GSEA) of participants in the lowest quartile of ALSFRS-R (representing more functional impairment) versus the highest quartile of ALSFRS-R scores (representing less functional impairment). We used the $-\text{Log}_{10}$(p-value) x sign(Log[fold-change]) as the statistic to rank the gene list for inputs to GSEA, where the p-values and sign(LogFC) were from our primary DEGs and from DEGs adjusted for immune cell type proportions.

We identified a set of enriched pathways that consistently distinguished low ALSFRS-R (more impairment) from high ALSFRS-R (less impairment) cases across both analyses (***new* Supplementary Fig. S11**). These shared pathways support a robust biological signal related to neurodegeneration, immune dysregulation, and metabolic stress. Key examples include:

Neurodegeneration-related pathways: Amyotrophic lateral sclerosis, Alzheimer's disease, Huntington's disease, Parkinson's disease, prion disease, and pathways of neurodegeneration – multiple diseases.

Metabolic dysfunction: Oxidative phosphorylation, glycolysis, fatty acid metabolism, carbon metabolism, adipogenesis, and diabetic cardiomyopathy.

Immune and inflammatory: Interferon gamma and alpha responses, IL6-JAK-STAT3 signaling, IL2-STAT5 signaling, allograft rejection, complement cascade, inflammatory response, and phagosome.

Stress and tissue remodeling: Mtorc1 signaling, hypoxia, epithelial-mesenchymal transition, cytoskeleton in muscle cells, and ECM-receptor interaction.

Notably, a few pathways from the primary analysis model, such as apoptosis, PI3K-AKT-mTOR signaling, and necroptosis, may still reflect biologically meaningful processes related to neuronal cell death and survival, and thus warrant further investigation. However, their absence in the adjusted analysis suggests that their signal may be partially confounded by cell-type heterogeneity rather than reflecting disease-intrinsic transcriptomic regulation.

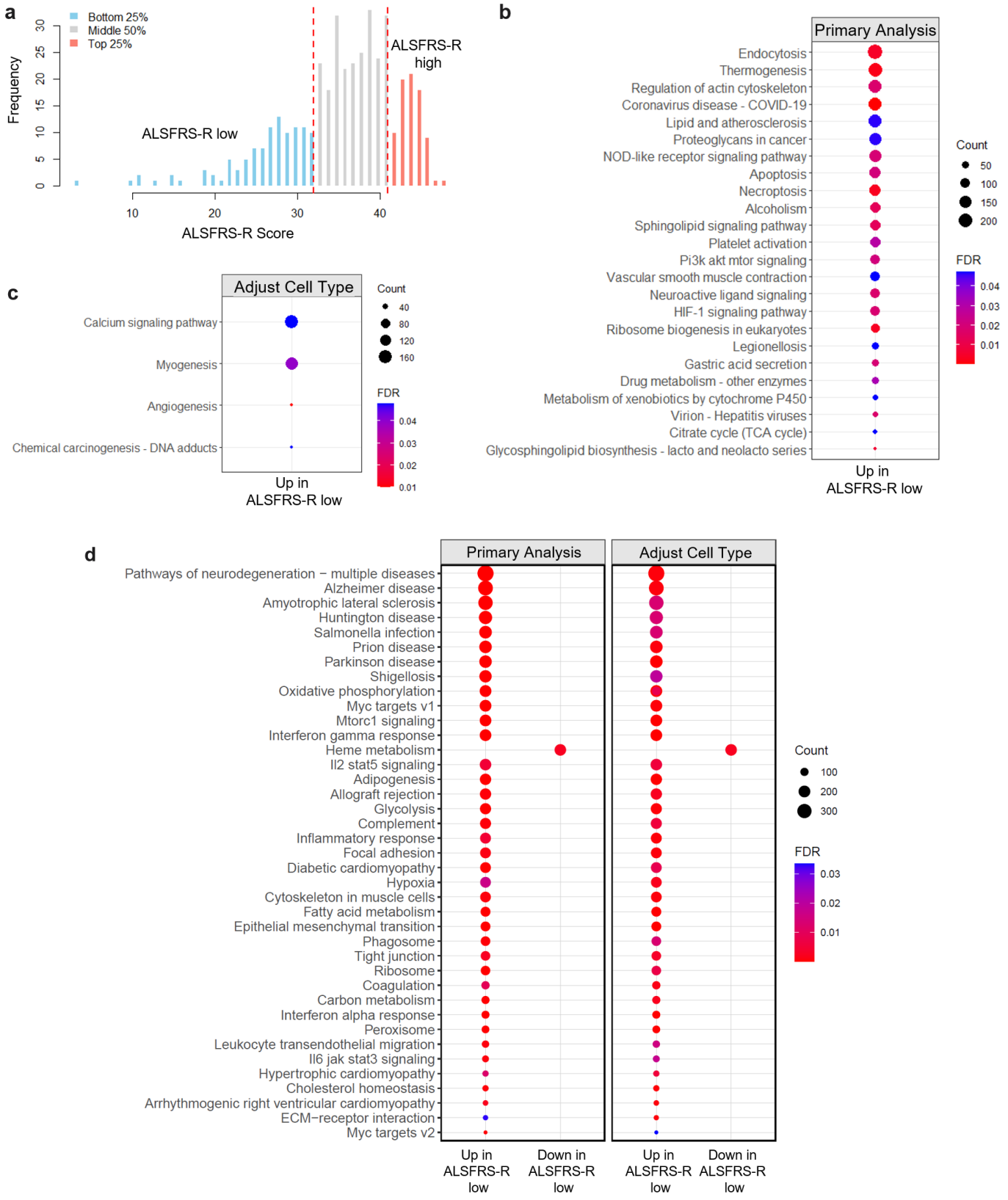
Overall, the dysregulated pathways associated with ALSFRS-R score support a robust biological signal related to neurodegeneration, immune dysregulation, and metabolic stress.

This analysis is now presented in the results, as follows, lines 297-302:

"We also examined pathway enrichment of primary DEGs and DEGs adjusted for cell proportions by disease severity by analyzing the lowest (most functionally impaired) versus highest (least functionally impaired) quartile of ALSFRS-R scores. Overall, we found pathways linked to more severe disease reflected in lower ALSFRS-R scores were related to neurodegeneration, immune dysregulation, and metabolic stress (**Supplementary Fig. S11**), key ALS pathophysiological processes."

We have added to the methods, as follows, lines 767-772:

"Gene set enrichment analysis was performed on participants in the lowest quartile of ALSFRS-R (most functionally impaired) versus the highest quartile of ALSFRS-R scores (least functionally impaired). $-\text{Log}_{10}$(p-value) x sign(Log[fold-change]) was used as the statistic to rank the gene list for inputs to gene set enrichment analysis, where p-values and sign(LogFC) were from primary DEGs and DEGs adjusted for cell type proportions."



***new* Supplementary Fig. S11. Pathway analysis by gene set enrichment analysis of the blood transcriptome in lower versus higher ALSFRS-R scores from ALS participants.**

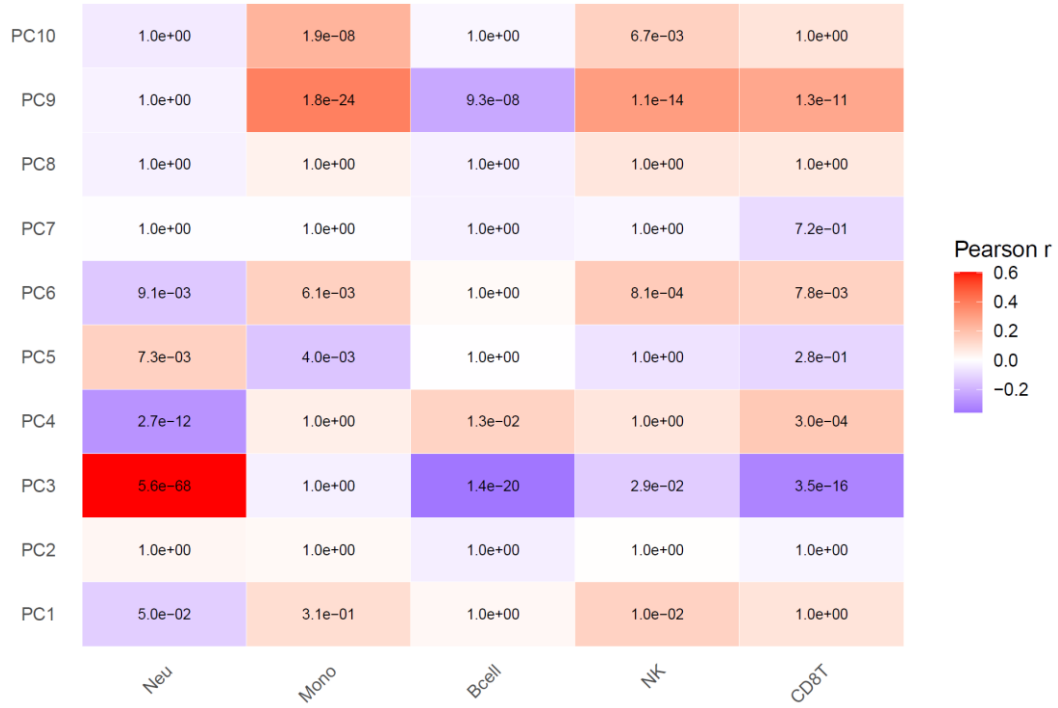
(**a**) Lower ALSFRS-R scores (n=107, ALSFRS-R≤32, bottom 25%) and higher ALSFRS-R scores (n=112, ALSFRS-R≥41, top 25%) from ALS participants based on ALSFRS-R score at first clinic visit. Pathways unique to enrichment analysis of (**b**) primary DEGs and (**c**) DEGs adjusted for cell type proportions. (**d**) Top 30 pathways shared by enrichment analysis of both primary DEGs and (**c**) DEGs adjusted for cell type proportions.

Reviewer 3, comment 8: Fig 4. Authors use the same RNA PCs in both their adjusted/unadjusted model. They should check whether any of the RNA PCs correlate with cell abundance (and thus would already be adjusted for in their model).

Our response: We appreciate the Reviewer's comment and have clarified our use of RNA principal components (PCs) across models. Specifically:

- For the model adjusted for cell proportions, we included 8 RNA PCs (PC_2, PC_4–PC_{10}). We excluded PC_1 and PC_3 because both correlated to case-control status after adjusting for cell proportions.
- For the primary model, we included 9 RNA PCs (PC_2–PC_{10}), excluding only PC_1, which correlated to case-control status.

In response to the Reviewer's suggestion, we evaluated whether any of the RNA PCs significantly correlated to estimated cell proportions (neutrophils, natural kill cells [NKs], CD8 T cells, monocytes, B cells). Pearson correlation identified seven PC–cell type pairs that met the criteria |r|>0.25 and Bonferroni-adjusted p<0.05 (***new* Supplementary Fig. S15a**). So, indeed, three of the PCs (3, 4, and 9) are estimated to account for a portion, but not all, of the cell type proportions effect (***new* Supplementary Fig. S15b**).

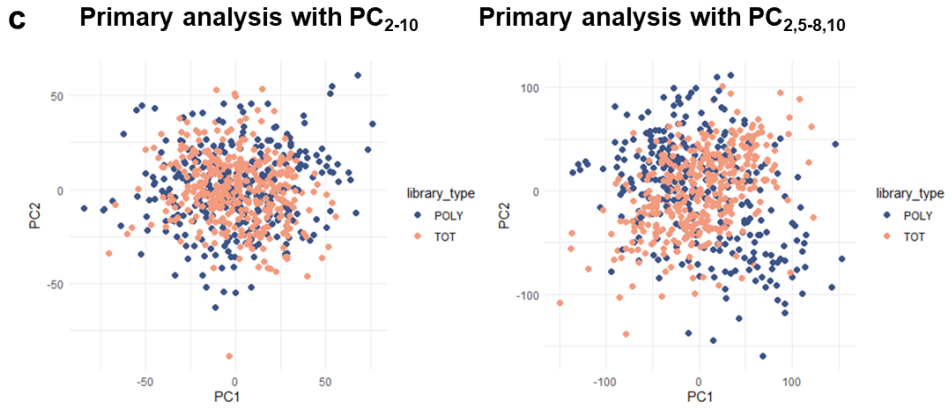


***new* Supplementary Figure S15a**. Heatmap of Pearson's correlation results between RNA PCs and cell type proportions. Legend and box colors indicate Pearson's correlation coefficients; box numbers indicate Bonferroni-adjusted p-values.

***new* Supplementary Figure S15b**. RNA PCs that correlate to cell type proportions.

Principal components	Cell types	r	Adjusted p-values
PC_3	Neutrophils	0.602098	5.56E-68
PC_3	B cells	-0.35685	1.43E-20
PC_3	CD8 T cells	-0.31906	3.46E-16
PC_4	Neutrophils	-0.28032	2.68E-12
PC_9	Monocytes	0.386587	1.85E-24
PC_9	NK cells	0.304874	1.07E-14
PC_9	CD8 T cells	0.272732	1.33E-11

We tried removing the PCs that were correlated with cell type proportions (PC_3, PC_4, & PC_9) but this reintroduced batch effects between library preparation types (***new* Supplementary Fig. S15c**), so we returned to the original model.



***new* Supplementary Figure S15c**. Principal component analysis (PCA) of gene expression residuals from (left) linear model with RNA $PC_{2\text{-}10}$ and (right) linear model with RNA PC_2, $PC_{5\text{-}8}$, and PC_{10}. Each dot is a sample colored by library preparation type, polyA selection (blue), rRNA depletion (orange).

Performing the same analyses without PC_3, PC_4, and PC_9 (i.e., reintroducing some batch effects, but without accounting for any cell type proportions) modestly shifted DEGs detection and minimally impacted pathway results, compared to the original model. For upregulated genes, 1,693 DEGs overlapped, with 306 unique to the primary analysis and 60 unique to the model without PC_3, PC_4, and PC_9; for downregulated genes, 1,412 overlapped, with 229 and 144 unique to each model, respectively. Pathway comparisons showed 24 shared pathways, with only minor differences in pathways gained or lost between models. These results suggest that while some modest changes occurred, the overall biological interpretation remains consistent.

This is now summarized in the methods, as follows, lines 647-654:

“Correlation analysis between RNA PCs and estimated cell proportions (neutrophils, NKs, CD8 T cells, monocytes, B cells) by Pearson correlation identified seven PC–cell type pairs that met the criteria $|r|>0.25$ and Bonferroni-adjusted $p<0.05$ (**Supplementary Fig. S15a-b**). Removing the PCs that correlated with cell type proportions (PC_3, PC_4, & PC_9) reintroduced batch effects between library preparation types (**Supplementary Fig. S15c**); therefore, the original model was retained, called the “primary” analysis, which accounted, in part only, for cell proportions, in contrast to the analysis that adjusted for the remainder of cell proportions (i.e., referred to as adjusted for cell proportions).”

Reviewer 3, comment 9: Fig 5. Unclear what the contribution of the whole blood transcriptomic dataset is to the drug perturbation analysis relative to the iPSC neurons and spinal cord datasets. Using core genes from each would clarify this concern and increase the confidence in the potential utility of these so-called therapeutic candidates.

Our response: We appreciate the Reviewer’s comments regarding core genes. Indeed, we completely agree with the Reviewer and this is the methodology we adopted. We identified “core genes” from the overlap of our blood transcriptomics DEGs (primary and adjusted for cell proportions) with iPSC-derived TDP-43 knockdown neurons and postmortem spinal cord DEGs. We called these “core genes” as the DEGs that were shared by blood with these two ALS-related cells/tissues, and, thus, were more related to the ALS disease process. We then used these “core genes” for the drug perturbation analysis to increase our odds of identifying therapeutic candidates with potential efficacy across all tissues, i.e., targeting the overall ALS disease process. We now realize that this was not clear for readers since the approach was buried in supplementary figures. We have now moved these figures to a ***new* Fig. 6a-d** to better illustrate our methodology to readers. We are very thankful Reviewer 3 brought this issue to our attention.

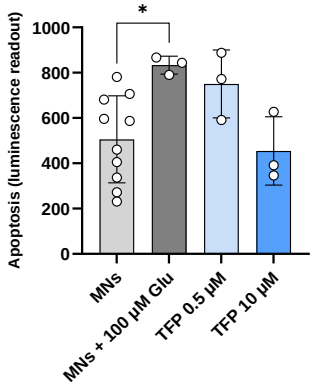
Reviewer 3, comment 10: Fig 5. Assessing the effects of 1-2 of these drugs on an in vitro model of ALS pathology is critically important to substantiate the utility of the in-silico analysis.

Our response: We thank Reviewer 3 for this excellent suggestion to validate our drug perturbation analysis. We employed our previously developed *in vitro* ALS model of primary motor neurons (MNs) with glutamate-induced excitotoxicity[33,34] to test trifluoperazine (TFP) identified by our drug perturbation analysis. TFP is a typical antipsychotic primarily used to treat schizophrenia, a neuropsychiatric disorder that may overlap with ALS.[10] Trifluoperazine is also a potent allosteric modulator of the human purinergic P2X7 receptor.[5] P2X7 receptor antagonism has been proposed as an ALS therapeutic strategy by modulating neuroinflammation.[6]

Briefly, on day 1, spinal cords were collected from E15 Sprague-Dawley rat embryos and perineural membranes were removed in L-15 media. Spinal cord tissue was minced, digested in trypsin, triturated, and a cellular gradient was established in Opti-Prep at 2,000 rpm (cat# D1556, Sigma). The top layer was collected, rinsed in L-15 media, and plated on poly-lysine coated wells in 96-well plates in MN plating media (MN feed media + L-glutamine). MNs were cultured in MN feed media (MN nutrient stock, NeuroBasal media, B-27, pen-strep) until day 7 when experiments were conducted in MN treatment media (MN feed media without B-27). Conditions included MNs alone, MNs with 100 µM glutamate for 6 h (ALS model), and MNs co-treated with 100 µM glutamate and 0.5 and 10 µM TFP for 6 h (cat# 15068, Cayman Chemical, Ann Arbor MI, reconstituted in DMSO at 20 mg/mL). Apoptosis was measured using a bioluminescence assay (cat# G8091, Caspase-Glo 3/7 Assay) on a plate reader.

TFP dose-dependently protected MNs from glutamate-induced excitotoxicity; at a TFP dose of 10 µM, the concentration we performed drug perturbation at, the extent of apoptosis did not differ from the MN control condition. Thus, this preliminary *in vitro* experiment supports the results of our drug perturbation analysis for TFP. Primary MN culture is work intensive and utilizes rat embryos; we are currently optimizing a screening approach in iNeurons, as we previously derived from ALS patient-derived inducible pluripotent stem cells.[35] We will be expanding on this future research direction.

Response Figure R4. TFP dose-dependently protects MNs from glutamate-induced excitotoxicity. MNs treated with TFP (0.5 and 10 µM) did not differ in extent of apoptosis to control MNs. Kruskal-Wallis with Dunn's multiple comparisons test versus MNs control. *p<0.05.



Reviewer 3, comment 11: Discussion. This section is well written and informative and the addition of a limitations and strengths section is great for the reader.

Our response: We thank Reviewer 3 for their favorable assessment of our manuscript's discussion section, which we hope contextualized our work relative to the current blood transcriptomics, diagnosis, and prognosis landscape in ALS.

Reviewer 3, comment 12: Figure 2a isn't an effective schematic. No idea what it means

Our response: We thank the Reviewer from bringing this to our attention. We asked our visualization expert to generate a new schematic, which we hope will be more informative. This ***new*** **Figure 1** shows the study

overview, and also clarifies the process for filtering DEGs and the resulting gene panels from **Figure 3a** (new number, **Figure 2a** old numbering). Additionally, we have updated **Figure 3a** (new number, **Fig. 2a** old numbering) itself to better explain our process for filtering DEGs, resulting in the 27-, 30-, and 29-gene panels.

The ***new*** **Figure 1** legend is as follows, lines 1,090-1,109:

"(**a**) Blood samples collected from cases and controls were processed for RNA-seq. (**b**) Differentially expressed genes (DEGs) in ALS cases versus controls were identified using both a primary (unadjusted) and an adjusted analysis for immune cell proportions. (**c**) (i) ALS case-control classifier: Comparison of seven machine learning (ML) algorithms (Part 1); Improving the performance of the best-performing ML algorithm XGBoost (Part 2); Testing the three gene panels plus a combined 46-gene panel on our internal and the external Grima et al.[3] datasets (Parts 3 and 4, respectively). Results shown in **Fig. 3**. (ii) ALS case survival classifier: Criteria applied for gene filtering (Part 1); Training two ML algorithms, XGBoost and stepwise, using gene features with clinical variables (Part 2); Comparing XGBoost and stepwise (gene features + clinical variables) to a clinical variables only model in our internal and the external Grima et al.[3] datasets (Parts 3 and 4, respectively). Results shown in **Fig. 4**. (iii) Pathway analysis to identify ALS disease signature in blood: Kyoto Encyclopedia of Genes and Genomes, Hallmark, and Gene Ontology pathway enrichment of primary DEGs (Part 1); Analyses on DEGs adjusted for cell proportions (Part 2); Pathway analyses of adjusted versus primary DEGs (Parts 3, 4). Results shown in **Fig. 5**. (iv) Drug perturbation analysis to identify ALS therapeutic candidates: Identification of "core genes" of adjusted and primary DEGs that overlap with DEGs from induced pluripotent stem cell (iPSC)-derived neurons with TDP-43 knockdown and ALS postmortem spinal cord (Part 1); Drug perturbation analysis of "core genes" (Part 2). Results shown in **Fig. 6**. Created in BioRender.com."

Reviewer 3, comment 13: Line 194: "Therefore, our classifier showed potential as a potential" is confusing

Our response: We apologize for the repetitive use of "potential" and have deleted the duplicate.

Finally, we would like to thank Reviewer 3 for their thoughtful review. Addressing Reviewer 3's comments has significantly strengthened the presentation of our data and the new requested analyses expanded and markedly improved the scope of our manuscript.

Reviewer 4

Review for "Gene Expression Signatures from Whole Blood Predict Amyotrophic Lateral Sclerosis Case Status and Survival"

Summary

This manuscript presents a promising analysis of gene expression signatures in whole blood for predicting ALS status and patient survival. The authors have assembled a substantial cohort and integrated blood data with iPSC-neuron and postmortem spinal cord datasets to develop potentially valuable prediction models. While the topic is clinically important and the approach innovative, there are several methodological concerns that should be addressed to strengthen the reliability of the findings.

Strengths

The study demonstrates several notable strengths, including a large sample size (225 ALS cases, 119 controls) providing adequate statistical power. The integration of blood signatures with CNS-relevant tissues offers important biological context. The authors show thoughtful consideration of cell type proportions in their analysis. The comprehensive pathway enrichment analyses further enhance the study's value.

Recommendation

This work addresses an important clinical need in ALS biomarker development and shows considerable promise. Addressing the methodological concerns outlined above would significantly strengthen the reliability of the findings and enhance their potential impact on clinical practice. I look forward to seeing a revised version that builds on the study's strengths while implementing these improvements.

Our response: We thank Reviewer 4 for their favorable assessment of our manuscript and for their comments and suggestions, which we address point-by-point below, and have helped us significantly improve the manuscript for the *Nature Communications* readership.

Reviewer 4, comment 1: Model development methodology: The gene panel derivation occurs on the same dataset used for model training, which may lead to inflated performance estimates. From a ML perspective identifying differential expression is feature selection. I suggest implementing proper nested cross-validation where feature selection occurs independently within each fold or clearly separating discovery and validation cohorts before any feature selection (i.e. perform differential expression on each CV split).

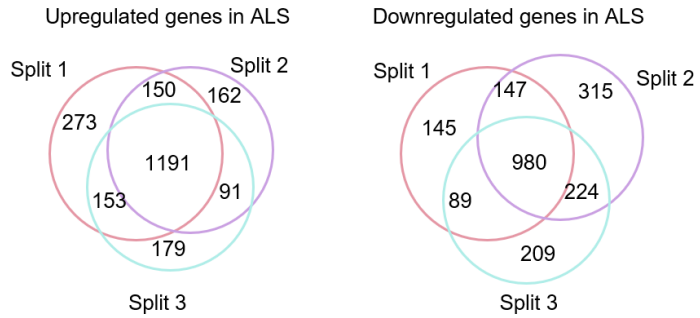
Our response: We appreciate the Reviewer's suggestion and have separated the discovery and validation cohorts before any feature selection. Specifically, to address this comment, we randomly split our internal dataset into discovery (70%) and validation (30%) subsets three independent times. Differential expression analysis was conducted within each discovery subset, and the resulting DEGs were compared across splits. Between 67-75% of upregulated and 59-72% of downregulated DEGs were shared across all three splits (**Response Figure R5a**).

Using the DEGs from random split #1, we trained three XGBoost models based on progressively stricter DEG filters: (i) FDR<0.01 (4,972 DEGs; 29-gene panel), (ii) FDR<0.01 with AvExpr>0 (4,789 DEGs; 28-gene panel), and (iii) FDR<0.01 with AvExpr>2 (4,414 DEGs, 26-gene panel). Mean 10-fold cross validation ROC-AUCs in the discovery subsets ranged from 96.27% to 97.03%, while ROC-AUCs in the corresponding validation subsets ranged from 94.81% to 95.01% (**Response Figure R5b**). These values closely match the 10-fold cross validation performance obtained when our entire internal data was used for training, demonstrating minimal overfitting caused by using DEGs from all 100% of our samples. Furthermore, they surpass performance on the external Grima et al.[3] cohort as expected, demonstrating the robustness of our feature selection and model building pipeline.

To maximize predictive power, we retained the original XGBoost models, which were trained on our full internal dataset (our largest available training set), and validated them on the external Grima et al.[3] RNA-seq dataset. We further evaluated these models on two additional external microarray datasets (GSE112676 and GSE112680 from van Rheenen et al.[1]/Swindell et al.[2]), now presented in ***new*** **Supplementary Figure S5** (refer to **Reviewer 1, comment 14**). Despite platform differences (RNA-seq versus microarray) and missing 12 out of 46 gene features, our XGBoost models achieved informative ROC-AUCs (65-75%) and consistent performance across multiple metrics. Taken together with the consistent DEG overlap observed in all random splits, these results demonstrate that our feature-selection strategy and XGBoost models generalize well beyond the training data.

Response Figure R5. **Feature selection and performance training with 70% of UM dataset.** (a) Overlap of ALS versus control DEGs across three random discovery subsets from 70/30 splits. (b) Predictive performance of the three gene panels derived from split #1: 10-fold cross validation within the 70% discovery subset (left) and validation on the held-out 30% subset (right).

a



b

10-fold cross-validation metrics from discovery cohort (70% of internal data, 296 ALS + 191 control)					Metrics from validation cohort (30% of internal data, 126 ALS + 81 control)					
# genes	Accuracy	Sensitivity	Specificity	ROC-AUC	Accuracy	Sensitivity	Specificity	ROC-AUC	Balanced accuracy	F1
29	90.11%	92.41%	86.33%	96.78%	87.08%	85.83%	89.02%	95.01%	87.43%	88.98%
28	90.72%	93.60%	85.61%	97.03%	88.04%	88.19%	87.80%	94.84%	88.00%	89.96%
26	88.87%	91.95%	83.25%	96.27%	88.04%	88.19%	87.80%	94.81%	88.00%	89.96%

Reviewer 4, comment 2: Performance evaluation metrics: Reliance on AUC alone may not fully characterize model performance given the dataset imbalance. I suggest including PR-curves, precision, recall, F1 scores, and balanced accuracy metrics to provide a more comprehensive assessment of model performance, particularly when comparing to the external dataset.

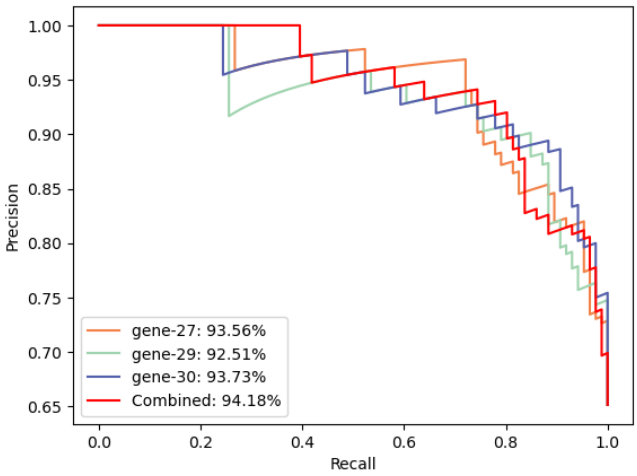
Our response: We appreciate the Reviewer's insightful suggestion regarding the limitations of AUC in the context of imbalanced datasets. We expanded our performance evaluation to include a broader set of metrics that provide a more nuanced view of model behavior. Specifically, we now report:

- Precision, recall (same as sensitivity), and F1 scores to assess the trade-off between false positives and false negatives,
- Balanced accuracy, which accounts for the performance of both the minority and majority classes, and
- Precision-recall curves (PR-curves) to complement ROC curves in highlighting performance under class imbalance.

The inclusion of these additional metrics confirms the robustness of our models and supports the consistency of performance when applied to the external Grima et al.[3] dataset (**Response Table R3** and **Figure R4**). The metrics from **Response Table R3** have been incorporated into main text **Figure 3f**.

Response Table R3. Performance metrics for predicting case status by the three gene panels plus the combined panel on the external Grima et al.[3] dataset.

Gene panel	Accuracy	Sensitivity	Specificity	Precision	Balanced accuracy	F1	ROC-AUC
27	75.00%	77.91%	69.57%	86.59%	73.74%	84.52%	87.23%
30	84.09%	86.05%	80.43%	88.37%	83.24%	88.37%	89.21%
29	80.30%	82.56%	76.09%	88.10%	79.33%	87.06%	87.69%
46	81.06%	82.56%	78.26%	87.65%	80.41%	85.03%	89.43%



Response Figure R6. Precision-recall curves and area under the precision-recall curves (PR-AUC) for predicting case status using three individual gene panels and the combined panel in the external Grima dataset.

Reviewer 4, comment 3: Model comparison framework: Current comparisons between clinical-only and clinical-only+genes-combined models (e.g. Figure 3) don't isolate the independent contribution of gene signatures. I suggest adding explicit genes only models to quantify the independent predictive value of the gene signatures beyond clinical variables alone.

Our response: We appreciate the Reviewer's suggestion to assess the independent contribution of the gene signatures in ALS survival prediction. To address this, we trained two additional genes only CoxPH models using the 18 stepwise- and 8 XGBoost-selected genes, which we evaluated on both the internal test data from five random train-test splits and one independent external Grima et al.[3] dataset.

In our cohort from the thirty 80/20 train-test splits, at year 4, the mean (±SD) AUCs were:

Clinical variables only model: 0.684 ± 0.067

18 stepwise genes (no clinical variables): 0.704 ± 0.062

Clinical variables + 18 stepwise genes: 0.748 ± 0.060

8 XGBoost genes (no clinical variables): 0.624 ± 0.069

Clinical variables + 8 XGBoost genes: 0.705 ± 0.067

Across years 2 to 8, the clinical + gene models consistently outperformed the clinical variables only model, as we also detail in our response to **Reviewer 3, comment 4**.

Furthermore, the genes only models also provided independent predictive value over all time points. In the independent external Grima et al.[3] dataset, the genes only XGBoost model showed moderate predictive ability (C-index 0.62; 4-year AUC 0.65), suggesting the model retains some independent signal. In contrast, the genes only stepwise model yielded poor predictive performance (C-index 0.51; 4-year AUC 0.56), indicating limited independent prognostic value. However, overall, gene signatures provide independent prognostic value and offer meaningful gains in predictive performance when integrated with clinical features.

This is stated in the results, as follows, lines 225-230:

"Next, we predicted survival in thirty random train-test splits of our internal dataset; compared to the clinical variables only model, stepwise had significantly larger AUC values for years 2 to 8, whereas XGBoost had higher AUCs for years 4 to 8 (**Fig. 4d**, **Supplementary Fig. S6a**, **Supplementary Table S4**). In fact, gene only models using the 18 stepwise- and 8 XGBoost-selected gene features significantly contributed to survival prediction independent of the clinical variables alone (**Supplementary Fig. S6b**)."

New **Supplementary Figure S6b**. Time-dependent ROC AUC performance for our dataset using 30 random train-test splits. Shaded green indicates year 4.

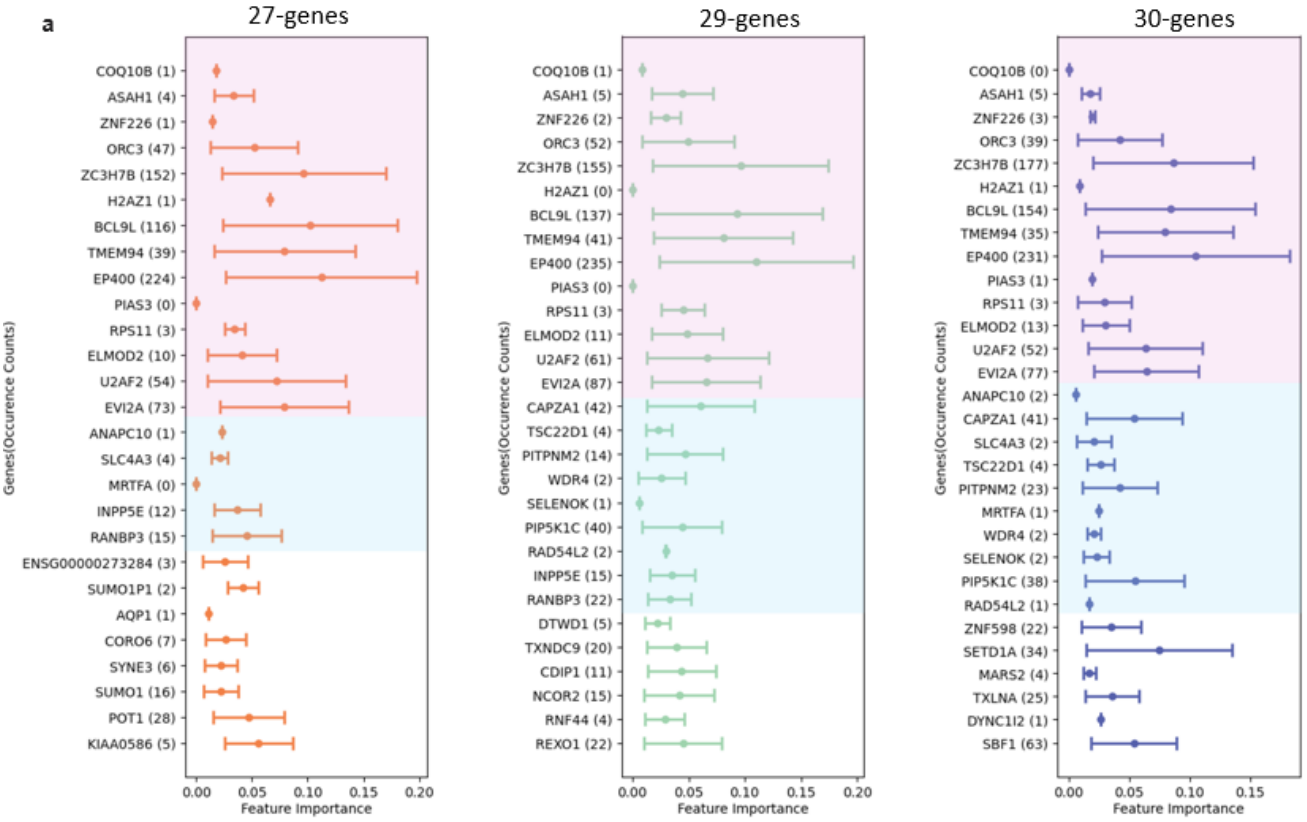
Time	Clinical variables only model		18 stepwise genes + clinical variables		8 XGBoost genes + clinical variables		18 stepwise genes only		10 XGBoost genes only	
Year	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
2	0.675	0.074	0.717	0.068	0.693	0.067	0.660	0.058	0.637	0.071
3	0.686	0.059	0.716	0.057	0.674	0.059	0.655	0.058	0.582	0.057
4	0.684	0.067	0.748	0.060	0.705	0.067	0.704	0.062	0.624	0.069
5	0.664	0.072	0.731	0.063	0.699	0.069	0.690	0.062	0.634	0.072
6	0.668	0.079	0.736	0.055	0.721	0.080	0.699	0.056	0.673	0.082
7	0.675	0.089	0.754	0.070	0.743	0.087	0.740	0.063	0.734	0.084
8	0.670	0.090	0.750	0.071	0.751	0.083	0.732	0.066	0.750	0.081

Reviewer 4, comment 4: Feature selection stability: The inconsistency in top features across different models (Fig 2b,c,d,e) raises questions about signature robustness. I suggest implementing bootstrap analysis to assess feature selection stability and provide confidence intervals for feature importance rankings.

Our response: We thank the reviewer for this suggestion to assess the feature stability. We performed 500 bootstraps on the training data, and based on the results, we added the confidence intervals of feature importance for each gene into our case-control classifiers (**Response Figure R7a**). The top 14 genes against a pink background are the common genes in all three classifiers, the middle genes against a blue background are the genes shared by two classifiers, and the bottom genes against a white background are unique genes to each classifier. The number inside the bracket next to each gene name is the number of times this gene was selected during the 500 bootstrap simulations. Overall, common genes tended to have higher importance score and tended to be more frequently selected by the model.

Taking both feature importance and feature selection robustness into consideration, we further built a robust combined model (**Response Figure R7b**). For each gene panel below, we selected the top 20 genes that were selected in 2 or more gene panels (pink or blue background), calculated their combined ranks (feature importance rank + selection frequency rank), and selected the top 35 based on the combined rank. We further trained a new robust 35-gene classifier by XGBoost, and tested its performance by 10-fold cross-validation in our internal dataset and external Grima et al. dataset. Notably, the new robust combined model did not outperform the original classifiers (DeLong's test $p>0.05$, **Response Figure R7c**), so we decided to retain our original classifiers.

a



b

	Internal							External						
Classifier	Accuracy	ROC-AUC	Recall	Specificity	F1	Precision	Balance Accuracy	Accuracy	ROC-AUC	Recall	Specificity	F1	Precision	Balance Accuracy
Robust-35-genes	92.93%	97.89%	95.49%	88.78%	94.24%	93.16%	92.14%	84.09%	90.42%	87.21%	78.26%	87.72%	88.24%	82.74%

c

Comparison	Z-score	P-value
Robust-C vs. Genes-27	-1.667	0.0955
Robust-C vs. Genes-29	-1.767	0.0772
Robust-C vs. Genes-30	-0.8497	0.3955
Robust-C vs. Combined	-0.8295	0.4068

Response Figure R7. (**a**) Plot of median feature importance score along with confidence intervals generated based on 5% and 95% importance scores across 500 bootstraps, for each of the three gene classifiers. (**b**) Performance metrics of the "robust" model evaluated by internal 10-fold cross-validation and the external Grima et al. dataset. (**c**) DeLong's test z-scores and p-values for comparisons between the "robust" model and the four models in our manuscript.

Reviewer 4, comment 5: The authors should make the underlying Code available to enhance reproducibility and transparency.

Our response: We thank the Reviewer for highlighting the importance of reproducibility. We have made the code available on our GitHub site for this manuscript, reflected in our code availability statement, added after the data availability statement, now reads: "Code has been made available at https://github.com/yzhao80/ALS_RNA."

We thank Reviewer 4 for their thoughtful review and helpful suggestions, which have strengthened our manuscript.

References

1. van Rheenen W, Diekstra FP, Harschnitz O, et al. Whole blood transcriptome analysis in amyotrophic lateral sclerosis: A biomarker study. *PLoS One*. 2018;13(6):e0198874. doi:10.1371/journal.pone.0198874
2. Swindell WR, Kruse CPS, List EO, Berryman DE, Kopchick JJ. ALS blood expression profiling identifies new biomarkers, patient subgroups, and evidence for neutrophilia and hypoxia. *J Transl Med*. May 22 2019;17(1):170. doi:10.1186/s12967-019-1909-0
3. Grima N, Liu S, Southwood D, et al. RNA sequencing of peripheral blood in amyotrophic lateral sclerosis reveals distinct molecular subtypes: Considerations for biomarker discovery. *Neuropathol Appl Neurobiol*. Dec 2023;49(6):e12943. doi:10.1111/nan.12943
4. Kadena K, Ouzounoglou E. Drug Repurposing for Amyotrophic Lateral Sclerosis Based on Gene Expression Similarity and Structural Similarity: A Cheminformatics, Genomic and Network-Based Analysis.
5. Hempel C, Nörenberg W, Sobottka H, et al. The phenothiazine-class antipsychotic drugs prochlorperazine and trifluoperazine are potent allosteric modulators of the human P2X7 receptor. *Neuropharmacology*. Dec 2013;75:365-79. doi:10.1016/j.neuropharm.2013.07.027
6. Ruiz-Ruiz C, Calzaferri F, García AG. P2X7 Receptor Antagonism as a Potential Therapy in Amyotrophic Lateral Sclerosis. *Front Mol Neurosci*. 2020;13:93. doi:10.3389/fnmol.2020.00093
7. Zheng C, Li W, Ali T, et al. Ibrutinib Delays ALS Installation and Increases Survival of SOD1(G93A) Mice by Modulating PI3K/mTOR/Akt Signaling. *J Neuroimmune Pharmacol*. Sep 2023;18(3):383-396. doi:10.1007/s11481-023-10068-9
8. Goutman SA, Hardiman O, Al-Chalabi A, et al. Emerging insights into the complex genetics and pathophysiology of amyotrophic lateral sclerosis. *Lancet Neurol*. Mar 22 2022;doi:10.1016/s1474-4422(21)00414-2
9. Duan QQ, Wang H, Su WM, et al. TBK1, a prioritized drug repurposing target for amyotrophic lateral sclerosis: evidence from druggable genome Mendelian randomization and pharmacological verification in vitro. *BMC Med*. Mar 5 2024;22(1):96. doi:10.1186/s12916-024-03314-1
10. McLaughlin RL, Schijven D, van Rheenen W, et al. Genetic correlation between amyotrophic lateral sclerosis and schizophrenia. *Nat Commun*. Mar 21 2017;8:14774. doi:10.1038/ncomms14774
11. Stommel EW, Graber D, Montanye J, Cohen JA, Harris BT. Does treating schizophrenia reduce the chances of developing amyotrophic lateral sclerosis? *Med Hypotheses*. 2007;69(5):1021-8. doi:10.1016/j.mehy.2007.02.041
12. Cortes-Flores H, Torrandell-Haro G, Brinton RD. Association between CNS-active drugs and risk of Alzheimer's and age-related neurodegenerative diseases. *Front Psychiatry*. 2024;15:1358568. doi:10.3389/fpsyt.2024.1358568
13. D'Ovidio F, d'Errico A, Farina E, Calvo A, Costa G, Chiò A. Amyotrophic Lateral Sclerosis Incidence and Previous Prescriptions of Drugs for the Nervous System. *Neuroepidemiology*. 2016;47(1):59-66. doi:10.1159/000448618
14. Imamura K, Izumi Y, Watanabe A, et al. The Src/c-Abl pathway is a potential therapeutic target in amyotrophic lateral sclerosis. *Sci Transl Med*. May 24 2017;9(391)doi:10.1126/scitranslmed.aaf3962
15. Imamura K, Izumi Y, Nagai M, et al. Safety and tolerability of bosutinib in patients with amyotrophic lateral sclerosis (iDReAM study): A multicentre, open-label, dose-escalation phase 1 trial. *EClinicalMedicine*. Nov 2022;53:101707. doi:10.1016/j.eclinm.2022.101707
16. Palomo V, Nozal V, Rojas-Prats E, Gil C, Martinez A. Protein kinase inhibitors for amyotrophic lateral sclerosis therapy. *Br J Pharmacol*. Mar 2021;178(6):1316-1335. doi:10.1111/bph.15221
17. Zhou F, Liu Y, Liu D, Xie Y, Zhou X. Identification of basement membrane-related signatures for estimating prognosis, immune infiltration landscape and drug candidates in pancreatic adenocarcinoma. *J Cancer*. 2024;15(2):401-417. doi:10.7150/jca.89665

18. Kondapuram SK, Coumar MS. Pan-cancer gene expression analysis: Identification of deregulated autophagy genes and drugs to target them. *Gene*. Nov 30 2022;844:146821. doi:10.1016/j.gene.2022.146821

19. Erb ML, Moore DJ. LRRK2 and the Endolysosomal System in Parkinson's Disease. *J Parkinsons Dis*. 2020;10(4):1271-1291. doi:10.3233/jpd-202138

20. Moore AS, Holzbaur EL. Dynamic recruitment and activation of ALS-associated TBK1 with its target optineurin are required for efficient mitophagy. *Proc Natl Acad Sci U S A*. Jun 14 2016;113(24):E3349-58. doi:10.1073/pnas.1523810113

21. Kubat Oktem E, Aydin B, Yazar M, Arga KY. Integrative Analysis of Motor Neuron and Microglial Transcriptomes from SOD1(G93A) Mice Models Uncover Potential Drug Treatments for ALS. *J Mol Neurosci*. Nov 2022;72(11):2360-2376. doi:10.1007/s12031-022-02071-1

22. Shefner JM, Al-Chalabi A, Baker MR, et al. A proposal for new diagnostic criteria for ALS. *Clin Neurophysiol*. Aug 2020;131(8):1975-1978. doi:10.1016/j.clinph.2020.04.005

23. Hannaford A, Pavey N, van den Bos M, et al. Diagnostic Utility of Gold Coast Criteria in Amyotrophic Lateral Sclerosis. *Ann Neurol*. May 2021;89(5):979-986. doi:10.1002/ana.26045

24. Pugdahl K, Camdessanché JP, Cengiz B, et al. Gold Coast diagnostic criteria increase sensitivity in amyotrophic lateral sclerosis. *Clin Neurophysiol*. Sep 8 2021;doi:10.1016/j.clinph.2021.08.014

25. Staats KA, Schönefeldt S, Van Rillaer M, et al. Beta-2 microglobulin is important for disease progression in a murine model for amyotrophic lateral sclerosis. *Front Cell Neurosci*. 2013;7:249. doi:10.3389/fncel.2013.00249

26. Verheijen MH, Peviani M, Hendricusdottir R, et al. Increased axonal ribosome numbers is an early event in the pathogenesis of amyotrophic lateral sclerosis. *PLoS One*. 2014;9(1):e87255. doi:10.1371/journal.pone.0087255

27. Poon HF, Hensley K, Thongboonkerd V, et al. Redox proteomics analysis of oxidatively modified proteins in G93A-SOD1 transgenic mice--a model of familial amyotrophic lateral sclerosis. *Free Radic Biol Med*. Aug 15 2005;39(4):453-62. doi:10.1016/j.freeradbiomed.2005.03.030

28. Jeon HJ, You SY, Park YS, Chang JW, Kim JS, Oh JS. TCTP regulates spindle microtubule dynamics by stabilizing polar microtubules during mouse oocyte meiosis. *Biochim Biophys Acta*. Apr 2016;1863(4):630-7. doi:10.1016/j.bbamcr.2016.01.012

29. Lee JS, Jang EH, Woo HA, Lee K. Regulation of Autophagy Is a Novel Tumorigenesis-Related Activity of Multifunctional Translationally Controlled Tumor Protein. *Cells*. Jan 20 2020;9(1)doi:10.3390/cells9010257

30. łłzecka J. Decreased serum-soluble TRAIL levels in patients with amyotrophic lateral sclerosis. *Acta Neurol Scand*. May 2008;117(5):343-6. doi:10.1111/j.1600-0404.2007.00947.x

31. Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res*. Sep 2008;18(9):1509-17. doi:10.1101/gr.079558.108

32. Syage AR, Ekiz HA, Skinner DD, Stone C, O'Connell RM, Lane TE. Single-Cell RNA Sequencing Reveals the Diversity of the Immunological Landscape following Central Nervous System Infection by a Murine Coronavirus. *J Virol*. Nov 23 2020;94(24)doi:10.1128/jvi.01295-20

33. Vincent AM, Mobley BC, Hiller A, Feldman EL. IGF-I prevents glutamate-induced motor neuron programmed cell death. *Neurobiol Dis*. Jul 2004;16(2):407-16. doi:10.1016/j.nbd.2004.03.001

34. Vincent AM, Backus C, Taubman AA, Feldman EL. Identification of candidate drugs for the treatment of ALS. *Amyotroph Lateral Scler Other Motor Neuron Disord*. Mar 2005;6(1):29-36. doi:10.1080/14660820510026171

35. Figueroa-Romero C, Monteagudo A, Murdock BJ, et al. Tofacitinib Suppresses Natural Killer Cells In Vitro and In Vivo: Implications for Amyotrophic Lateral Sclerosis. *Front Immunol*. 2022;13:773288. doi:10.3389/fimmu.2022.773288

Reviewer 1

Zhao et al. have thoroughly revised the manuscript, putting substantial efforts into improving the underlying study, not only the manuscript. We sincerely appreciate the efforts made and the way that all concerns have been addressed extensively. Although some generalization concern may still remain, we think the revision of this work has rendered it significantly more valuable for publication. In particular, the publishing of the data and the code increased its value as a resource. Similarly, the revisions of the drug perturbation data and the clarifications in the text as to how this study improves the field - and prepares for the next step - now highlight its value.

Our response: We thank Reviewer 1 very much for their earlier valuable and insightful comments and suggestions, which helped us improve our study and manuscript. We hope that our study will serve as a springboard for further investigations leveraging molecular technologies for ALS diagnosis, prognosis, and drug development.

Reviewer 2

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Reviewer 3

The authors have addressed most of the critiques raised previously and the manuscript is substantially improved.

Our response: We express much gratitude to Reviewer 3 for their prior significant and astute remarks and suggestions on the submitted version of our manuscript, which assisted us in our efforts to revise and enhance our study and manuscript. Our work highlights the feasibility of bringing molecular technologies, such as gene expression data, to bear on ALS diagnosis, prognosis, and drug development. We hope this will catalyze further research, with the overarching goal for clinical use. We address the Reviewer's remaining comments below, and hope they will now find our manuscript suitable for publication in *Nature Communications*.

Rev 1 Comment 14 and Rev 3 Comment 2: As requested the authors have tested their classifier on the Swindell dataset but did not go beyond that although both reviewers requested testing against more datasets, which is disappointing.

Our response: We are very grateful for the suggestion from the first review of our submitted manuscript to apply our classifier to the Swindell et al.[1] as well as the Grima et al.[2] dataset. The ability of our classifier to differentiate ALS cases from controls in the Swindell dataset, collected using a different gene expression platform (microarray), showcased the strength of our classifier. However, to our knowledge, Swindell and Grima are the only publicly available gene expression datasets of ALS peripheral whole blood (**Response Table R1**).

Although in a distinct tissue, we also considered applying our classifier to ALS gene expression datasets of peripheral blood mononuclear cells (PBMCs); however, besides the distinct tissue, none are sufficiently large, limiting our ability to generate test performance metrics. Current PBMC datasets, to our knowledge, include Zhang et al. 2011 (PMID 20884065, ALS n=20, controls n=22), Lam et al. 2015 (PMID 26807342, ALS n=9, controls n=1), Gagliardi et al. 2018 (PMID 29402919, ALS n=7, controls n=4), and Zucca et al. 2019 (PMID 29402919, ALS n=24, controls n=7). Finally, given the only moderate correlation between peripheral whole blood to brain[3] and spinal cord tissue (please see also our **Figure 6**), which constitute the majority of other ALS gene expression datasets, we felt our classifier trained on peripheral whole blood would not necessarily be generalizable CNS tissues. Therefore, in our first revision, we limited analysis only to the Swindell et al.[1] as well as the Grima et al.[2] datasets. As the research community is increasingly appreciating the utility of blood gene expression for classification of neurodegenerative diseases, we anticipate more whole blood gene expression datasets of ALS will emerge with future opportunities to further test our classifier.

Response Table R1. Available ALS gene expression datasets in blood and various blood components

Study	PubMed	Tissue	n ALS	n control	Platform	GEO Series	Note
Saris 2009	19712483	Whole blood	30	30	Illumina HumanRef-8 BeadChip	None	Not compared due to data unavailability.
Zhang 20112	20884065	PBMCs	20	22	Affymetrix HG-U133 Plus 2.0 Array	None	Not compared due to data unavailability.
Mougeot 2011	22027401	Lymphocytes	11	11	Agilent Whole Genome 4 × 44k Array	GSE28253	Not compared due to use of a different tissue type.
Lam 2015	26807342	PBMCs	9	1	Illumina HiSeq 2000	None	Not compared due to data unavailability.
Zhao 2017	28437540	Monocytes	43	22	Illumina HiSeq 1500	None	Not compared due to data unavailability.
Gagliardi 2018	29402919	PBMCs	7	4	Illumina NextSeq 500	GSE115259	Not compared due to use of a different tissue type.
van Rheenen 2018	29939990	Whole blood	233	508	Illumina HumanHT-12 V3.0 Beadchip	GSE112676	Not compared due to lack of directional information (up/down regulation) for DEGs.
van Rheenen 2018	29939990	Whole blood	164	137	Illumina HumanHT-12 V4.0 Beadchip	GSE112680	Not compared due to lack of directional information (up/down regulation) for DEGs.
Zucca 2019	29402919	PBMCs	24	7	Illumina NextSeq 500	GSE106443, 115259	Not compared due to use of a different tissue type.
Swindell 2019	31118040	Whole blood	396	645	Illumina HumanHT-12 V3.0 & V4.0 Beadchip	GSE112676, 112680	Compared to our study.
Grima 2023	37818590	Whole blood	96	48	Illumina HiSeq 2000	GSE234297	Compared to our study.

Reviewer 3, comment 3: The authors claim that “building a classifier for ALS-ALS mimic classification, as well as a PCR format, are both future avenues that are outside the scope of the present study.” While their points are well taken, I do believe that without this crucial step their study is not novel or significant enough beyond the prior work that has been, which they cite. The main novelty here is that “this study has shown blood transcriptomics can accurately predict case-control status in external cohorts”.

Our response to “*The main novelty here is that “this study has shown blood transcriptomics can accurately predict case-control status in external cohorts*”.:

We fully appreciate the Reviewer’s perspective. We acknowledge that our milestone is that we can leverage whole blood transcriptomics to accurately predict case-control status in external cohorts. Nevertheless, it is a *significant milestone*. The Swindell et al. paper, published in 2019, is a re-analysis of an older study from 2018 by van Reehan et al.[4] Therefore, since the initial concept of potentially diagnosing ALS from blood in 2018, advances have stalled in the intervening half decade since no classifier was ever been shown to be useful in independent cohorts. We were able to move the field forward by, (i) transcriptomics analysis of a particularly large number of ALS cases (n=422) and controls (n=272) using RNA-seq, which is more granular than microarray, (ii) application of a modified surrogate variable analysis to eliminate batch effects and unknown sources of variation in gene expression, which better revealed the underlying gene expression attributable to ALS, (iii) application of stringent FDR and fold-change criteria, and (iv) application of a suite of machine learning algorithms to extract the best and most stable gene features.

Our response to *The authors claim that “building a classifier for ALS-ALS mimic classification, as well as a PCR format, are both future avenues that are outside the scope of the present study.” While their points are well taken, I do believe that without this crucial step their study is not novel or significant enough*

We respectfully disagree with the Reviewer. The journey from bench-to-beside for diagnostic tests is usually long and incremental due to the numerous challenging steps involved. Our long-term goal is to develop a PCR panel for diagnostic purposes for our patients, and *the milestone we reached in this paper is the significant critical first step* to achieve our goal. For the commercial PAM50 test, a 50-gene panel for breast cancer subtype classification, discovery started in 2000 based on Perou et al., published in Nature, which identified four intrinsic breast cancer subtypes from tumor gene expression data.[5] The 50-gene panel, based on a PCR platform, was reported 9 years later in 2009 by Parker et al.,[6] followed by a gene transcript-counting technology 4 years later in 2013, in the commercial incarnation of the panel, Prosigna PAM50.[7] Although this assay is FDA approved, research continues to evaluate the concordance between quantifying gene expression using more conventional counting methods and PCR-based Prosigna PAM50.[8,9] In our rebuttal letter, we similarly stated this as a separate future aim, by assessing concordance of our RNA-seq data to a newly developed multiplex PCR platform.

In our manuscript, we also stated the need to improve our classifier as a separate future aim, by introducing the capability of differentiating ALS cases from controls as well as ALS mimics. We have applied for funding for this future aim (R01 NS148128, under review) with the goal of entering 75 participants for each of three different ALS mimic conditions (n=225 total), which we anticipate will take two years.

In summary, we firmly believe our manuscript represents a vital step forward along the usual development pipeline for diagnostic and prognostic tests and hope the Reviewer, in this context, will agree. As the corresponding author and an ALS physician for 30 years, I am also hopeful that making this valuable dataset publicly available now will also advance our understanding of ALS pathogenesis, a disease where survival from diagnosis to death remains a dismal two to four years.

Reviewer 4

The Discovery and Validation Split follows best-practices. The independent validation on two external datasets with platform differences is strong. Great addition with the added metrics. The test of the clinical variable only model gives a good idea on the contribution of the gene model, which shows a high predictive power of the clinical variables or genes alone. The bootstrap analysis is great and shows the variability of the top features. The added analysis add a lot of value and show a diligent characterization of the data and gives a good perspective of the variability. The authors addressed my concerns and I can recommend publication in Nature Communications. I could access the GitHub code - I would recommend to add a Readme document to allow exploration.

Our response: We are very grateful to Reviewer 4 for their earlier constructive and meaningful critiques and recommendations on the submitted version of our manuscript. They allowed us to increase the robustness and transparency of our study and manuscript. We completely agree with their suggestion to add a Readme document to the study's GitHub code, which has now been deposited. Furthermore, once our manuscript has been published, we will add **Figure 1**, our study overview figure, to the GitHub welcome page. This figure provides a detailed overview of the study design as well as the various analyses we performed.

References

1. Swindell WR, Kruse CPS, List EO, Berryman DE, Kopchick JJ. ALS blood expression profiling identifies new biomarkers, patient subgroups, and evidence for neutrophilia and hypoxia. *J Transl Med*. May 22 2019;17(1):170. doi:10.1186/s12967-019-1909-0
2. Grima N, Liu S, Southwood D, et al. RNA sequencing of peripheral blood in amyotrophic lateral sclerosis reveals distinct molecular subtypes: Considerations for biomarker discovery. *Neuropathol Appl Neurobiol*. Dec 2023;49(6):e12943. doi:10.1111/nan.12943
3. Sullivan PF, Fan C, Perou CM. Evaluating the comparability of gene expression in blood and brain. *Am J Med Genet B Neuropsychiatr Genet*. Apr 5 2006;141b(3):261-8. doi:10.1002/ajmg.b.30272
4. van Rheenen W, Diekstra FP, Harschnitz O, et al. Whole blood transcriptome analysis in amyotrophic lateral sclerosis: A biomarker study. *PLoS One*. 2018;13(6):e0198874. doi:10.1371/journal.pone.0198874
5. Perou CM, Sørlie T, Eisen MB, et al. Molecular portraits of human breast tumours. *Nature*. Aug 17 2000;406(6797):747-52. doi:10.1038/35021093
6. Parker JS, Mullins M, Cheang MC, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol*. Mar 10 2009;27(8):1160-7. doi:10.1200/jco.2008.18.1370
7. https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K130010.
8. Nielsen T, Wallden B, Schaper C, et al. Analytical validation of the PAM50-based Prosigna Breast Cancer Prognostic Gene Signature Assay and nCounter Analysis System using formalin-fixed paraffin-embedded breast tumor specimens. *BMC Cancer*. Mar 13 2014;14:177. doi:10.1186/1471-2407-14-177
9. Wallden B, Storhoff J, Nielsen T, et al. Development and verification of the PAM50-based Prosigna breast cancer gene signature assay. *BMC Med Genomics*. Aug 22 2015;8:54. doi:10.1186/s12920-015-0129-6