ANNALS of Clinical and Translational Neurology
Open Access
AMERICAN NEUROLOGICAL ASSOCIATION

RESEARCH ARTICLE

Identification of potential pathways and biomarkers linked to progression in ALS

Roland G. Huber[1], Swapnil Pandey[2], Deepak Chhangani[2], Diego E. Rincon-Limas[2], Nathan P. Staff[3] & Crystal Jing Jing Yeo[4,5,6,7,8,9]

[1]Bioinformatics Institute (BII), Agency for Science, Technology and Research (A*STAR), Matrix #07-01, 30 Biopolis Street, Singapore, 138671, Singapore
[2]Department of Neurology, McKnight Brain Institute, and Norman Fixel Institute for Neurological Diseases, University of Florida, Gainesville, Florida, 32611, USA
[3]Department of Neurology, Mayo Clinic, Rochester, Minnesota, 55905, USA
[4]Agency for Science, Technology and Research (A*STAR), IMCB, 61 Biopolis Drive, Proteos, Singapore, 138673, Singapore
[5]Department of Neurology, Feinberg School of Medicine, Northwestern University, Chicago, Illinois, 60611, USA
[6]Lee Kong Chian School of Medicine, Imperial College London and NTU Singapore, Singapore, 308232, Singapore
[7]School of Medicine, Medical Sciences and Nutrition, University of Aberdeen, AB243FX,Scotland, UK
[8]National Neuroscience Institute, TTSH Campus, 11 Jalan Tan Tock Seng, Singapore, 308433, Singapore
[9]Duke NUS Medical School, 8 College Road, Singapore, 169857, Singapore

Correspondence
Crystal Jing Jing Yeo, Agency for Science, Technology and Research (A*STAR), IMCB, 61 Biopolis Drive, Proteos, Singapore 138673, Singapore. Tel: +65 63306363; Fax: +65 63577103; E-mail: crystaljjyeo@cantab.net

Received: 28 September 2022; Accepted: 24 October 2022

Annals of Clinical and Translational Neurology **2023; 10(2): 150–165**

doi: 10.1002/acn3.51697

Abstract

Objective: To identify potential diagnostic and prognostic biomarkers for clinical management and clinical trials in amyotrophic lateral sclerosis. **Methods**: We analysed proteomics data of ALS patient-induced pluripotent stem cell-derived motor neurons available through the AnswerALS consortium. After stratifying patients using clinical ALSFRS-R and ALS-CBS scales, we identified differentially expressed proteins indicative of ALS disease severity and progression rate as candidate ALS-related and prognostic biomarkers. Pathway analysis for identified proteins was performed using STITCH. Protein sets were correlated with the effects of drugs using the Connectivity Map tool to identify compounds likely to affect similar pathways. RNAi screening was performed in a Drosophila TDP-43 ALS model to validate pathological relevance. A statistical classification machine learning model was constructed using ridge regression that uses proteomics data to differentiate ALS patients from controls. **Results**: We identified 76, 21, 71 and 1 candidate ALS-related biomarkers and 22, 41, 27 and 64 candidate prognostic biomarkers from patients stratified by ALSFRS-R baseline, ALSFRS-R progression slope, ALS-CBS baseline and ALS-CBS progression slope, respectively. Nineteen proteins enhanced or suppressed pathogenic eye phenotypes in the ALS fly model. Nutraceuticals, dopamine pathway modulators, statins, anti-inflammatories and antimicrobials were predicted starting points for drug repurposing using the connectivity map tool. Ten diagnostic biomarker proteins were predicted by machine learning to identify ALS patients with high accuracy and sensitivity. **Interpretation**: This study showcases the powerful approach of iPSC-motor neuron proteomics combined with machine learning and biological confirmation in the prediction of novel mechanisms and diagnostic and predictive biomarkers in ALS.

Introduction

Amyotrophic lateral sclerosis (ALS) is the most common adult-onset motor neuron disease, with an estimated prevalence of 4.42 per 100,000 population and an incidence of 1.59 per 100,000 person-years.[1] It is characterized by progressive deterioration of motor neurons in the brain and spinal cord and is invariably fatal. Heterogenous in clinical presentation and disease mechanisms, there is much variability in the site of onset, upper

Summary for Social Media if Published

1. If you and/or a co-author has a Twitter handle that you would like to be tagged, please enter it here: @CrystalYeoMDPhD

2. What is the current knowledge on the topic? (one to two sentences): ALS progression is highly heterogenous, posing challenges for diagnosis, monitoring, clinical trial design and therapeutic interventions.

3. What question did this study address? (one to two sentences): This study looked at proteomic changes between patient iPSC-derived motor neurons from fast and slow, early and late ALS patients and normal controls combined with machine learning and biological confirmation.

4. What does this study add to our knowledge? (one to two sentences): This study identified proteins and pathways which could be potential ALS-related and prognostic biomarkers for ALS, confirmed suppressor and enhancer genes in a fly model of the disease, and suggested possible repurposed drugs.

5. How might this potentially impact the practice of neurology? (one to two sentences): ALS-related and prognostic biomarkers could be used by clinicians in managing patient care and when selecting patients for treatment or clinical trial enrolment. Selected causal genes and repurposed drugs could be targeted as therapeutic avenues.

or lower motor neuron involvement and progression rate.[2] About three in four patients first report limb-onset weakness, and one in four report bulbar-onset difficulties of swallowing and/or speaking. Progressive weakness and difficulties with speech, swallowing and breathing are followed by respiratory failure and death, usually in 2 to 5 years, although some forms exhibit protracted survival. Whilst ALS is primarily driven by the degeneration of motor neurons, cognitive and behavioural changes corresponding to frontotemporal dysfunction occur in up to 40% of cases.[3]

There is also vast genetic heterogeneity. 10% of ALS patients have a clear familial inheritance pattern which may be associated with a pathogenic gene, the most common being intronic hexanuclotide expansions in *C9orf72* and missense mutations in *SOD1*, *TARDBP* and *FUS*.[4] Many pathogenic ALS genes are associated with protein quality control, RNA metabolism, or cytoskeletal disruptions in the motor axon.[4] Many other genes confer susceptibility and not necessarily causation, such as *ATXN2*, where mid-range CAG expansions increase the risk of developing ALS.[5] In about 90% of patients, ALS is sporadic with no known inheritance pattern. A range of disease mechanisms have also been reported, including protein aggregation, ER stress, disruption of protein degradation, abnormal DNA and RNA processing, mitochondrial dysfunction,[6] issues with oxidative homeostasis[7] and inflammation.[8] These mechanisms may be cell-autonomous and non-cell-autonomous since the defective genes which cause ALS are expressed in different cell types including motor neurons, microglial, astrocytes and oligodendrocytes.

The broad clinical and biological heterogeneity in ALS has implications on patient risk assessment, selection for clinical trials, and timing of therapeutic interventions, all of which challenge therapeutic development.[9] Since ALS patients differ in age, clinical features and genetic makeup, personalized medicine approaches will likely be needed in the future to develop effective treatments across the spectrum of ALS, which is already occurring in the area of gene therapies for familial ALS. The only two medications currently approved by the FDA for ALS, riluzole,[10] an antagonist of glutamate transmission, and edaravone,[11–13] a superoxide scavenger, only modestly prolong survival or slow disease progression. Biomarkers are also needed for the diagnosis and stratification of patients for prognostication, enrolment in clinical trials, and to establish drug-responsive biomarkers or diagnostics.[9]

In this study, we investigated proteomic changes in induced pluripotent stem cells (iPSC)-derived motor neurons from ALS patients at the extremes of clinical phenotypes to uncover mechanistic pathways associated with the motor and cognitive/behavioural phenotypes. Two classes of potential biomarkers, namely ALS-related and prognostic biomarkers, and new therapeutic approaches were identified using statistical and machine learning approaches and a RNAi screen of candidates in Drosophila confirmed the biological importance of putative causal and susceptibility genes.

Materials and Methods

Answer ALS data portal

Clinical and proteomics data from induced pluripotent stem cells (iPSCs) and motor neurons for this study was obtained from the Answer ALS Data Portal, available at answerals.org. iPSC generation for AnswerALS is handled by the Cedars-Sinai biomanufacturing centre and details on processing for each cell line can be retrieved via their web portal (https://www.cedars-sinai.edu/research/areas/biomanufacturing/ipsc.html). The specific protocol is

outlined in Baxi et al.[14] Proteomics data was acquired using the SWATH method. We extracted the quantifications and used Limma[15] (version 3.50.0; Bioconductor[16] version 3.14, R[17] version 4.1.1) to build a model of differential protein expression on log2-normalized protein abundance for different patient and control groups. The AnswerALS database also provides whole-exome sequencing data for all patient and control samples discussed in this study.

Categorization as fast or slow ALS progressors and milder or severe ALS phenotypes

We stratified patients according to their ALSFRS-R[18] and CBS[19] scale baseline values and progression slopes. Weakness and loss of function are measured by the ALSFRS-R, which comprises 12 items to assess daily functioning mediated by the cervical, trunk, lumbosacral and respiratory muscles, with a scale ranging from 0 to 4 for each item, giving a total score of between 0 and 48, with 48 being normal.[18] Scores of less than 30 are considered as having moderate to severe impact. ALS CBS cognitive score measures cognitive change.[19] It ranges between 0 and 20, with scores ≤10 classified as possible Frontal Temporal Lobal Dementia (FTLD), cognitive type, and scores 11–16 classified as cognitively impaired. Baseline values are evaluations obtained at first diagnosis whilst slope is calculated from the observed change in scale values at subsequent visits. Negative slopes are associated with the worsening condition and steeper slopes indicate faster progression. ALSFRS-R baseline values were split at 30, with patients >30 regarded as high (milder clinical phenotype) and ≤30 as low (severe clinical phenotype). CBS baseline values were split at 15 with patients >15 regarded as high and values ≤15 as low. ALSFRS-R slopes were classified as fast progression (<−1.5/month) and slow progression (>−0.5/month) with values in between classed as intermediate. CBS slopes were equally stratified as fast progression (<−0.5/month), slow progression (>0.5/month) and intermediate progression for the remainder. ALSFRS-R progression slopes in particular have been shown to be predictive of survival in ALS.[20] Cut-offs for these classifications were chosen to identify the extremes of phenotype and are not necessarily associated with specific clinical criteria. Our aim is to identify genes determinative of particular progressions and hence we compare these extremes and exclude intermediate progression slopes. This approach of contrasting favourable and unfavourable phenotypes has previously been applied to the study of cancer genomics.[21] As we do not use the cutoffs of early- and late-stage disease to directly look for biomarkers, these cutoffs are chosen somewhat arbitrarily to yield larger group sizes for the late-stage patients, thus enhancing statistical confidence for the findings. We then particularly identify features that share a contrast with the healthy control group from both early- and late-stage cases. Patients for whom either baseline or slope values were unavailable were excluded from the analysis. Demographics data for all classes were calculated and are given in Tables 1 and 2.

Identification of diagnostic and prognostic biomarker proteins

In order to identify both ALS-related and prognostic biomarkers we calculated the differential protein concentration between all groups and extracted significant contrasts that showed at least >2-fold concentration change and a p-value of <0.05 using Limma[15] on log2-normalized protein expression values obtained for differentiated motor neurons available within the AnswerALS public dataset. Each set of proteins matching the specific criteria outlined above is shown in Table S2. To identify only high-confidence candidates, we looked at the intersections of these gene sets for both ALS-related and prognostic proteins which are depicted in the Venn diagrams of Figs. 1, 2. Candidate proteins for ALS-related biomarkers were extracted by contrasting high and low classes for baseline values, or fast and slow progressors for slopes, against controls and taking the intersection of these gene sets (Figs. 1, 2, upper central element in the Venn diagram). Prognostic biomarker candidates were identified as proteins that differ between high and low baseline or fast and slow progression, but are in common with either group's contrast to control (Figs. 1, 2, lower lateral elements in the Venn diagram). Heatmaps for identified ALS-related or prognostic candidate proteins are presented in Figs. 1, 2 and a table of each set of genes is presented in the Table S2. Heatmaps depict per-protein normalized values and changes in expression are shown as standard deviations from the mean of all samples for a specific protein. Analysis was performed using Python version 3.9.7, using the pandas[22] (version 1.3.5), scipy[23] (version 1.7.1) libraries and visualization were performed using matplotlib[24] (version 3.4.2). In order to understand the role these proteins play in the organism, we performed a STITCH[25,26] analysis which identifies interaction networks between proteins and provides associated gene ontology terms from a variety of gene ontology databases. We opted to allow STITCH to expand the network with up to 5 linked proteins not present in the query set. The resulting networks are presented alongside the expression data in Figs. 1, 2. Gene ontology terms associated with the various protein sets are presented in ST1.

Table 1. Clinical features of early versus late disease and fast versus slow progressors based on ALSFRS-R.

Clinical features of early versus late disease using ALSFRS-R baselines							
ALSFRSR-baseline <30 (late)				ALSFRS-R baseline > or =30 (early)			
Clinical features	Mean	Standard deviation	Number	Mean	Standard deviation	Number	*p* value
Age at symptom onset	50.74	11.99	*N* = 38	55.82	11.06	*N* = 101	0.0206
Age at death	59.71	9.17	*N* = 17	63.11	8.31	*N* = 46	0.287
Sex (% Female)	34.21		*N* = 38	44.23		*N* = 104	0.854
Bulbar onset (%)	26.32		*N* = 38	27.88		*N* = 104	0.0478[a]
Axial onset (%)	0		*N* = 38	9.62		*N* = 104	0.25
Limb onset (%)	68.42		*N* = 38	77.88		*N* = 104	1.40 E-39[a]
ALSFRS-R Baseline	20.45	5.72	*N* = 38	37.17	4.29	*N* = 104	2.98 E-10[a]
ALSFRS-R Latest	16.57	7.7	*N* = 37	28.79	9.85	*N* = 104	0.0165[a]
ALSFRS-R Progression Slope	−0.34	0.39	*N* = 23	−0.68	0.64	*N* = 74	0.631
CBS Baseline	15.53	3.98	*N* = 34	15.84	2.81	*N* = 88	0.202
CBS Latest	15.09	4.79	*N* = 35	16	3.05	*N* = 97	0.763
CBS Progression Slope	0.04	0.3	*N* = 11	0.01	0.32	*N* = 45	0.413
C9orf72 gene hexanucleotide repeat count	4	2.12	*N* = 4	117.93	266.27	*N* = 30	0.784
ATXN2 gene trinucleotide repeat count	22.25	0.43	*N* = 4	22.63	2.69	*N* = 30	0.287
Clinical features of fast versus slow progressors using ALSFRS-R							
ALSFRS-R progression > −0.5/month (Slow)				ALSFRS-R progression < −1.5/month (fast)			
Clinical features	Mean	Standard deviation	Number	Mean	Standard deviation	Number	*p* value
Age at symptom onset	52.4	12.0	*N* = 55	59.4	6.5	*N* = 8	0.12
Age at death	58.8	10.1	*N* = 6	62.4	6.2	*N* = 9	0.44
Sex (% Female)	40.0		*N* = 55	33.3		*N* = 9	0.71
Bulbar onset (%)	10.9		*N* = 55	33.3		*N* = 9	0.07
Axial onset (%)	0.0		*N* = 55	33.3		*N* = 9	0.00[a]
Limb onset (%)	92.7		*N* = 55	77.8		*N* = 9	0.16
ALSFRS-R Baseline	33.2	9.1	*N* = 54	36.7	3.1	*N* = 9	0.27
ALSFRS-R Latest	29.8	10.0	*N* = 55	16.6	6.0	*N* = 9	0.00[a]
ALSFRS-R Progression Slope	−0.2	0.2	*N* = 55	−2.0	0.4	*N* = 9	0.00[a]
CBS Baseline	16.4	2.2	*N* = 47	14.3	3.9	*N* = 6	0.06
CBS Latest	16.8	2.6	*N* = 51	14.1	4.6	*N* = 7	0.03[a]
CBS Progression Slope	0.1	0.2	*N* = 33	−0.2	0.2	*N* = 3	0.02[a]
C9orf72 gene hexanucleotide repeat count[b]	88.6	257.0	*N* = 11	378.7	437.1	*N* = 3	0.20
ATXN2 gene trinucleotide repeat count[c]	22.2	0.4	*N* = 11	22.0	0.0	*N* = 3	0.46

Abbreviations: ALSFRS-R, Amyotrophic Lateral Sclerosis Functional Rating Scale-Revised; ALS CBS, Amyotrophic Lateral Sclerosis Cognitive Behavioural Screen; *C9orf72*, chromosome 9 open reading frame 72; *ATXN2*, ataxin 2.

[a]Represents significant values.

[b]Abnormal hexanucleotide repeat counts are >30.

[c]Abnormal trinucleotide repeat counts associated with ALS are between 29 and 33.

Statistical machine learning model to determine ALS status from protein expression

Based on the identified protein candidates we pooled all possible ALS-related and prognostic proteins respectively and created a statistical model to infer ALS status from proteomics data. As our data set is confined to a limited set of patients and controls, we used a Ridge classifier[27] model with internal cross-validation as implemented in scikit-learn[28] (1.0). After splitting the data into a 75% training and 25% test set, we were able to fit a statistical model on raw expression data that was able to identify ALS with perfect accuracy on the training set and 97% precision and 78% sensitivity on the unseen test set (Table 4A). We assessed precision (positive predictive

Table 2. Clinical features of early versus late disease and fast versus slow progressors based on ALS-CBS.

Clinical features of early versus late disease using ALS CBS baselines							
ALS CBS baseline < 15 (Late)				ALS CBS baseline > or = 15 (Early)			
Clinical features	Mean	Standard deviation	Number	Mean	Standard deviation	Number	*p* value
Age at symptom onset	59.42	9.63	*N* = 33	53.51	−11.25	*N* = 88	0.851
Age at death	65.17	−6.96	*N* = 18	61.95	8.62	*N* = 38	0.18
Sex (% Female)	47.06		*N* = 34	43.96		*N* = 91	0.759
Bulbar onset (%)	32.35		*N* = 34	26.37		*N* = 91	0.512
Axial onset (%)	5.88		*N* = 34	7.69		*N* = 91	0.73
Limb onset (%)	61.76		*N* = 34	79.12		*N* = 91	0.0483
ALSFRS-R Baseline	31.41	8.02	*N* = 34	33.16	8.04	*N* = 88	0.288
ALSFRS-R Latest	23.15	11.13	*N* = 34	27.12	9.63	*N* = 89	0.0544
ALSFRS-R Progression Slope	−0.79	0.7	*N* = 21	−0.5	0.55	*N* = 61	0.059
CBS Baseline	11.68	2.69	*N* = 34	17.32	1.51	*N* = 91	1.11 E-28[a]
CBS Latest	12.06	4.21	*N* = 34	17.24	1.97	*N* = 91	8.90 E-16[a]
CBS Progression Slope	0.09	0.38	*N* = 13	−0.01	0.29	*N* = 41	0.31
C9orf72 gene hexanucleotide repeat count	239.57	370.21	*N* = 7	74.29	217.32	*N* = 21	0.179
ATXN2 gene trinucleotide repeat count	22.29	0.45	*N* = 7	22.81	−3.19	*N* = 21	0.68
Clinical features of fast versus slow progressors using ALS CBS							
ALS CBS progression > 0.5 (slow)				ALS CBS progression < −0.5 (fast)			
Clinical features	Mean	Standard deviation	Number	Mean	Standard deviation	Number	*p* value
Age at symptom onset	53.33	13.07	*N* = 3	58	7	*N* = 2	0.74
Age at death	66	8	*N* = 2	63.5	3.5	*N* = 2	0.80
Sex (% Female)	100		*N* = 3	50		*N* = 2	0.27
Bulbar onset (%)	33.33		*N* = 3	100		*N* = 2	0.22
Axial onset (%)	0		*N* = 3	0		*N* = 2	NA
Limb onset (%)	66.67		*N* = 3	0		*N* = 2	0.22
ALSFRS-R Baseline	30.67	1.89	*N* = 3	36	1	*N* = 2	0.07
ALSFRS-R Latest	20.33	7.72	*N* = 3	18	10	*N* = 2	0.84
ALSFRS-R Progression Slope	−0.89	0.57	*N* = 3	−1.8	0.63	*N* = 2	0.29
CBS Baseline	11	2.16	*N* = 3	16.5	2.5	*N* = 2	0.14
CBS Latest	18	1.41	*N* = 3	9.5	0.5	*N* = 2	0.01[a]
CBS Progression Slope	0.7	0.13	*N* = 3	−1.02	0.48	*N* = 2	0.02[a]
C9orf72 gene hexanucleotide repeat count[b]	nil	nil	*N* = 0	321	321	*N* = 1	N.A.
ATXN2 gene trinucleotide repeat count[c]	nil	nil	*N* = 0	22	22	*N* = 1	N.A.

Abbreviations: ALSFRS-R, amyotrophic lateral sclerosis functional rating scale-revised; ALS CBS, amyotrophic lateral sclerosis cognitive behavioural screen; *C9orf72*, chromosome 9 open reading frame 72; *ATXN2*, ataxin 2.

[a]Represents significant values.

[b]Abnormal hexanucleotide repeat counts are >30.

[c]Abnormal trinucleotide repeat counts associated with ALS are between 29–33.

value), recall (sensitivity), f1 score (the harmonic mean of precision and recall) and accuracy (the mean of f1 scores) for our model. We proceeded to sort the classifier coefficients by the magnitude and identify the top 10 most contributing proteins to the model. The absolute values of these coefficients account for approximately 70% of the total model. We then proceeded to restrict our model to the 10 identified genes which resulted in a precision of 97% and a sensitivity of 52% on the training set and a precision of 89% and sensitivity of 62% on the unseen test set (Table 4B).

Drosophila genetics, RNAi screen and phenotypic analysis

To validate protein candidates, we used a transgenic fly model of ALS based on the expression of human TDP-43^{M337V}.[29] We first recombined this line with the

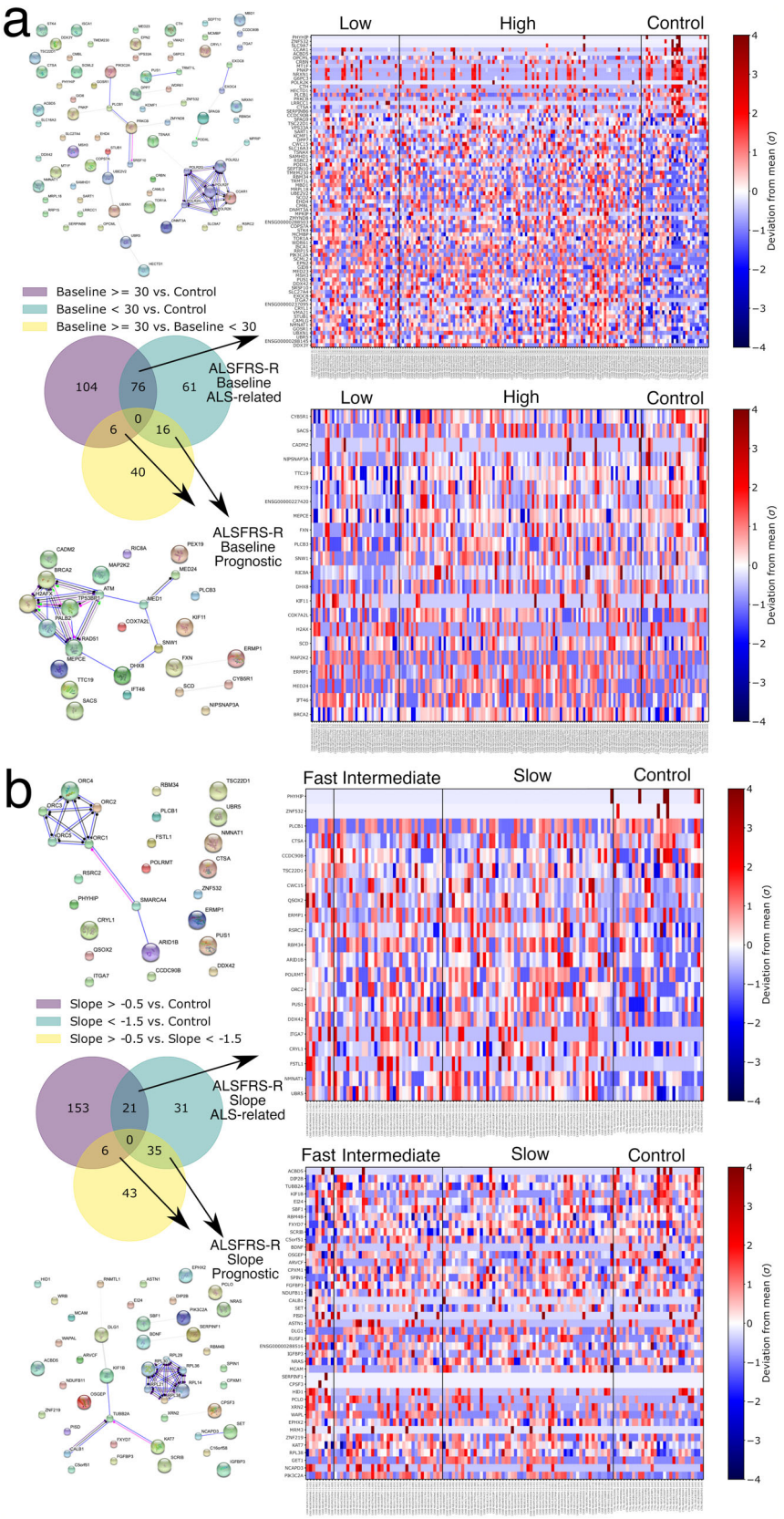


Figure 1. Protein contrasts for ALS-related and prognostic biomarkers based on ALSFRS-R. We identified proteins that show significantly different abundance (>2-fold change, $p \leq 0.05$) in three classes, namely high initial score or shallow progression slope versus control, low initial score or steep progression slope versus control and high initial score or shallow slope versus low initial score or steep slope respectively. We then designated candidate ALS-related biomarkers as proteins that are shared between the contrasts with the control group (upper central section of the Venn diagrams), whilst we required candidate prognostic biomarkers to present in the outcome contrast and at least one contrast versus control (left and right and central intersections in the Venn diagrams) STITCH analysis with up to 5 additional included genes are shown for ALS-related and prognostic biomarkers whilst heatmaps depict normalized expression levels. (A) Classification by ALSFRS-R baseline score. (B) Classification by ALSFRS-R progression slope.

GMR-Gal4 driver in the second chromosome to target expression to photoreceptor neurons in the fly eye. Then, we crossed the resulting recombinant stock with the innocuous Luciferase RNAi control line (BDSC# 35788) along with RNAi transgenes against the fly homologues of candidate genes (Table S3). All RNAi stocks used were obtained from the Bloomington Drosophila Stock Center in Indiana or the Vienna Drosophila Resource Center. Crosses were performed at 27°C and subsequent progenies were collected 24 h after eclosion and frozen overnight at −80°C. Eye images were acquired as Z-stacks with the Leica Z16 APO zoom system and single in-focus images were generated with the Montage Multifocus module of the Leica Application Software. Those genes that improved the eye phenotype upon knock-down were scored as suppressors and those that aggravated it were deemed enhancers. Eye phenotypes were quantified using a flynotyper score for suppressors and a severity score for enhancers as previously outlined.[29,30]

Identification of possible disease-modifying therapies

We imported the identified gene sets corresponding to our protein sets into the connectivity map tool (CMAP)[31–33] provided by the Broad Institute. The connectivity map uses gene expression signatures to connect gene expression patterns to the effects of drugs on these genes. For a set of query genes, a ranking of drugs and pathways is returned that measures the goodness of fit between the known effects of the drugs and the genes in the query based on a percentile ranking within a large set of queries. A detailed explanation of how this score is calculated is given in Subramanian et al.[34] This ranking allows us to pinpoint candidate drugs that affect proteins and pathways highlighted in our differential expression analysis. We then extracted the top 20 compounds for each of our gene sets that showed the highest inverse correlation with the differential gene expression pattern, of which selected compounds in prominent drug categories are presented in Table 3.

Results

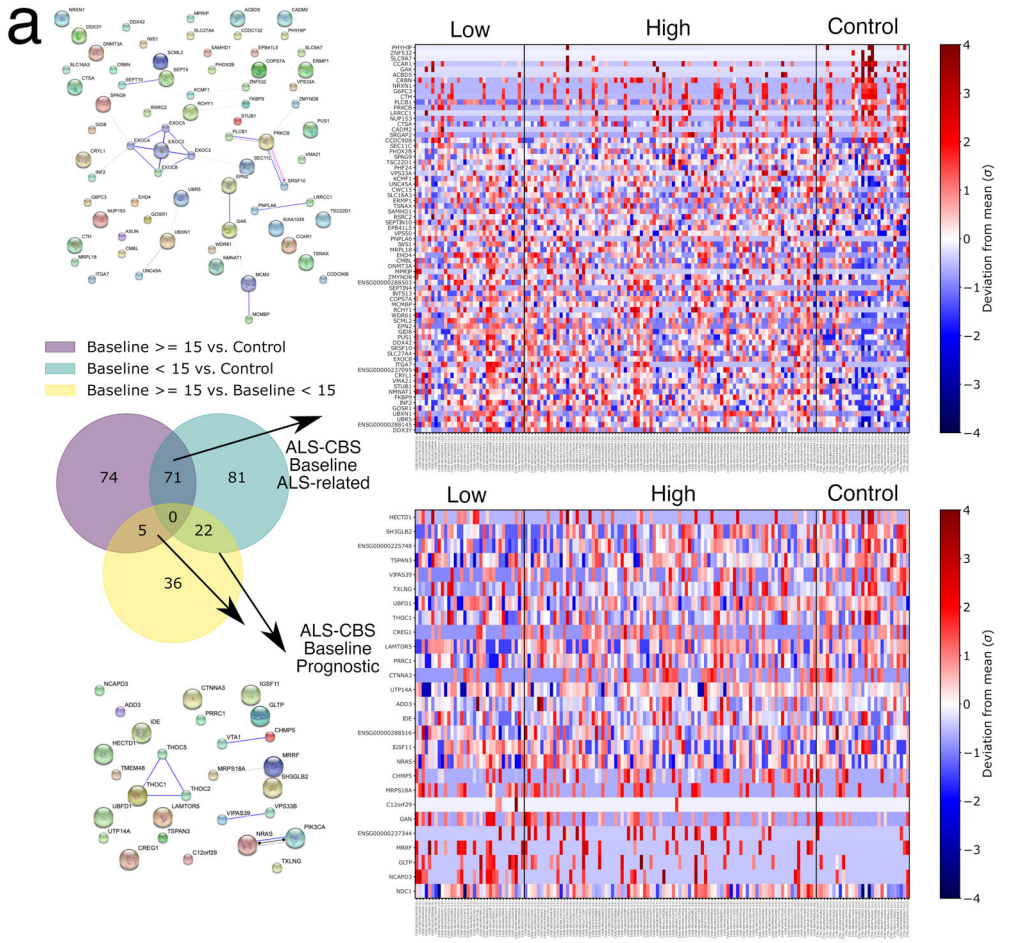
Clinical features of milder versus severe phenotypes and fast versus slow progressors

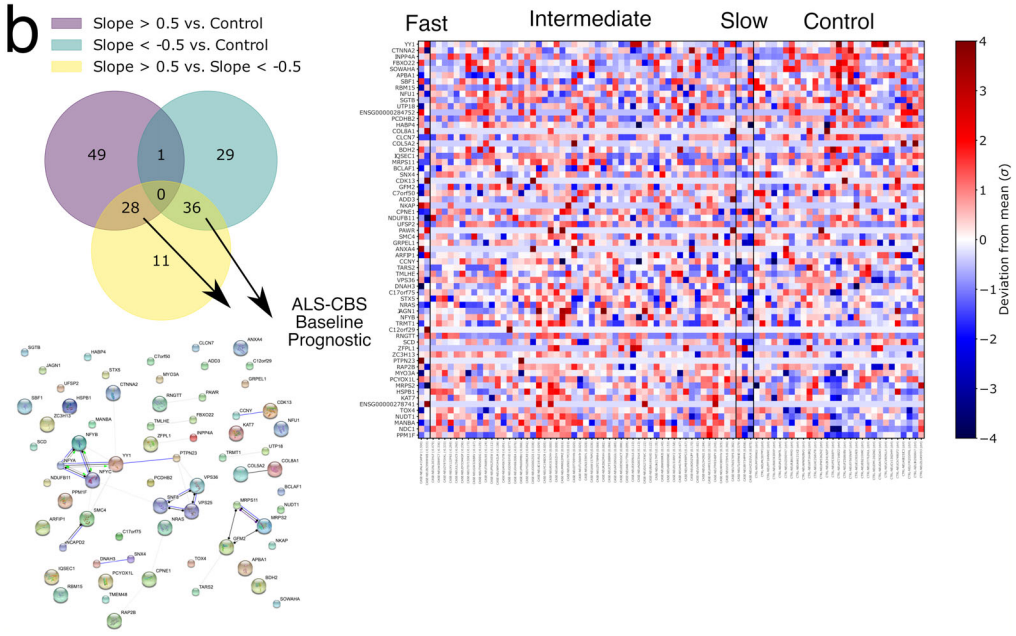
433 iPSC-derived motor neurons were available within the AnswerALS cohort, comprising 366 ALS patients and 67 healthy controls. Out of these, 149 patient samples and 29 healthy controls were associated with proteomics data in December 2021 and hence were included in this study. Out of 149 patients, 38 had ALSFRSR <30, 101 had ALSFRSR ≥30 and 7 patients had no ALSFRSR available; 55 had ALSFRS progression >−0.5/month, 8 had ALSFRSR progression <−1.5/month, 35 had intermediate slopes between these values and 50 patients had undefined slopes; 33 had ALS CBS baseline <15, 88 had ALS CBS baseline ≥15 and 24 had undefined ALS CBS baselines; 3 had ALS CBS Progression >0.5, 2 had ALS CBS progression <−0.5, 52 had intermediate slopes between these values, and 92 had undefined slopes. Demographics were well-matched when comparing the extremes of the motor and cognitive phenotypes (Tables 1, 2). No significant difference was noted in age of death, sex and number of pathogenic nucleotide repeat in the C9orf72 and ATXN2 genes (Tables 1, 2) with the only significant difference being the age of onset.

Identification of potential ALS-related and prognostic biomarkers from proteomics analysis of motor neurons differentiated from patient-derived iPSC

Contrasting patients based on ALSFRS-R[18] baseline scores higher or lower than 30 and each of these groups with the healthy controls yields a total of 186 significantly altered proteins for ALSFRS-R baseline score ≥30 versus control, 153 proteins for ALSFRS-R baseline score < 30 versus control and 62 proteins for contrasting high versus low baseline ALSFRS-R (Fig. 1A). The intersection between high and low baseline score versus control

Figure 2. Protein contrasts for ALS-related and prognostic biomarkers based on ALS-CBS. We identified proteins that show significantly different abundance (>2-fold change, $p \leq 0.05$) in three classes, namely high initial score or shallow progression slope versus control, low initial score or steep progression slope versus control and high initial score or shallow slope versus low initial score or steep slope respectively. We then designated candidate ALS-related biomarkers as proteins that are shared between the contrasts with the control group (upper central section of the Venn diagrams), whilst we required candidate prognostic biomarkers to present in the outcome contrast and at least one contrast versus control (left and right and central intersections in the Venn diagrams) STITCH analysis with up to 5 additional included genes are shown for ALS-related and prognostic biomarkers whilst heatmaps depict normalized expression levels. (A) Classification by ALS-CBS baseline score. (B) Classification by ALS-CBS progression slope. Only a single protein, ENO3, is classified as an ALS-related candidate and the heatmap and STITCH analysis are omitted for this category.


a
Baseline >= 15 vs. Control
Baseline < 15 vs. Control
Baseline >= 15 vs. Baseline < 15
74
71
81
0
5
22
36
ALS-CBS Baseline ALS-related
ALS-CBS Baseline Prognostic
Low
High
Control
Deviation from mean (σ)
4
3
2
1
0
−1
−2
−3
−4


b
Slope > 0.5 vs. Control
Slope < -0.5 vs. Control
Slope > 0.5 vs. Slope < -0.5
49
1
29
28
0
36
11
ALS-CBS Baseline Prognostic
Fast
Intermediate
Slow
Control
Deviation from mean (σ)
4
3
2
1
0
−1
−2
−3
−4

comprises of 76 proteins that are candidate ALS-related biomarkers as these are significantly different proteins between patients and healthy controls regardless of clinical severity of the disease. The intersection of high versus low with either high or low versus control yields 22 proteins as candidates for prognostic biomarkers as these are proteins that significantly differ between the milder or severe clinical phenotypes (Fig. 1A).

Classification by ALSFRS-R progression slope of either >−0.5 for slow progressors or <−1.5 for fast progressors and the corresponding contrasts in proteins yields 180 proteins for slow progression versus control, 87 proteins for fast progression versus control and 84 proteins for the contrast of fast versus slow progression. There are 21 candidate ALS-related biomarker proteins in the intersection of the slow versus control and fast versus control groups and 41 candidate prognostic biomarker proteins from either the intersection of fast versus slow with fast versus control or slow versus control (Fig. 1B).

Classification by ALS-CBS baseline score of either ≥15 (high), or <15 (low) yields a total of 150 proteins for high baseline versus control, 174 for low baseline versus control and 63 proteins for high baseline versus low baseline. A total of 71 candidate ALS-related biomarker proteins overlap between the high versus control and low versus control groups. A total of 27 prognostic biomarker candidates are identified from high versus low intersecting either with high versus control or low versus control (Fig. 2A).

Classification of patients by ALS-CBS[19] progression slope using >0.5 for slow progression and <−0.5 for fast progression results in 78 proteins that are significantly altered in slow progression versus control, 66 altered proteins in fast progressors versus control and 75 proteins that are different between fast progressors versus slow progressors. Only a single protein, ENO3, is shared between the fast progression versus control and slow progression versus control groups and hence was not included in further gene set-based analysis. A total of 64 proteins are shared by the fast progression versus slow progression contrast and either fast versus control or slow versus control contrasts and hence constitute the candidate prognostic biomarker set based on the ALS-CBS slope (Fig. 2B).

Figures 1 and 2 include a STITCH[25,26] maps of associated gene networks for the respective candidate biomarkers and Table S1 shows gene ontology terms resulting from these networks for each prognostic or ALS-related set, respectively.

Table 3. Candidate repurposed drugs.

Classes	Compounds	Norm_CS	Pattern
Nutraceuticals	Curcumin	−2.0061	ALS CBS baseline prognostic
	Yohimbine	−1.9514	ALSFRS-R progression ALS-related
	Ginkgolide-b	−1.7385	ALSFRS-R progression prognostic
	Forskolin	−1.7441	ALSFRS-R baseline ALS-related
	Ipriflavone	−1.6879	ALSFRS-R baseline ALS-related
	Ginsenoside	−1.6878	ALSFRS-R progression prognostic
	Luteolin	−1.6402	ALSFRS-R baseline ALS-related
Statins	Simvastatin	−2.0035	ALS CBS baseline prognostic
	Atorvastatin	−1.7351	ALSFRS-R progression prognostic
Dopamine pathway modulators	Fenoldopam	−2.1363	ALSFRS-R progression ALS-related
	Raclopride	−2.028	ALS CBS baseline prognostic
	Clozapine	−1.9439	ALSFRS-R progression ALS-related
	Dopamine	−1.6483	ALSFRS-R baseline ALS-related
Anti-inflammatories	Mesalazine	−2.047	ALS CBS baseline prognostic
	Balsalazide	−1.9469	ALSFRS-R progression ALS-related
	Sulfafurazole	−1.8996	ALSFRS-R progression ALS-related
	Cinalukast	−1.7893	ALSFRS-R progression prognostic
	Indobufen	−1.7429	ALSFRS-R progression prognostic
	Etodolac	−1.7245	ALSFRS-R baseline ALS-related
	Sulfasalazine	−1.6673	ALSFRS-R baseline ALS-related
	Econazole	−1.7727	ALS CBS progression prognostic
	Primaquine	−1.7497	ALS CBS progression prognostic
	Terbinafine	−1.6951	ALS CBS progression prognostic
	Clindamycin	−1.6537	ALS CBS baseline ALS-related
	Tobramycin	−1.646	ALSFRS-R baseline ALS-related
Thiazolidinedione	Ciglitazone	−1.6988	ALSFRS-R progression prognostic
Sodium channel blocker (FDA-approved ALS drug)	Riluzole	−1.38, 0.18	Mean score across all patterns, standard deviation
Antioxidant (FDA-approved ALS drug)	Edaravone	−1.27, 0.20	Mean score across all patterns, standard deviation

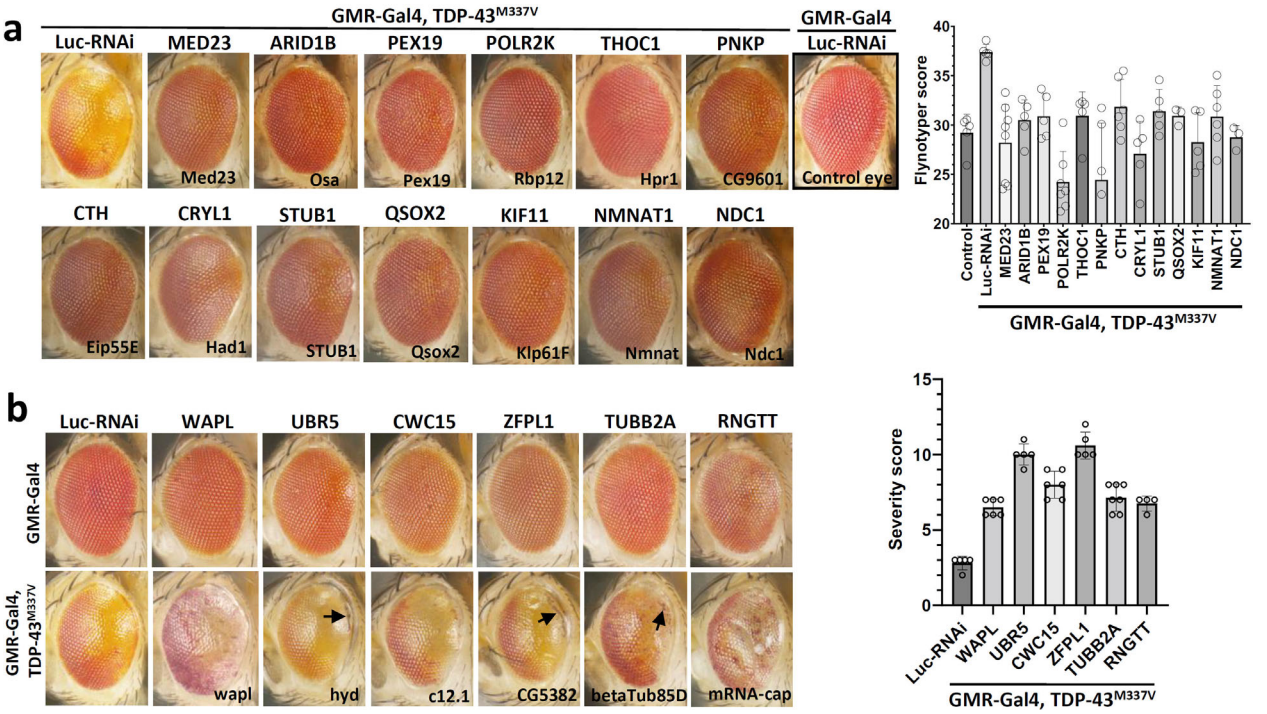


Figure 3. Identification of genetic suppressors and enhancers of human TDP-43 M337V toxicity in the Drosophila eye. Co-expression of human TDP-43^{M337V} with the innocuous Luciferase (Luc) RNAi transgene in the eye leads to disorganization of the external surface and posterior depigmentation (top left corner) compared to control flies expressing the Luc RNAi transgene alone (top right corner). This phenotype was suppressed (A) or enhanced (B) when RNAi constructs against the indicated genes were included. Note that the enhancers do not perturb the morphology of the eye when knocked down in the absence of TDP-43^{M337V}, confirming their specific interactions. Names of the fly genes are shown at the bottom of each panel and the corresponding human homologs are indicated at the top. Expression of transgenes was induced with the eye-specific GMR-Gal4 driver at 27°C. Eye phenotypes of suppressors and enhancers were quantified using the flynotyper and severity scores, respectively, and histograms were prepared using Graphpad Prism ($n \geq 5$).

In vivo identification of genes with therapeutic potential

To define the disease-modifying potential of selected proteomics hits, we capitalized on a well-established *Drosophila* model of ALS that expresses human mutant TDP-43^{M337V}.[29] Expression of the TDP-43^{M337V} transgene in photoreceptor neurons of the fly eye triggers posterior depigmentation and disorganization of the ommatidial lattice when compared to flies expressing an innocuous Luciferase transgene (Fig. 3A, compare panels at the top left and right corners). This easy-to-score phenotype is highly consistent and thus provides a reliable platform to test fly homologues of human genes for their ability to modify TDP-43^{M337V} toxicity upon RNAi silencing. From the top 157 candidates, we screened 57 genes which had Drosophila homologues and available RNAi lines for testing. Amongst these, we found 13 suppressors and 6 enhancers of TDP-43^{M337V} toxicity. Since suppressor genes improve the eye phenotype when knocked down, they may act as causative genes or risk factors, whilst enhancers of the phenotype may play protective or compensatory roles. Fig. 3A shows all 13 suppressors. These include fly homologues for MED23, a conserved subunit of the Mediator complex that regulates transcription of RNA pol II-dependent genes; ARIDB1, a component of the SWI/SNF chromatin complex involved in cell-cycle activation; PEX19, a factor that facilitates the import of peroxisomal membrane proteins; POLR2K, participates in mRNA synthesis; THOC1, is a member of the TREX mRNA export machinery; PNKP, an enzyme involved in DNA repair; CTH, an enzyme that converts cystathionine into cysteine in the trans-sulfonation pathway; CRYL1 which mediates D-glucuronate degradation, a step related to the alternative glucose metabolic pathway; STUB1 a ubiquitin ligase which modulates the activity of several chaperones; QSOX2 a member of the sulfhydryl oxidase/quiescin-6 family which participates in disulfide bond formation of secreted proteins; KIF11, a motor protein involved in spindle dynamics during mitosis; NMNAT1, which plays a key role in the biosynthesis of nicotinamide adenine dinucleotide (NAD), and lastly, NDC1, a

transmembrane nucleoporin that participates in the formation of the nuclear pore complex.

The enhancers also correspond to genes from diverse ontology groups (Fig. 3B). Amongst these, we found homologues for WAPL, a conserved cohesion release factor that regulates sister chromatid cohesion in mitosis; UBR5, an E3 ubiquitin-protein ligase involved in protein degradation; CWC15, a protein involved in pre-mRNA splicing and a component of the spliceosome; ZFPL1, a factor involved in ER to Golgi transport and essential for cis-Golgi integrity; TUBB2A, encodes the beta-tubulin subunit of the microtubules; and RNGTT, an enzyme involved in mRNA capping. Importantly, none of the enhancer RNA lines disrupts the morphology of the eye when expressed in the absence of mutant TDP-43, confirming their specific interactions (Fig. 3B, top row).

Identification of compounds with therapeutic potential

We identified compounds that showed anti-correlated perturbation patterns using the CMAP tool, which forms a starting point for drug repurposing efforts. Anti-ALS candidate drugs which could be considered for drug repurposing are presented in Table 3. Amongst our top candidates are nutraceuticals such as ginsenoside, ginkgolide B, curcumin, luteolin, forskolin and ipriflavone, statins including atorvastatin and simvastatin, drugs which modulate dopamine receptors and pathways including dopamine, clozapine, fenoldopam, and raclopride, the thiazolidinedione ciglitazone, and anti-inflammatories and antimicrobials including mesalazine, sulfasalazine, indobufen, etodolac, cinalukast, vancomycin, econazole, sulfisoxazole, terbinafine and tobramycin. Currently, FDA-approved ALS drugs, edaravone and riluzole, ranked within the top 13% of the 59,820 compounds tested. We want to empathize that none of the identified candidate drugs has been proven to be an effective treatment for ALS at this stage and we strongly advise to not consume the named compounds for this purpose. Further studies using iPSC-based assays are necessary to determine whether the thus identified candidate drugs are effective in patients, followed by an adequately powered and designed placebo-controlled clinical trial.

Potential Biomarkers from machine learning classification model

We used the identified ALS-related biomarker proteins as a basis to create a classifier that differentiates between ALS and non-ALS patients. The results of our classifier models are shown in Table 4. Using a ridge regression classifier, we were able to create a model that is able to fit the training set perfectly. We selected a ridge classifier as this type of classifier is more resistant to overfitting in the context of a limited number of samples and a high amount of features. On the unseen test set, it has precision of 97% for confirmation of ALS status and identifies patients with a recall of 78%. The high precision indicates some degree of usefulness of this model to provide a biomarker-based confirmation of ALS status to what is generally a clinical diagnosis. The recall of 78% indicates that only 4 out of 5 patients are picked up by this model and we were unable to enhance performance, as our fit with the training set reaches 100% and hence no further optimization takes place. We are confident that the inclusion of additional patients or control samples would improve classifier performance, but not fundamentally alter the involved proteins. Moreover, we created a minimal classifier based on just the top 10 most influential proteins (Table 4) in the first ALS-related classifier. The coefficients indicate that RCHY1, a RING finger and zinc finger domain containing ubiquitin ligase (log2FC −19.200, p 0.0188), HECTD1, an ubiquitin ligase (log2FC −17.412, p 0.0201), RRP15, a ribosomal RNA processing homolog (log2FC −14.430, p 0.0085), GAK, a cyclin G associated kinase (log2FC −20.775, p 0.0037), PHYHIP, a

Table 4. Machine learning classifier for ALS-related genes.

(A) All ALS-related genes

Training set	Precision	Recall	F1 score	*N*
Control	1.00	1.00	1.00	24
ALS	1.00	1.00	1.00	109
		Accuracy	1.00	

Test set	Precision	Recall	F1 score	*N*
Control	0.31	0.80	0.44	5
ALS	0.97	0.78	0.86	40
		Accuracy	0.78	

(B) TOP 10 highest coefficient ALS-related genes

RCHY1, HECTD1, RRP15, GAK, PHYHIP, IWS1, MT1F, SLC9A7, NUP153, TOR1A

Training set	Precision	Recall	F1 score	*N*
Control	0.30	0.92	0.45	24
ALS	0.97	0.52	0.68	109
		Accuracy	0.59	

Test set	Precision	Recall	F1 score	*N*
Control	0.12	0.40	0.18	5
ALS	0.89	0.62	0.74	40
		Accuracy	0.60	

phytanoyl-CoA 2-Hydroxylase interacting protein (log2FC −21.979, p 0.0168), IWS1, a SUPT6H interacting protein (log2FC −8.057, p 0.0230), MT1F, a Metallothionein (log2FC −18.960, p 0.0096), SLC9A7, a solute carrier family 9 protein (log2FC −22.068, p 0.0050), NUP153, a nucleoporin (log2FC −16.684, p 0.0355) and TOR1A, a torsin family 1 protein (log2FC −2.7245, p 0.0195) are most effective at differentiating between ALS patients and healthy individuals. Negative signs for log2FC indicate lower expression levels in ALS patients compared to healthy individuals. This second model showed still considerable discriminatory ability, although with a slight degradation in performance on the unseen test set in terms of precision which drops to 89%, and a reduced recall rate of 62%. Such a restricted model might lend itself to more economical test panels than the model based on 100 proteins.

Discussion

The integration of clinical data with computational biology, machine learning and *in vivo* approaches to address the heterogeneity of ALS disease mechanisms and progression is a strength of our study. Based on the premise that dysfunction of molecular pathways in motor neurons leads to neurodegeneration and each patient's unique clinical presentation, we integrated the clinical history of each patient with the proteomics signature of their iPSC-derived motor neurons. We chose to focus on proteomics data as it is the most direct measurement of cellular conditions. Modelling ALS using iPSCs provides an opportunity to examine molecular changes at the earliest stages of neurodegeneration, unconfounded by changes associated with terminal degeneration, representing early/prodromal therapeutic windows where drugs may be most effective. We recognize that such iPSC-derived cells are unable to replicate all features of ALS-affected human tissues, e.g. the inability to capture the effects of non-cell-autonomous toxicity, but also provide certain advantages as they would be generally representative of an earlier stage of the disease. This is desirable, as late-stage degraded cells are no longer amenable to treatment. Another confounding factor is the exposure of patients to different treatment regimens. Unfortunately, details of prescribed medication and the duration of such treatment are unavailable for our patient samples. However, analysis of iPSC-derived motor neurons should minimize such effects with the exception of persistent epigenetic changes induced through the use of medication.

We contrasted proteomic signatures at the extremes of clinical phenotypes to look for shared pathological pathways or those which differentiate the extremes. These may represent possible ALS-related and prognostic biomarkers respectively, and potential therapeutic targets. Many promising preclinical and early phase ALS clinical trials were unable to be replicated in larger phase 3 confirmatory trials, and biological heterogeneity of ALS has been cited as a contributory factor.[9] We included both C9orf72-positive and cases with undetermined C9orf72 status as the number of such cases in our data set is small (only 7/149 cases have confirmed C9orf72 high repeat counts with a further 10 cases self-reporting C9orf72 positive) and does not easily allow us to distinguish the effects of C9orf72 from effects observed in spontaneous ALS. Moreover, the data set does not contain asymptomatic C9orf72 carriers in the control group. Unmet medical need to develop biomarkers which could reduce clinical heterogeneity and improve clinical trial design exists. No biomarkers that accurately serve in an ALS-related or prognostic capacity in ALS are available.[9] ALS-related biomarkers may select for cases which are "true" ALS, which addresses the median diagnostic delay in ALS of about 1 year, which affects the initiation of treatment and care planning. Prognostic biomarkers could select patients with a specific disease stage and/or rate of progression, which could aid phenotypic stratification for clinical trial enrolment and enrich clinical trials with patients more likely to show clinical effects within a limited time period, thereby reducing costs of conducting clinical trials. As a case in point, initial phase 3 clinical trial results for edaravone failed to show treatment benefits,[12] possibly due to clinical heterogeneity in the ALSFRS-R slopes obscuring treatment effect,[13] but a subsequent clinical trial which enrolled a particular subset of patients in early disease and ALSFRS-R decline of −1 to −4 over 12 weeks showed a significant 33% reduction in the rate of disease progression with edaravone treatment,[11] which eventually led to FDA-approval. Our study has predicted biomarkers which could be further tested in more easily accessed saliva, tears, cerebrospinal fluid or serum samples from ALS patients. GO Biological pathways of identified proteins were enriched in RNA metabolism, DNA damage repair, and cell cycle mechanisms, which are known pathogenic mechanisms in ALS,[35–37] and therefore validate our approach (Tables S1, S2).

RNA-binding proteins such as TDP-43 and FUS are known to be involved in ALS pathogenesis.[33] Using the Drosophila eye expressing human TDP-43^{M337} as a discovery platform,[29] we identified 19 hits out of 57 candidates from our proteomics analysis. Two hits were recently identified in an independent genome-wide screen in TDP-43-expressing flies and correspond to fly homologues of ARIDB1 and WAPL,[38] confirming their specificity. MED23 is relevant because subunits of the Mediator complex suppressed TDP-43 toxicity in another screen[39] and interact with WAPL,[40] suggesting its

importance in disease pathogenesis. TUBB2A transcripts were dysregulated in the axonal compartment of TDP-43-depleted motor neurons.[41] The identification of modifiers involved in RNA synthesis and processing (POLR2K and CWC15), RNA export (THOC1) and nuclear pore complex formation (NDC1) implicates these biological pathways. Results from the fly do not necessarily translate to human cells but are useful starting points for hypothesis generation.

Not all human genes have fly homologues and available RNAi lines for testing, hence it is important to combine insights from patient iPSC and fly models. Of the proteins screened in the fly model, TOR1A was the only protein screened amongst the top 10 diagnostic biomarkers predicted by machine learning and had no effect on the fly eye phenotype, which suggests that the TOR1A signature is downstream of the causative protein. Overlaps in neurodegenerative pathways mediated by novel biomarkers identified by machine learning and Drosophila screen deserve further study. Protein ubiquitination is promising as it is mediated by UBR5 and STUB1 (Drosophila screen) and RCHY1 and HECTD1 (Machine learning predictions). UBR5 is protective in Huntington's disease[42] and STUB1 mutations have been linked to ataxia and cognitive decline.[43] RCHY1 and HECTD1 have not yet been associated with neurodegeneration. Concurring with the Drosophila screen, machine learning predictions suggest that RNA processing (RRP15) and nuclear pore complex (NUP153) are important biological pathways for further study.

Genes which, when silenced, rescue the ALS phenotype, may be promising therapeutic targets. Advances in gene silencing and editing technologies can permit personalized therapeutic programs involving selective knockdown of these proteins. Improved methods of delivering gene-modifying therapy to the central nervous system in another prototypical motor neuron disease, Spinal Muscular Atrophy (SMA),[44] support the viability of such strategies for ALS.

Anti-ALS candidate drugs which we identified included nutraceuticals, statins, drugs which modulate dopamine pathways, anti-inflammatories and antimicrobials. We again want to empathize that none of the identified candidate drugs has been proven to be an effective treatment for ALS at this stage. Ginseng alleviates neurological symptoms in the SOD1 mouse model of ALS.[45] A possible role for platelet-activating factor receptor (PAFR) specific inhibitors has been postulated based on overexpression of PAFR in the spinal cords of SOD1 mice,[46] and ginkgo is a powerful PAFR inhibitor. Curcumin might have a beneficial effect on neurodegenerative diseases through anti-inflammatory and antioxidant properties.[47] Ropinirole, a dopamine agonist, was recently identified as a candidate therapeutic agent for ALS using iPSC-based drug discovery,[48] which supports the consideration of drugs which modulate the dopamine pathway. Given that inflammation has been identified in ALS pathogenesis,[8] anti-inflammatories and anti-microbials may be attractive anti-ALS candidates. In this class, celecoxib, a Cox2 inhibitor and ceftriazone, an antibiotic, had previously been tested in ALS clinical trials but did not show clinical efficacy,[49,50] perhaps due to clinical and biological heterogeneity of ALS patients. However, it is also possible, that the drugs thus identified are ineffective for ALS and the similarity in affected pathways is purely coincidental. Further studies should determine whether candidate drugs are effective in ALS patients using iPSC-based assays, followed by a placebo-controlled clinical trial.

Potential limitations of our study arise from the available data included in this study, which is primarily based on iPSC-derived differentiated motor neurons that have been derived with a single protocol. These cells are patient-derived, but are not direct patient samples and, hence, we cannot rule out that the preparation protocol introduces biases reflected in the data presented here. Unfortunately, omics data for the primary patient samples is currently not available. Patient ages in the ALSFRS-R baseline group differ significantly in age between early- and late-stage patients with late-stage patients on average about 5 years younger, which may result in confounding effects with the observed levels of neurodegeneration. Additionally, the limited number of C9orf72 samples and the lack of asymptomatic C9orf72 carriers in the controls, alongside similar limitations concerning ATXN2 carriers, constrain our ability to differentiate these genetic sub-populations from the remainder of the patient samples. Our set of 10 proteins identified as an ALS-related using machine learning methods have been determined purely through statistical criteria and have not been confirmed to play a role in ALS pathogenesis through biological studies. Lastly, the TDP-43 Drosophila model employed here may not accurately reflect the progression of ALS and FTD in humans, and hence our results related to genes having disease-altering effects in TDP-43-expressing flies may not be directly transferrable to humans. A follow-up study will be needed to define the specific role of the candidate proteins in motor neurons of the TDP-43 fly model to identify targets with higher translational potential.

Conclusion

Our study integrates clinical data with computational biology, machine learning and *in vivo* approaches to address the heterogeneity of ALS disease mechanisms. Overall, these results illustrate the power of coupling

proteomics and bioinformatics platforms with in vivo validation in flexible model organisms. Since ALS clinical presentation is highly heterogeneous, single drugs are unlikely to be effective across all cases. Therefore, precision medicine approaches will be critical to identify treatment responders.

Acknowledgements

This work was partially supported by National Institutes of Health grant R01AG059871 to DER-L, A*STAR to C.J.J.Y.

Author Contributions

Crystal Jing Jing Yeo contributed to the conception and design of the study; Crystal Jing Jing Yeo, Roland G. Huber, Diego E. Rincon-Limas, Swapnil Pandey, Deepak Chhangani, Nathan P. Staff contributed to the acquisition and analysis of data; Crystal Jing Jing Yeo, Roland G. Huber, Diego E. Rincon-Limas, Swapnil Pandey, Deepak Chhangani, Nathan P. Staff contributed to the drafting of text or preparing the figures.

Conflicts of Interest

The authors would like to declare that no potential conflicts of interest with any commercial entities relating to this study.

References

1. Xu L, Liu T, Liu L, et al. Global variation in prevalence and incidence of amyotrophic lateral sclerosis: a systematic review and meta-analysis. J Neurol. 2020;267(4):944-953.
2. Westeneng H-J, Debray TPA, Visser AE, et al. Prognosis for patients with amyotrophic lateral sclerosis: development and validation of a personalised prediction model. Lancet Neurol. 2018;17(5):423-433.
3. Raaphorst J, Beeldman E, De Visser M, et al. A systematic review of behavioural changes in motor neuron disease. Amyotroph Lateral Scler. 2012;13(6):493-501.
4. Taylor JP, Brown RH, Cleveland DW. Decoding ALS: from genes to mechanism. Nature. 2016;539(7628):197-206.
5. Elden AC, Kim H-J, Hart MP, et al. Ataxin-2 intermediate-length polyglutamine expansions are associated with increased risk for ALS. Nature. 2010;466 (7310):1069-1075.
6. Muyderman H, Chen T. Mitochondrial dysfunction in amyotrophic lateral sclerosis–a valid pharmacological target? Br J Pharmacol. 2014;171(8):2191-2205.
7. Pollari E, Goldsteins G, Bart G, et al. The role of oxidative stress in degeneration of the neuromuscular junction in amyotrophic lateral sclerosis. Front Cell Neurosci. 2014;8:131.
8. Lyon MS, Wosiski-Kuhn M, Gillespie R, Caress J, Milligan C. Inflammation, immunity, and amyotrophic lateral sclerosis: I. Etiology and Pathology. Muscle Nerve. 2019;59 (1):10-22.
9. Goyal NA, Berry JD, Windebank A, et al. Addressing heterogeneity in amyotrophic lateral sclerosis CLINICAL TRIALS. Muscle Nerve. 2020;62(2):156-166.
10. Saitoh Y, Takahashi Y. Riluzole for the treatment of amyotrophic lateral sclerosis. Neurodegener Dis Manag. 2020;10(6):343-355.
11. Abe K, Aoki M, Tsuji S, et al. Safety and efficacy of edaravone in well defined patients with amyotrophic lateral sclerosis: a randomised, double-blind, placebo-controlled trial. Lancet Neurol. 2017;16(7):505-512.
12. Abe K, Itoyama Y, Sobue G, et al. Confirmatory double-blind, parallel-group, placebo-controlled study of efficacy and safety of edaravone (MCI-186) in amyotrophic lateral sclerosis patients. Amyotroph Lateral Scler Front Degener. 2014;15(7–8):610-617.
13. Group E (MCI-186) ALS 16 S. A post-hoc subgroup analysis of outcomes in the first phase III clinical study of edaravone (MCI-186) in amyotrophic lateral sclerosis. Amyotroph. Lateral Scler. Front. Degener. 2017;18(Supp 1):11-19.
14. Baxi EG, Thompson T, Li J, et al. Answer ALS, a large-scale resource for sporadic and familial ALS combining clinical and multi-omics data from induced pluripotent cell lines. Nat Neurosci. 2022;25(2):226-237.
15. Smyth GK. limma: linear models for microarray data. In: Gentleman R, Carey VJ, Huber W, Irizarry RA, Dudoit S, eds. Bioinformatics and computational biology solutions using R and bioconductor. Statistics for biology and health. Springer; 2005:397-420.
16. Gentleman RC, Carey VJ, Bates DM, et al. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. 2004;5(10):1-16.
17. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing; 2013.
18. Cedarbaum JM, Stambler N, Malta E, et al. The ALSFRS-R: a revised ALS functional rating scale that incorporates assessments of respiratory function. J Neurol Sci. 1999;169 (1–2):13-21.
19. Woolley SC, York MK, Moore DH, et al. Detecting frontotemporal dysfunction in ALS: utility of the ALS cognitive behavioral screen (ALS-CBS™). Amyotroph Lateral Scler. 2010;11(3):303-311.
20. Labra J, Menon P, Byth K, et al. Rate of disease progression: a prognostic biomarker in ALS. J. Neurol. Neurosurg. Psychiatry. 2016;87(6):628-632.
21. Pérez-Gracia JL, Gúrpide A, Ruiz-Ilundain MG, et al. Selection of extreme phenotypes: the role of clinical observation in translational research. Clin Transl Oncol off

Publ Fed Spanish Oncol Soc Natl Cancer Inst Mex. 2010;12(3):174-180.
22. McKinney W. Pandas: a foundational python library for data analysis and statistics. Python High Perform Sci Comput. 2011;14(9):1-9.
23. Virtanen P, Gommers R, Oliphant TE, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods. 2020;17(3):261-272.
24. Hunter JD. Matplotlib: a 2D graphics environment. Comput Sci Eng. 2007;9(3):90-95.
25. Kuhn M, von Mering C, Campillos M, et al. STITCH: interaction networks of chemicals and proteins. Nucleic Acids Res. 2007;36(suppl_1):D684-D688.
26. Kuhn M, Szklarczyk D, Pletscher-Frankild S, et al. STITCH 4: integration of protein–chemical interactions with user data. Nucleic Acids Res. 2014;42(D1):D401-D407.
27. Hoerl AE, Kennard RW. Ridge regression: biased estimation for nonorthogonal problems. Dent Tech. 1970;12(1):55-67.
28. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in python. J Mach Learn Res. 2011;12:2825-2830.
29. Ritson GP, Custer SK, Freibaum BD, et al. TDP-43 mediates degeneration in a novel Drosophila model of disease caused by mutations in VCP/p97. J Neurosci. 2010;30(22):7729-7739.
30. Singh MD, Jensen M, Lasser M, et al. NCBP2 modulates neurodevelopmental defects of the 3q29 deletion in Drosophila and Xenopus laevis models. PLoS Genet. 2020;16(2):e1008590.
31. Lamb J, Crawford ED, Peck D, et al. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. Science. 2006;313 (5795):1929-1935.
32. Lamb J. The connectivity map: a new tool for biomedical research. Nat Rev Cancer. 2007;7(1):54-60.
33. Qu XA, Rajpal DK. Applications of connectivity map in drug discovery and development. Drug Discov Today. 2012;17(23–24):1289-1298.
34. Subramanian A, Narayan R, Corsello SM, et al. A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. Cell. 2017;171(6):1437-1452.
35. Kok JR, Palminha NM, Dos Santos SC, et al. DNA damage as a mechanism of neurodegeneration in ALS and a contributor to astrocyte toxicity. Cell Mol Life Sci. 2021;78(15):5707-5729.
36. Butti Z, Patten SA. RNA dysregulation in amyotrophic lateral sclerosis. Front Genet. 2019;9:712.
37. Ranganathan S, Bowser R. Alterations in G1 to S phase cell-cycle regulators during amyotrophic lateral sclerosis. Am J Pathol. 2003;162(3):823-835.
38. Kankel MW, Sen A, Lu L, et al. Amyotrophic lateral sclerosis modifiers in drosophila reveal the phospholipase d pathway as a potential therapeutic target. Genetics. 2020;215(3):747-766.
39. Azpurua J, El-Karim EG, Tranquille M, Dubnau J. A behavioral screen for mediators of age-dependent TDP-43 neurodegeneration identifies SF2/SRSF1 among a group of potent suppressors in both neurons and glia. PLoS Genet. 2021;17(11):e1009882.
40. Haarhuis JHI, van der Weide RH, Blomen VA, et al. A mediator-cohesin axis controls heterochromatin domain formation. Nat Commun. 2022;13(1):1-10.
41. Briese M, Saal-Bauernschubert L, Lüningschrör P, et al. Loss of Tdp-43 disrupts the axonal transcriptome of motoneurons accompanied by impaired axonal translation and mitochondria function. Acta Neuropathol Commun. 2020;8(1):1-16.
42. Koyuncu S, Saez I, Lee HJ, et al. The ubiquitin ligase UBR5 suppresses proteostasis collapse in pluripotent stem cells from Huntington's disease patients. Nat Commun. 2018;9(1):1-22.
43. Chen D-H, Latimer C, Yagi M, et al. Heterozygous STUB1 missense variants cause ataxia, cognitive decline, and STUB1 mislocalization. Neurol Genet. 2020;6(2):1-13.
44. Yeo CJJ, Simmons Z, De Vivo DC, Darras BT. Ethical perspectives on treatment options with spinal muscular atrophy patients. Ann. Neurol. 2022;91(3):305-316.
45. Nam SM, Choi JH, Choi S-H, et al. Ginseng gintonin alleviates neurological symptoms in the G93A-SOD1 transgenic mouse model of amyotrophic lateral sclerosis through lysophosphatidic acid 1 receptor. J Ginseng Res. 2021;45(3):390-400.
46. Briones MRS, Snyder AM, Ferreira RC, Neely EB, Connor JR, Broach JR. A possible role for platelet-activating factor receptor in amyotrophic lateral sclerosis treatment. Front Neurol. 2018;9:39.
47. Monroy A, Lithgow GJ, Alavez S. Curcumin and neurodegenerative diseases. Biofactors. 2013;39(1):122-132.
48. Okano H, Yasuda D, Fujimori K, Morimoto S, Takahashi S. Ropinirole, a new ALS drug candidate developed using iPSCs. Trends Pharmacol Sci. 2020;41(2):99-109.
49. Cudkowicz ME, Shefner JM, Schoenfeld DA, et al. Trial of celecoxib in amyotrophic lateral sclerosis. Ann Neurol. 2006;60(1):22-31.
50. Cudkowicz ME, Titus S, Kearney M, et al. Safety and efficacy of ceftriaxone for amyotrophic lateral sclerosis: a multi-stage, randomised, double-blind, placebo-controlled trial. Lancet Neurol. 2014;13(11):1083-1091.
51. Laferrière F, Maniecka Z, Pérez-Berlanga M, et al. TDP-43 extracted from frontotemporal lobar degeneration subject brains displays distinct aggregate assemblies and neurotoxic effects reflecting disease progression rates. Nat Neurosci. 2019;22(1):65-77.
52. Iridoy MO, Zubiri I, Zelaya MV, et al. Neuroanatomical quantitative proteomics reveals common pathogenic biological routes between amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD). Int J Mol Sci. 2018;20(1):4.

53. Gozal YM, Dammer EB, Duong DM, et al. Proteomic analysis of hippocampal dentate granule cells in frontotemporal lobar degeneration: application of laser capture technology. Front Neurol. 2011;2:24.
54. Umoh ME, Dammer EB, Dai J, et al. A proteomic network approach across the ALS-FTD disease spectrum resolves clinical phenotypes and genetic vulnerability in human brain. EMBO mol Med. 2018;10(1):48-62.
55. Pun FW, Liu BHM, Long X, et al. Identification of therapeutic targets for amyotrophic lateral sclerosis using PandaOmics–an AI-enabled biological target discovery platform. Front Aging Neurosci. 2022;14:914017.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Figure S1. Comparison with human tissue studies. Venn diagrams showing the overlap of proteins identified in this study with previous studies by Laferriére et al.,[51] Iridoy et al.,[52] Gozal et al.,[53] Umoh et al.[54] and Pun et al.,[55] in human tissues. Our study uses iPSC-derived motor neuron data provided by the AnswerALS consortium. We compare our findings with previous literature studying proteomics in human tissues and find a degree of overlap, particularly with the study of Umoh et al. In general, the overlap between different human tissue proteomics studies is limited as evidenced by the comparison of various human tissue studies.

Figure S2. Comparison between previous human tissue studies. Overlap of proteins identified in previous studies of ALS and FTLD literature. We find limited overlap between proteins in proteomics studies by Laferriére et al.,[51] Iridoy et al.,[52] Gozal et al.,[53] Umoh et al.[54] and Pun et al.[55]

Table S1. GO of heatmaps. The top five gene ontology terms resulting from STITCH analysis are shown for GO biological processes.

Table S2. Protein lists from heatmaps. Protein lists, log2-fold change and *p*-values from Limma for all contrasts. Individual lists correspond to the three groups in the Venn diagrams of Figure 1.

Table S3. RNAi screening results in flies expressing mutant TDP-43^{M337V}. Columns indicate the names of the human and fly homologues, the catalogue number of each stock used and the corresponding score in the validation.

