

RESEARCH

Open Access



Identification and validation of a tear fluid-derived protein biomarker signature in patients with amyotrophic lateral sclerosis

Lena-Sophie Scholl^{1†}, Antonia F. Demleitner^{1†}, Jenny Riedel^{2,3}, Seren Adachi^{4,5}, Lisa Neuenroth⁶, Clara Meijs^{2,3}, Laura Tzeplaeff¹, Lucas Caldi Gomes¹, Ana Galhoz^{2,3}, Isabell Cordts¹, Christof Lenz^{6,7*†}, Michael Menden^{2,4*†} and Paul Lingor^{1,8,9*†}

Abstract

The diagnosis of Amyotrophic Lateral Sclerosis (ALS) remains challenging, particularly in early stages, where characteristic symptoms may be subtle and nonspecific. The development of disease-specific and clinically validated biomarkers is crucial to optimize diagnosis. Here, we explored tear fluid (TF) as a promising ALS biomarker source, given its accessibility, anatomical proximity to the brainstem as an important site of neurodegeneration, and proven discriminative power in other neurodegenerative diseases. Using a discovery approach, we profiled protein abundance in TF of ALS patients ($n=49$) and controls ($n=54$) via data-independent acquisition mass spectrometry. Biostatistical analysis and machine learning identified differential protein abundance and pathways in ALS, leading to a protein signature. These proteins were validated by Western blot in an independent cohort (ALS $n=51$; controls $n=52$), and their discriminatory performance was assessed in-silico employing machine learning. 876 proteins were consistently detected in TF, with 106 differentially abundant in ALS. A six-protein signature, including CRYM, PFKL, CAPZA2, ALDH16A1, SERPINC1, and HP, exhibited discriminatory potential. We replicated significant differences of SERPINC1 and HP levels between ALS and controls across the cohorts, and their combination yielded the best in-silico performance. Overall, this investigation of TF proteomics in ALS and controls revealed dysregulated proteins and pathways, highlighting inflammation as a key disease feature, strengthening the potential of TF as a source for biomarker discovery.

Keywords Amyotrophic lateral sclerosis, Biomarker, Tear fluid, Diagnosis, Proteomics, Neurodegeneration

[†]Lena-Sophie Scholl, Antonia F. Demleitner, Christof Lenz, Michael Menden and Paul Lingor contributed equally to this work.

*Correspondence:

Christof Lenz

christof.lenz@med.uni-goettingen.de

Michael Menden

michael.menden@unimelb.edu.au

Paul Lingor

paul.lingor@tum.de

¹Department of Neurology, TUM University Hospital Munich, Munich, Germany

²Computational Health Center, Helmholtz Munich, Neuherberg, Germany

³Department of Biology, Ludwig-Maximilians University, Munich, Martinsried, Germany

⁴Department of Biochemistry and Pharmacology, Bio21 Molecular Science and Biotechnology Institute, The University of Melbourne, Parkville, Victoria, Australia

⁵Melbourne Bioinformatics, Faculty of Medicine, Dentistry and Health Sciences, The University of Melbourne, Parkville, Victoria, Australia

⁶Department of Clinical Chemistry, University Medical Center Göttingen, Göttingen, Germany

⁷Bioanalytical Mass Spectrometry Group, Max Planck Institute for Multidisciplinary Sciences, Göttingen, Germany

⁸German Center for Neurodegenerative Diseases (DZNE), Munich, Germany

⁹Munich Cluster for Systems Neurology (SyNergy), Munich, Germany



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Introduction

Amyotrophic Lateral Sclerosis (ALS) is the most common motor neuron disease, marked by progressive degeneration of both the upper (UMN) and lower motor neurons (LMN), leading to muscle weakness and atrophy [1]. ALS diagnosis primarily relies on clinical assessments of UMN and LMN dysfunction [2, 3], but its heterogeneous presentation in early stages often delays a definite diagnosis by 10–16 months after symptom onset [4], imposing a substantial burden on patients and impending prompt treatment. The prognosis for most ALS patients remains poor [5]. Early diagnosis and timely care are essential for patient quality of life, highlighting the urgent need for biomarkers to enable rapid, accurate diagnosis and to monitor disease progression and treatment response.

Previous studies have identified several biomarker candidates in cerebrospinal fluid (CSF) and blood, including neurofilaments [6], chitinases [7], TDP-43 [8], cystatin C [9], creatine kinase [10], and tau protein [11]. The most convincing data for urinary biomarkers exists for neopterin [12] and p75ECD [13]. Among these, neurofilaments and particularly neurofilament light (NfL), are the most extensively studied in neurodegenerative diseases (NDD) as they are believed to reflect axonal damage and degeneration [14]. NfL has shown promise in diagnostics [15, 16], prediction of prognosis [17], but most importantly as a surrogate marker for therapeutic response in clinical trials [18]. Recently, chitinases CHIT1, CHI3L1, and CHI3L2 have emerged as potential CSF and blood biomarkers for ALS, as their expression is associated with neuroinflammation [19], with elevated CSF chitinase levels observed in ALS patients compared to controls and ALS mimics [20].

High-throughput omics technologies, particularly proteomics, have transformed biomarker research by enabling large-scale detection of several proteins simultaneously through mass spectrometry. This advancement has facilitated the identification of differentially abundant proteins associated with ALS [21, 22]. Several of these proteins have been validated as disease biomarkers and successfully translated into clinical applications [23].

While most biomarkers have been identified in CSF and blood, both sources have limitations. Blood's selective permeability across the blood-brain barrier restricts direct contact with the site of neurodegeneration. Although CSF contains molecules that are more relevant to central nervous system (CNS) pathology, its collection is invasive, costly, and less accepted by patients. In contrast, tear fluid (TF) is readily and rapidly accessible. Unstimulated tears can be collected using Schirmer strips, a non-invasive and cost-efficient method suitable for bedside clinical testing [24].

TF production is stimulated by parasympathetic fibers from the superior salivatory nucleus in the brainstem [25]. This anatomical connection, along with TF being an ultrafiltrate of blood, suggests its potential as an indicator for disease-specific alterations [26]. Elevated levels of alpha-synuclein and NfL have been found in the TF of Parkinson's disease patients compared to controls [27, 28]. Additionally, Alzheimer's disease-specific biomarkers, including amyloid-beta peptides and tau proteins, have been detected in TF [29].

To investigate TF as a biomarker source in ALS, we conducted a study evaluating a protein signature for ALS prediction. Here, we present the characterization of the TF proteome in a cohort of 49 ALS and 54 control subjects using data-independent mass spectrometry (DIA-MS). A protein signature was identified through biostatistical and machine learning models and subsequently validated with Western blot analysis and machine learning models in an independent cohort of 51 ALS and 52 controls.

Methods

Tear fluid (TF) collection

TF was collected between September 2019 and June 2022 in the in- and outpatient clinics of the Department of Neurology, TUM University Hospital, Germany. The study adheres to the Declaration of Helsinki of the World Medical Association. Permission was obtained from the Ethics Committee of the Technical University of Munich, School of Medicine (approval number: 9/15S). Unstimulated TF was collected with Schirmer strips (Madhu Instruments Pvt. Ltd., New Delhi, India) according to a standardized protocol [30]. Prior to sampling, participants were educated about the procedure, and written informed consent was acquired. For the sampling process, the upper part of the strip was inserted into the lower conjunctival sac at a one-third distance from the lateral canthus of both eyes. TF collection lasted 5 min per patient. Once finished, the Schirmer strips were removed, and the wetting length (WL) in millimeters was recorded for each eye. The strips were then separately transferred into collection tubes and kept on ice before being immediately stored at -80 °C until further use.

Study population

For the discovery cohort, 49 patients with probable or definite ALS, according to the revised El Escorial criteria, were selected and compared to 54 age-matched control subjects without evidence of neurodegeneration (Fig. 1a). Due to the exploratory nature of this analysis, the sample size was determined based on previously published cohort sizes as well as the availability of samples in a rare disease [26]. Since secreted tear volume is known to be positively correlated with tear protein concentration [31,

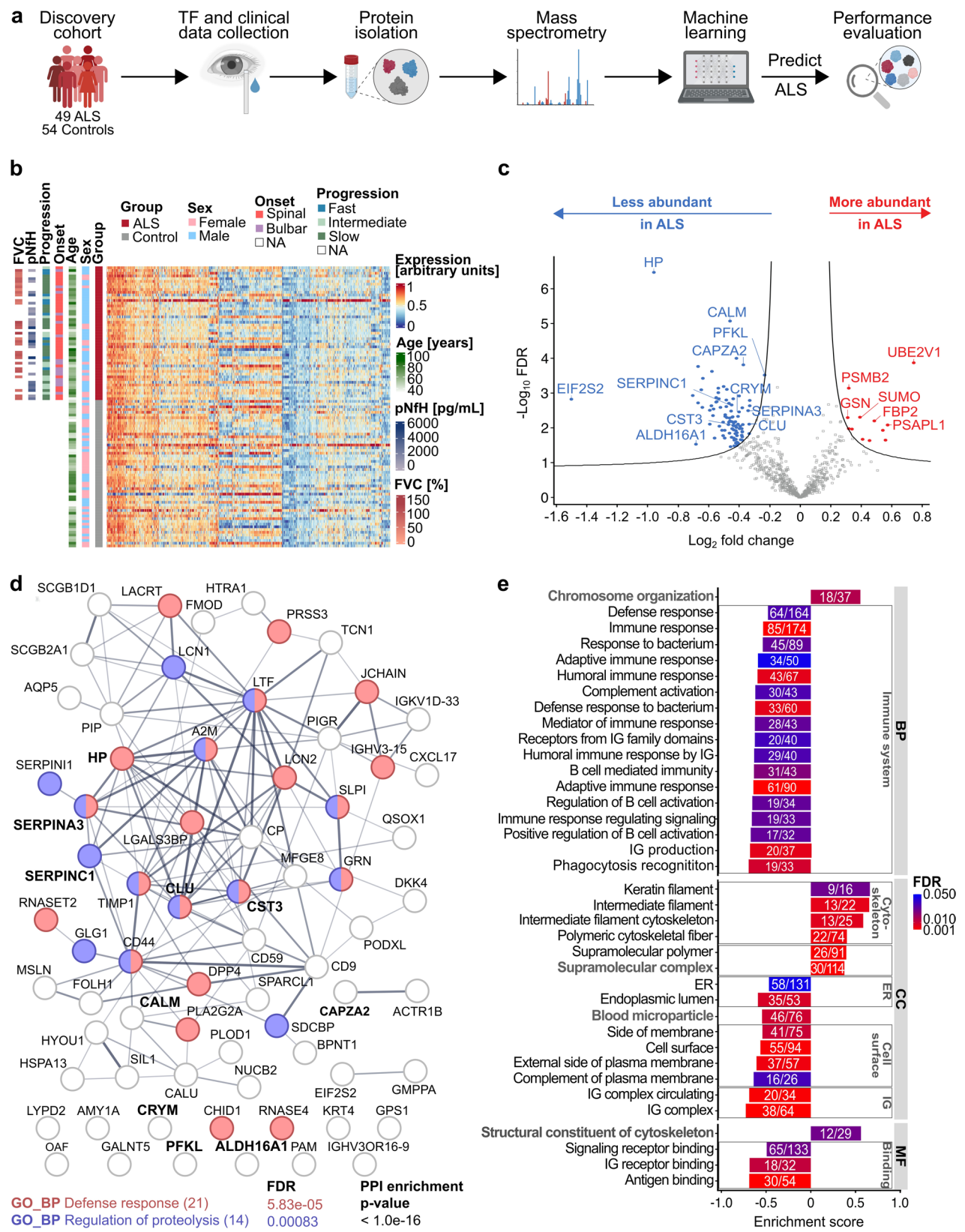


Fig. 1 (See legend on next page.)

(See figure on previous page.)

Fig. 1 Overview of the TF proteomics results. **a** Experimental workflow for the discovery approach. TF and clinical data were collected from ALS ($n=49$) and control subjects ($n=54$). TF proteins were analyzed by mass spectrometry, and machine learning models were used to predict ALS/control status. Created with BioRender.com. **b** Heatmap of normalized TF protein abundance in the discovery cohort, with unsupervised hierarchical clustering of 876 consistently quantified proteins. Rows represent proteins, columns represent samples. Clinical features, including sex, age [years], onset, progression rate, CSF pNfH levels [pg/mL], and FVC [%], are annotated. **c** Volcano plot of the differential protein abundance between ALS and control subjects. x-axis represents $\log_2$ (fold change), y-axis represents $-\log_{10}$ (FDR). Significantly less abundant proteins (blue) and more abundant proteins (red) in ALS are shown. Significance was estimated by PERSEUS standard settings, including FDR < 0.1 and $S_0=0.1$. **d** STRING PPI network analysis for significantly less abundant proteins. Hub proteins are associated with “defense response” (red) and “regulation of proteolysis” (purple). The number of proteins in parentheses indicates the count of proteins assigned to each pathway. PPI enrichment p -value ($<1.0e-16$) is displayed. **e** GSEA of TF proteome in ALS and controls, showing significant GO terms for biological processes, cellular components, and molecular functions. Enrichment scores (x-axis) and annotated GO terms (y-axis) are shown. Bar colors reflect FDR, with pathway counts listed. Pathways referring to the same category were manually summarized under generic terms (bolded grey). Benjamini-Hochberg correction and an FDR cut-off of 5% were applied. **Abbreviations:** ALS, Amyotrophic Lateral Sclerosis; TF, tear fluid; FVC, forced vital capacity; CSF, cerebrospinal fluid; pNfH, phosphorylated neurofilament heavy chain; GO, gene ontology; BP, biological process; PPI, protein-protein interaction; CC, cellular component; MF, molecular function; ER, endoplasmic reticulum; IG, immunoglobulins; HP, haptoglobin; CALM, calmodulin; PFKL, phosphofructokinase, liver type; CAPZA2, f-actin-capping protein subunit alpha-2; EIF2S2, eukaryotic translation initiation factor 2 subunit 2; SERPINC1, antithrombin-III; CRYM, mu-crystallin; SERPINA3, alpha-1-antichymotrypsin; CST3, cystatin C; ALDH16A1, aldehyde dehydrogenase family 16 member A1; CLU, clusterin; UBE2V1, ubiquitin-conjugating enzyme E2 variant 1; PSMB2, proteasome subunit beta type-2; SUMO, small-ubiquitin modifier; GSN, gelsolin; FBP2, fructose-1,6-bisphosphatase isozyme 2; PSAPL1, proactivator polypeptide-like 1

32], only TF samples with sufficient WL (≥ 5 mm) were included in the study.

An independent validation cohort, comprising 51 ALS and 52 age- and sex-matched control patients without signs of neurodegenerative disease adhering to the same inclusion criteria as those of the discovery cohort, was selected. The demographic and clinical characteristics of the validation cohort were matched to those of the discovery cohort (Table 1).

Demographic and clinical data analysis

The following clinical data were extracted from medical records and interviews: age, sex, eye diseases, eye medication and/or lubricants, and the use of contact lenses. For ALS patients, we additionally assessed ALS-associated genetic variants, the time point of first symptoms and diagnosis, and the revised ALS Functional Rating Scale (ALSFRS-R) score at two time points (first recorded and at TF sampling). According to initial symptoms, patients were divided into two phenotypic groups: spinal and bulbar onset. Disease duration was calculated as the difference between the self-reported appearance of symptoms and the date of TF collection. Disease severity was assessed by the ALSFRS-R sum score [33] and the forced vital capacity (FVC, expressed as a percentage of normal) at the time of TF sampling. Disease progression was determined by calculating the difference between the ALSFRS-R at symptom onset and the first recorded ALSFRS-R divided by the time (in months) between both visits (Δ FRS) [34]. ALS patients were assigned accordingly to three subgroups: fast progressors (Δ FRS > 1.11 points/month), intermediate progressors (Δ FRS 0.47–1.11 points/month), and slow progressors (Δ FRS < 0.47 points/month) [35]. CSF levels of phosphorylated neurofilament heavy chain (pNfH) were measured as part of routine diagnostic procedures when the diagnosis of ALS was suspected.

Mass spectrometry (MS) sample preparation

Proteomic analyses were performed in a blinded manner without knowledge of the cohort affiliation of the samples. Proteins were extracted from Schirmer strips without additional washing or pre-conditioning steps to avoid selective loss of soluble proteins. Individual strips were extracted overnight using RIPA buffer augmented with a phosphatase inhibitor cocktail (Roche, Mannheim, Germany) at 37 °C, 800 rpm in a thermoshaker. Following centrifugation ($17,000 \times G$, 4 °C, 30 min), the supernatant was collected, and protein concentration was determined by bicinchoninic acid (BCA) assay. Aliquots containing 20 μ g protein were filled up with 100 mM aqueous ammonium bicarbonate pH 8.0 to a volume of 80 μ L. Proteins were reduced and alkylated with tris-(2-carboxyethyl)phosphine (TCEP) (20 mM) and iodoacetamide (50 mM) in a thermoshaker (600 rpm, 37 °C, 30 min, in the dark). A modified SP3 protocol [36] was used to perform protein digestion with trypsin. 20 μ g paramagnetic beads (SpeedBeads magnetic carboxylate modified, Cytiva, Marlborough, USA) were mixed with the samples in 0.2 mL tubes, and 100% ethanol (Merck, Darmstadt, Germany) was added to an ethanol concentration of 50% (volume per volume (v/v)). The samples were incubated in a thermoshaker (800 rpm, 24 °C, 5 min), then transferred to a magnetic rack, and magnetic beads captured on the magnets for 2 min. The supernatant was removed, and the bound beads were washed three times with 80% ethanol in water. Samples were allowed to air-dry, then removed from the rack and incubated with porcine trypsin (Promega, Madison, USA) in 100 mM ABC at an enzyme-to-protein ratio of 1:20 (600 rpm, 37 °C, overnight) for digestion. Following digestion, the supernatant was removed, the beads washed again with 100 mM ABC, and the combined supernatant and wash solution acidified with 10% aqueous trifluoroacetic acid (TFA) to a final TFA concentration of 0.4%. The acidified samples

Table 1 Demographic and clinical data of the discovery and validation cohort

	Discovery cohort			Validation cohort			DC vs. VC
	Controls	ALS	<i>p</i> -value	Controls	ALS	<i>p</i> -value	<i>p</i> -value
Demographics							
Patients <i>n</i>	54	49		52	51		
Male / female <i>n</i> (% female)	22 / 32 (58.26)	35 / 14 (28.57)	0.003 ^a	29 / 23 (44.23)	27 / 24 (47.06)	0.844 ^a	0.020 ^c
Age [years]							
Median (IQR)	60 (48, 74)	60 (53, 73)	0.302 ^b	63 (57, 76)	64 (57, 75)	0.832 ^b	0.094 ^d
Disease-specific parameters							
Subgroup							1 ^a
sALS <i>n</i> (%)	NA	47 (95.92)		NA	48 (94.12)		
fALS <i>n</i> (%)	NA	2 (4.08)		NA	3 (5.88)		
Gene mutation							0.098 ^a
C9orf72 <i>n</i> (%)	NA	1 (2.04)		NA	6 (11.77)		
SOD1 <i>n</i> (%)	NA	1 (2.04)		NA	0 (0)		
Variants associated with ALS <i>n</i> (%)	NA	1 (2.04)		NA	2 (3.92)		
VUS <i>n</i> (%)	NA	1 (2.04)		NA	3 (5.88)		
Not determined <i>n</i> (%)	NA	36 (73.45)		NA	26 (50.98)		
Stratum of onset							1 ^a
Spinal <i>n</i> (%)	NA	34 (69.39)		NA	36 (70.59)		
Bulbar <i>n</i> (%)	NA	15 (30.61)		NA	15 (29.41)		
Disease duration [years] at TF collection							0.247 ^b
Median (IQR)	NA	20 (9, 37)		NA	17 (12, 20)		
ALSFERS-R at TF collection							0.177 ^e
Median (IQR)	NA	35 (29, 41)		NA	37 (31, 42)		
Progression rate							0.135 ^a
Fast <i>n</i> (%)	NA	10 (20.41)		NA	8 (15.69)		
Intermediate <i>n</i> (%)	NA	12 (24.45)		NA	22 (43.14)		
Slow <i>n</i> (%)	NA	26 (53.06)		NA	18 (35.29)		
Not determined <i>n</i> (%)	NA	1 (2.04)		NA	3 (5.88)		
CSF pNfH [pg/mL]							0.135 ^b
Median (IQR)	NA	1,931 (1,175, 3,155)		NA	2,830 (1,454, 4,744)		
Not determined <i>n</i> (%)	NA	21 (42.86)		NA	13 (25.49)		
FVC [%]							0.703 ^e
Median (IQR)	NA	84 (65, 93)		NA	77 (55, 98)		
Not determined <i>n</i> (%)	NA	20 (40.82)		NA	16 (31.38)		
Ophthalmological parameters							
Cumulative WL of both strips [mm/5 min]							
Median (IQR)	20 (9, 35)	15 (10, 30)	0.486 ^b	21 (11, 30)	19 (11, 28)	0.651 ^b	0.749 ^d
Eye disease <i>n</i> (%)	11 (20.37)	5 (10.20)	0.182 ^a	12 (23.08)	16 (31.37)	0.382 ^a	0.079 ^c
Topical eye substances <i>n</i> (%)	6 (11.11)	3 (6.12)	0.493 ^a	3 (5.77)	2 (3.92)	1 ^a	0.549 ^a
Contact lenses <i>n</i> (%)	3 (5.56)	1 (2.04)	0.619 ^a	1 (1.92)	4 (7.84)	0.205 ^a	0.430 ^a

Abbreviations: DC, discovery cohort; VC, validation cohort; IQR, interquartile range; ALS, amyotrophic lateral sclerosis; sALS, sporadic ALS; fALS, Familial ALS; VUS, variant of uncertain significance; TF, tear fluid; ALSFRS-R, ALS functional rating scale revised; CSF, cerebrospinal fluid; pNfH, phosphorylated neurofilament heavy chain; FVC, forced vital capacity in percent of normal; WL, wetting length; NA, not applicable

^aFisher's exact test ^bMann Whitney test ^cChi-square test ^dKruskal-Wallis test ^eUnpaired t-test

were then desalted on C18 tips (Pierce, Rockford, USA), and the eluates were dried in a SpeedVac and stored at -20 °C until analysis.

MS analysis

For the generation of an MS/MS spectral library, equal aliquots of each sample were pooled and fractionated on a high-performance liquid chromatography system by basic pH reversed-phase C18 chromatography (Thermo Hypersil Gold C18 150 × 1.0 mm, gradient 5–35% aqueous acetonitrile pH 8.0, 200 $\mu\text{L min}^{-1}$). The sample was fractionated into 12 fractions using a concatenated pooling scheme, which were used to generate the library. For MS analysis, samples and reverse phase fractions were taken up in 20 μL loading buffer (0.1% aqueous formic acid, 2% acetonitrile), spiked with iRT peptides (Biognosys, Schlieren, Switzerland) [37], and separated on a nanoElute chromatography system (Bruker, Bremen, Germany). Peptides were enriched on a reversed-phase pre-column (0.3 mm ID × 5 mm PepMap C18 5 Å, Thermo Fisher Scientific, Waltham, USA) and separated on an analytical RP-C18 column (0.075 mm ID × 250 mm Aurora C18, IonOpticks, Collingwood, Australia) using a 60 min (data-dependent acquisition, DDA) or 30 min (data-independent acquisition, DIA) linear gradient of 5–35% acetonitrile/0.1% formic acid (v/v) at 300 nL min^{-1} . The eluents were analyzed on a hybrid ion mobility-quadrupole-time of flight mass spectrometer (timsTOF Pro 2, Bruker, Bremen, Germany). Reversed phase fractions were then analyzed by DDA using a DDA-PASEF acquisition method [38], with a survey scan of m/z 100–1700 and an ion mobility enrichment of charges states 2+ to 4+. Precursors selected for MS/MS were isolated at 1.5 FWHM resolution and fragmented using nitrogen as a collision gas using default collision energy settings and up to ten repetitions per precursor depending on intensity. Four replicate injections per reversed phase fraction were analyzed. For quantitative DIA-MS, MS/MS data were acquired using 32 evenly-sized windows staggered in the ion mobility dimension to cover the 2+ to 4+ precursor population across the 400–1250 m/z range (DIA-PASEF) [39]. Fragments were produced using default collision energy settings at a cycle time of 1.9 s. Two replicate injections per biological sample were acquired.

MS data processing

Data were analyzed using Spectronaut SW v14.2 (Biognosys, Schlieren, Switzerland) [40]. For generation of a spectral library, the DDA-PASEF data were searched against the UniProtKB Homo sapiens reference proteome v01/2020 using default settings at a false discovery rate (FDR) of 1%. Quantitative analysis of DIA-PASEF data was achieved by peak extraction on the precursor/

fragment level using the above spectral library, combining six fragments to the precursor and up to 10 peptides to the protein for obtaining protein area values, each at estimated FDR levels of 1%.

Preprocessing proteomics raw data

The proteomics raw data was filtered to remove proteins with over 66% missing values for at least one disease status. Log₂-transformation was applied to the remaining values, followed by an imputation step done with the open-source software Perseus (v1.6.15.0) under default settings [41]. The data was then normalized using min-max scaling between 0 and 1.

Proteomics differential abundance analysis

Differential abundance analysis was conducted with Perseus v1.6.15.0 using standard settings. Perseus assumes data to be missing not at random (MNAR), with missingness largely driven by a protein's abundance in a given sample. It therefore imputes a normal distribution of values which is down-shifted by 1.8 standard deviations from the distribution of valid values [41]. In brief, after defining the experimental groups and handling the imputation of missing values, differential abundance analysis was performed using a permutation-based FDR method for multiple hypothesis testing, as described in the application manual. The FDR values were derived by comparing the proportion of significant results in the permuted data to those in the observed data. For the current analysis, proteins were deemed statistically differentially expressed with an FDR < 0.1 and S0 = 0.1 (Data S1).

Gene set enrichment and protein-protein interaction network analyses

To investigate the pathways associated with the expressed proteins identified in TF, we performed gene set enrichment analysis (GSEA) using the “GSEA()” function from the R package clusterProfiler (v4.6.2) [42]. For the GSEA, we rank the proteins based on log-fold change, and the GSEA is conducted with a Benjamin-Hochberg correction and an FDR cut-off of 5%. To construct a tissue-specific background for the analysis, we utilize the C5 category from the ‘msigdb’ (v7.5.1) package [43] and filter for proteins that we detected in our samples. Furthermore, following the recommendations of the GSEA User Guide, which advises that gene sets with as few as 2–3 genes can yield statistically significant results [44], we applied the more stringent threshold and only included pathways with a minimum gene set size of three. This approach ensured the inclusion of small but biologically meaningful gene sets while maintaining robustness in the enrichment analysis (Data S2).

Protein-protein interaction (PPI) network analysis with differentially abundant proteins was performed and

visualized using the STRING database v12.0 with default settings [45]. Functionally enriched pathways for biological processes (BP) from gene ontology (GO) were highlighted and annotated to corresponding proteins.

Biomarker identification leveraging machine learning

Unsupervised clustering analysis using uniform manifold approximation and projection (UMAP) and heatmap (R packages ComplexHeatmap v2.20, UMAP v0.2.10.0, R v4.2.3), was performed to stratify samples by disease status, sex, and age. The ALS cohort was further organized by clinical features such as stratum of onset, disease progression rate, FVC, and pNfH levels. Machine learning models, including lasso regression, support vector machine (SVM), and random forest, were applied for biomarker identification, using the caret package (v6.0-94) [46] with 80% training, 20% testing. A nested 10-fold cross-validation to estimate models' hyperparameters as well as a 500 times bootstrapping procedure for performance assessment were employed to prevent overfitting. While group sizes were relatively balanced, stratified sampling was employed to account for the limited sample sizes and to ensure proportional representation during model training and evaluation. The models were selected based on their complementary strengths in biomarker discovery: Lasso regression offers embedded feature selection and interpretability; support vector machines (SVM) are well suited for high-dimensional classification tasks with limited sample size; and random forests, as ensemble models, are robust to overfitting and capable of capturing non-linear relationships. For linear regression, a binomial distribution with a lambda range of 0 to 100 (step size 0.1) was used. The Lasso regression model was run with nested 10-fold cross-validation. The optimal lambda value was selected through hyperparameter tuning within each inner cross-validation fold, and the corresponding model weights of the predictor variables were obtained. SVM analysis (e1071 package, v1.7-13) [47] employed both radial and linear kernels. For random forest, the tune length was set to 15, the tune grid ranged from 1 to 5, and 500 trees were used. Receiver operating characteristic (ROC) curves assessed the performance (Table S1).

To identify the promising biomarker proteins for further validation, we firstly created a tier list of proteins selected by the linear regression model during bootstrapping. Proteins selected in at least 20% of runs and with an SD lower than the mean were considered targets. Only significantly differentially abundant proteins were retained as candidates (Data S3). Next, linear regression models were built on these candidate proteins using their average feature importance from the initial analysis. The optimal protein signature was selected based on AUROC performance (Supplementary Methods).

Protein isolation for western blotting

The upper part of the Schirmer strip, which was placed into the lower conjunctival sac, was discarded to prevent contamination with ocular mucosa proteins. The remaining Schirmer strip was cut into small fragments and transferred into 0.5 mL tubes. Subsequently, RIPA lysis buffer containing phosphatase and protease inhibitors (both Roche, Mannheim, Germany) was added, adapted to the WL of both the Schirmer strips (<40 mm WL: 60 µL RIPA, 40–60 mm WL: 70 µL RIPA, >60 mm WL: 80 µL RIPA). The samples were then incubated for 60 min at 4 °C. The tube containing the incubated fragments was then placed into a bigger 1.5 mL tube, and the base of the smaller tube was perforated. Following centrifugation at 16,000 × G for 30 min at 4 °C, the aqueous proteinaceous phase was carefully transferred to a new tube. Finally, total protein concentration was measured using a BCA assay according to the manufacturer's instructions.

Western blot

Since there is no well-established housekeeping protein for TF, a pooled TF standard comprising samples from patients without ALS served for normalization. 30 µg of TF standard and samples were prepared with 100 mM Dithiothreitol (DTT) reducing agent and diluted in Novex Tris-Glycine SDS sample buffer (both Thermo Fisher Scientific, Waltham, USA). Samples were loaded in triplicates in Novex 10–20% Tris-Glycine precast gels (Thermo Fisher Scientific, Waltham, USA) and wet-transferred to PVDF (Polyvinylidene fluoride) membranes using the mini blot module (Thermo Fisher Scientific, Waltham, USA). To detect all six proteins, the three membranes, each harboring two proteins of interest, were sectioned once at specific molecular weight markers as determined by the prestained protein ladder (Figure S1a–c). Membranes were blocked in 5% skimmed milk diluted in PBS-Tween 20 0.05% for 60 min (for CRYM, CAPZA2, SERPINC1, and HP) or 5% bovine serum albumin (BSA) for 180 min in PBS-Tween 20 0.05% (for PFKL and ALDH16A1). Subsequently, membranes were incubated in blocking buffer overnight at 4 °C with the following primary antibodies: CRYM mouse monoclonal antibody (1:100, Santa Cruz Biotechnology Cat# sc-376687, RRID: AB_11150103); SERPINC1 rabbit monoclonal antibody (1:10,000, Abcam Cat# ab126598, RRID: AB_11128908); CAPZA2 rabbit polyclonal antibody (1:500, Proteintech Cat# 15948-1-AP, RRID: AB_2070148); PFKM and PFKL rabbit monoclonal antibody (1:750, Abcam Cat# ab181064, RRID: AB_3675949); ALDH16A1 rabbit monoclonal antibody (1:1,000, Abcam Cat# ab137082, RRID: AB_2861363); HP rabbit monoclonal antibody (1:200, Abcam Cat# ab131236, RRID: AB_11157376). After washing,

Horseradish peroxidase (HRP) horse anti-mouse IgG H&L secondary antibody (1:10,000, Vector Laboratories Cat# PI-2000-1, RRID: AB_2336177) was used for anti-CRYM, and HRP goat anti-rabbit IgG H&L secondary antibody (1:10,000, Vector Laboratories Cat# PI-1000-1, RRID: AB_2916034) for all the other proteins. Membranes were washed, incubated with Super Signal West Pico PLUS Chemiluminescent Substrate (Thermo Fisher Scientific, Waltham, USA), and developed in a Fusion SL chemiluminescence imaging system (Vilber Lourmat, Collégien, France) for protein detection (full-length as well as complete Western blots of the six proteins are provided in Figure S1d-i and Figure S2). Band intensities of the specific proteins were quantified automatically with ImageJ/Fiji2 (Version 2.1.0/1.53c) without knowledge of the cohort affiliation [48]. Signal intensities were normalized to the pooled standard to account for inter-membrane variability, and subsequently to the total protein amount of each sample to normalize to the different amounts of protein entailed by the different collection volumes.

Statistical analysis

GraphPad Prism (Version 10) and RStudio (version 2024.12.0.467 [49]; R version 4.4.2 (2024-10-31), R Core Team [50]) were used for statistical analysis. The significance level for all analyses was set to $\alpha=0.05$ unless stated otherwise. Regarding the clinical and demographic cohort information, quantitative data are shown as median and interquartile range (IQR: Q1, Q3). Categorical data are presented as absolute and relative frequencies. Quantitative data were tested for normality with the D'Agostino-Pearson, and Shapiro-Wilk tests. The Mann-Whitney test was used for non-parametric data. Unpaired Student's t-test was used for parametric data. Fisher's exact test was used for categorical data, such as sex, gene mutations, stratum of onset, progression rate, or ophthalmological characteristics within the cohorts. For inter-cohort comparisons between the discovery and validation cohorts, Kruskal-Wallis testing was performed to compare multiple quantitative data. The Chi-Square test was used for the comparison of qualitative data. Concerning the Western blot validation data, $\log_{10}$ -transformed protein abundance data are represented as median and IQR (Q1, Q3). Missing values were imputed using mindet imputation (CRYM ALS $n=5$, control $n=4$, CAPZA2 ALS $n=2$, control $n=1$, ALDH16A1 ALS $n=2$, control $n=1$, PFKL ALS $n=2$, control $n=3$, SERPINC1 ALS $n=2$, control $n=0$, HP no missing). The differences in protein abundance levels between ALS and control subjects were calculated using the Student's t-test after testing for normality and Bonferroni for multiple testing correction.

In-silico validation of the western blot results

Based on our Western blot validation, we employed machine learning to evaluate the performance of individual proteins as a signature for differentiating ALS from controls. We trained linear regression (lasso), SVM with a linear kernel, and random forest using the same approach as previously described for biomarker identification. The normalized Western blot abundance data with mindet imputation served as input. Embedded feature selection within the lasso model was employed to systematically identify the most important proteins for prediction. AUROCs were quantified to assess protein performances across different applied models.

Results

Cohorts and data description

This study included a discovery and validation cohort (Table 1). In the discovery cohort, we characterized the TF-proteome of 49 subjects with ALS and 54 control subjects without evidence of neurodegeneration (Fig. 1a), matched for age and Schirmer strip wetting length (WL). The ALS group comprised 34 spinal- and 15 bulbar-onset cases with varying progression rates (10 fast, 12 intermediate, 26 slow, 1 not determined). Median disease duration was 20 months (interquartile range (IQR): 9, 37 months), and median ALSFRS-R at TF collection was 35 months (IQR: 29, 41 months).

The independent validation cohort consisted of TF samples from 51 ALS patients and 52 control subjects, with no significant differences in age and WL between groups. To address the male overrepresentation in the ALS group of the discovery cohort ($p=0.003$, Fisher's exact test), the validation cohort was balanced for sex. ALS patients included 36 spinal- and 15 bulbar-onset cases with varying progression rates (7 fast, 22 intermediate, 18 slow, 3 not determined). Median disease duration was 17 months (IQR: 12, 20 months), and median ALSFRS-R at TF collection was 37 months (IQR: 31, 42 months). Apart from sex distribution, no significant differences in the demographic and clinical features were observed between cohorts.

Individual proteomic profiles of ALS patients and control subjects

DIA-MS identified 2,205 proteins across the TF of all ALS and control subjects ($FDR \leq 0.01$). Of these, 876 proteins were consistently detected and quantified in at least 2/3 ($\geq 66\%$) of the samples per group and were used for further analysis. Unsupervised hierarchical clustering and uniform manifold approximation and projection (UMAP) did not reveal a distinct separation between ALS and controls (Figure S3; S4) or stratification by age or sex (Figure S4). In ALS patients, clustering showed no associations with demographics or clinical variables,

including age, sex, genetic status, stratum of onset, disease progression rate, FVC, or pNfH levels (Fig. 1b; Figure S5). This suggests that molecular diversity in this cohort of ALS patients may be independent of demographic and clinical parameters.

Functional enrichment analysis highlights immunological pathways in ALS

Differential abundance analysis using Perseus identified 106 significantly dysregulated proteins in ALS: 94 proteins (89%) were less abundant, whereas 12 proteins (11%) were more abundant in ALS compared to controls (Fig. 1c). Notably, we identified proteins shown to be differentially regulated in ALS before [9, 21, 51–54], e.g., cystatin C (CST3: $-\log_{10}$ FDR=2.11, $\log_2$ fold change (FC) = -0.41, FC = -1.33), calmodulin (CALM: $-\log_{10}$ FDR=5.08, $\log_2$ FC = -0.46, FC = -1.38), alpha-1-antichymotrypsin (SERPINA3: $-\log_{10}$ FDR=2.50, $\log_2$ FC = -0.34, FC = -1.27) and clusterin (CLU: $-\log_{10}$ FDR=2.10, $\log_2$ FC = -0.38, FC=1.30) were significantly lower abundant, while gelsolin isoform 2 (GSN: $-\log_{10}$ FDR=2.30, $\log_2$ FC=0.31, FC=1.24) was more abundant in ALS TF compared to controls (Data S1).

STRING analysis was used to investigate interactions among differentially abundant proteins in ALS. Protein-protein interaction (PPI) network analysis of proteins with reduced abundance in ALS revealed multiple hub proteins, including haptoglobin (HP), CLU, SERPINA3, CST3, and antithrombin-III (SERPINC1). The number of interactions was significantly higher than expected ($p < 1.0 \times 10^{-16}$), indicating strong biological relation (Fig. 1d). Subnetworks enrichment analysis revealed the GO biological processes (BP) “regulation of proteolysis” (FDR=0.00083) and “defense response” (FDR=5.83e-05; Fig. 1d). No significant PPI were found for proteins with increased abundance in ALS ($p = 0.295$; Figure S6).

Gene set enrichment analysis (GSEA) was performed to systematically investigate biological processes (BP), cellular components (CC), and molecular functions (MF) that differed between ALS and control TF protein compositions. GO terms were annotated for all expressed proteins. Of the 37 pathways analysed, 29 (78%) were significantly downregulated in ALS, primarily related to immune system functions (Fig. 1e; Data S2). Multiple pathways related to the same generic category were manually summarized (Fig. 1e, Data S2).

Among the BP significantly altered in ALS, 17 out of 18 pathways linked to the “immune system” were downregulated. The most significantly downregulated pathways included immune response (Normalized enrichment score (NES) = -1.95, $p_{\text{adj}} = 0.0005$) and adaptive immune response (NES = -1.67, $p_{\text{adj}} = 3.77 \times 10^{-6}$). Notably, four pathways (23.5%) were associated with B cell-mediated

processes, and five pathways (29.4%) were related to humoral immune response.

Focusing on CC, we observed a significant enrichment of the “cytoskeleton” category in ALS, particularly pathways associated with intermediate filaments (intermediate filaments NES=2.25, $p_{\text{adj}} = 0.0001$; intermediate filament cytoskeleton NES=2.09, $p_{\text{adj}} = 0.0002$) and keratin filaments (NES=2.08, $p_{\text{adj}} = 0.0001$).

Regarding MF, pathways related to “binding” were downregulated, highlighting a dysregulation of immune processes such as antigen binding (NES = -2, $p_{\text{adj}} = 8.34 \times 10^{-6}$) and immunoglobulin receptor binding (NES = -1.86, $p_{\text{adj}} = 0.012$) pathways. Conversely, the “structural constituent of cytoskeleton” category exhibited significant enrichment.

The most abundant protein family in ALS was immunoglobulins, which were represented in four of the ten generic categories (40%): “immune system”, “cell surface”, “immunoglobulins”, and “binding”. Indeed, the majority of proteins identified in the GSEA belonged to the immunoglobulin family. Among the other proteins associated with reduced immune system pathways, we identified several protein superfamilies, including complement factors (C3, C4BPA, C9 and CD55), serine protease inhibitors (SERPING1 and SERPINA3), fibrinogens (FGA and FGB), and apolipoproteins (APOB and APOE), but also proteins like haptoglobin (HP) and angiogenin (ANG).

Machine learning distinguishes ALS from controls based on the TF proteome

Using the DIA-MS data, we aimed to identify a protein signature distinguishing ALS from controls. Four supervised machine learning algorithms were employed: linear regression, random forest, and support vector machine (SVM) with linear kernel or radial kernel, trained on the abundance data of the previously identified 876 TF proteins. Receiver operating characteristic (ROC) curves were generated in order to evaluate the performance via area under the ROC (AUROC). The SVM with linear kernel performed best, with an AUROC of 0.77 (± 0.09), sensitivity of 0.81 (± 0.11) and specificity of 0.74 (± 0.10), outperforming the other models (Fig. 2a, Table S1).

Despite differences in AUROC performance, the averaged feature importance assigned by the SVM (the highest-performing model) and linear regression (the lowest-performing model) for disease classification were strongly correlated, suggesting that both relied on a shared set of discriminative proteins to differentiate disease status (Fig. 2b). Although the SVM achieved slightly higher performance, the difference was negligible. Based on this concordance and the interpretability and embedded feature selection of the Lasso regression model, we prioritized its results for the identification of a protein signature.

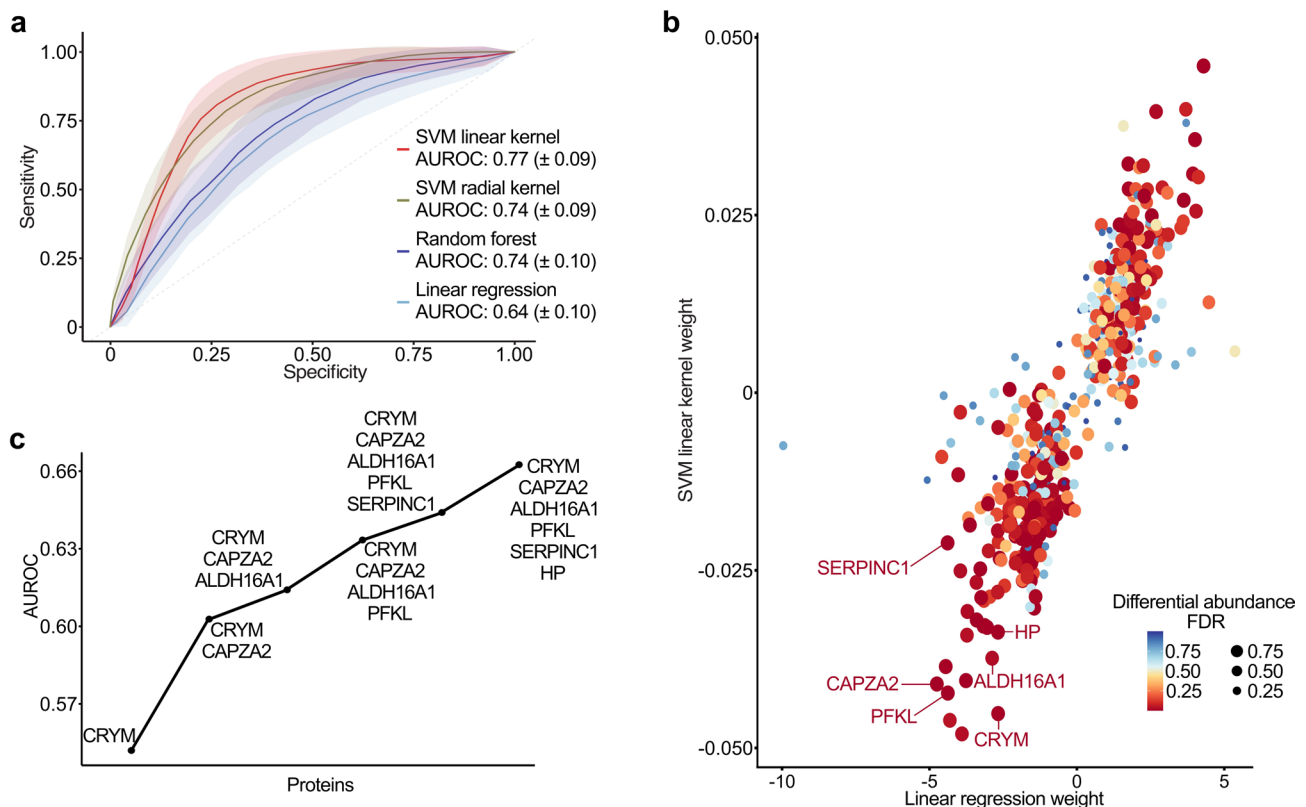


Fig. 2 Leveraging machine learning to identify a six-protein signature for optimal discrimination between ALS and controls. **a** ROC curves for four machine learning algorithms (SVM with linear and radial kernels, random forest, and linear regression) to differentiate ALS from controls. The x-axis represents specificity, the y-axis sensitivity, and performance is quantified by AUROC, with mean $\pm$ SD and the 95% confidence interval shaded. **b** Protein importance for ALS/control classification, as determined by SVM linear kernel and linear regression, alongside differential abundance. The x- and y-axes show averaged protein weights from 500 bootstrap runs of the linear regression and SVM models, respectively. Point color and size correspond to the FDR of the differential abundance analysis. **c** AUROC curve for the six-protein signature optimized for ALS/control discrimination. The x-axis represents proteins selected by the linear regression model, and the y-axis shows the corresponding AUROC values. **Abbreviations:** SVM, support vector machine; AUROC, area under the receiver operating characteristics curve; FDR, false discovery rate; CRYM, mu-crystallin; CAPZA2, f-actin-capping protein subunit alpha-2; ALDH16A1, aldehyde dehydrogenase family 16 member A1; PFKL, phosphofructokinase, liver type; SERPINC1, antithrombin-III; HP, haptoglobin

A six-protein signature in TF differentiates ALS from controls

To identify a more concise subset of biomarkers with high discriminatory diagnostic power, we extracted the averaged protein weights of the previously mentioned linear regression model, attributed to each protein across all 500 bootstrap runs. The models used a total of 83 proteins for the classification of ALS or control subjects. Among these, 24 were significantly differential abundant ($FDR < 0.1$), making them suitable for further validation as potential biomarkers (Data S3).

To identify a biomarker signature that could be suitable for clinical routine, we evaluated the performance of several linear regression models based on combinations of the previously selected 24 proteins. The optimal performance with an AUROC of 66.2% was achieved with the following six proteins: mu-crystallin (CRYM), phosphofructokinase, liver type (PFKL), f-actin-capping protein subunit alpha-2 (CAPZA2), aldehyde dehydrogenase family 16 member A1 (ALDH16A1), SERPINC1, and HP

(Fig. 2c, detailed identification of the six-protein signature is provided in Supplementary Methods). All these six proteins were found to be significantly less abundant in the TF of ALS patients (Fig. 1c) and exhibited high absolute weights distinguishing ALS from controls (Fig. 2b).

Validation of HP and SERPINC1 in an independent cohort

To validate the hypothesis that the protein signature comprising CRYM, PFKL, CAPZA, ALDH16A1, SERPINC1, and HP can distinguish ALS from control subjects, we quantified these six proteins using Western blots in an independent validation cohort (Fig. 3a). All six proteins were detectable in TF of ALS and controls (Fig. 3b-g). $\log_{10}$ -transformed normalized abundance of SERPINC1 (ALS: -0.12, IQR: -0.48, 0.13; controls: 0.08, IQR: -0.18, 0.29; $p_{\text{adj}} = 0.037$) and HP (ALS: 0.04, IQR: -0.21, 0.66; controls: 0.54, IQR: 0, 0.91; $p_{\text{adj}} = 0.050$) was significantly lower in ALS compared to controls. A directional difference was observed for CRYM, with lower values in ALS

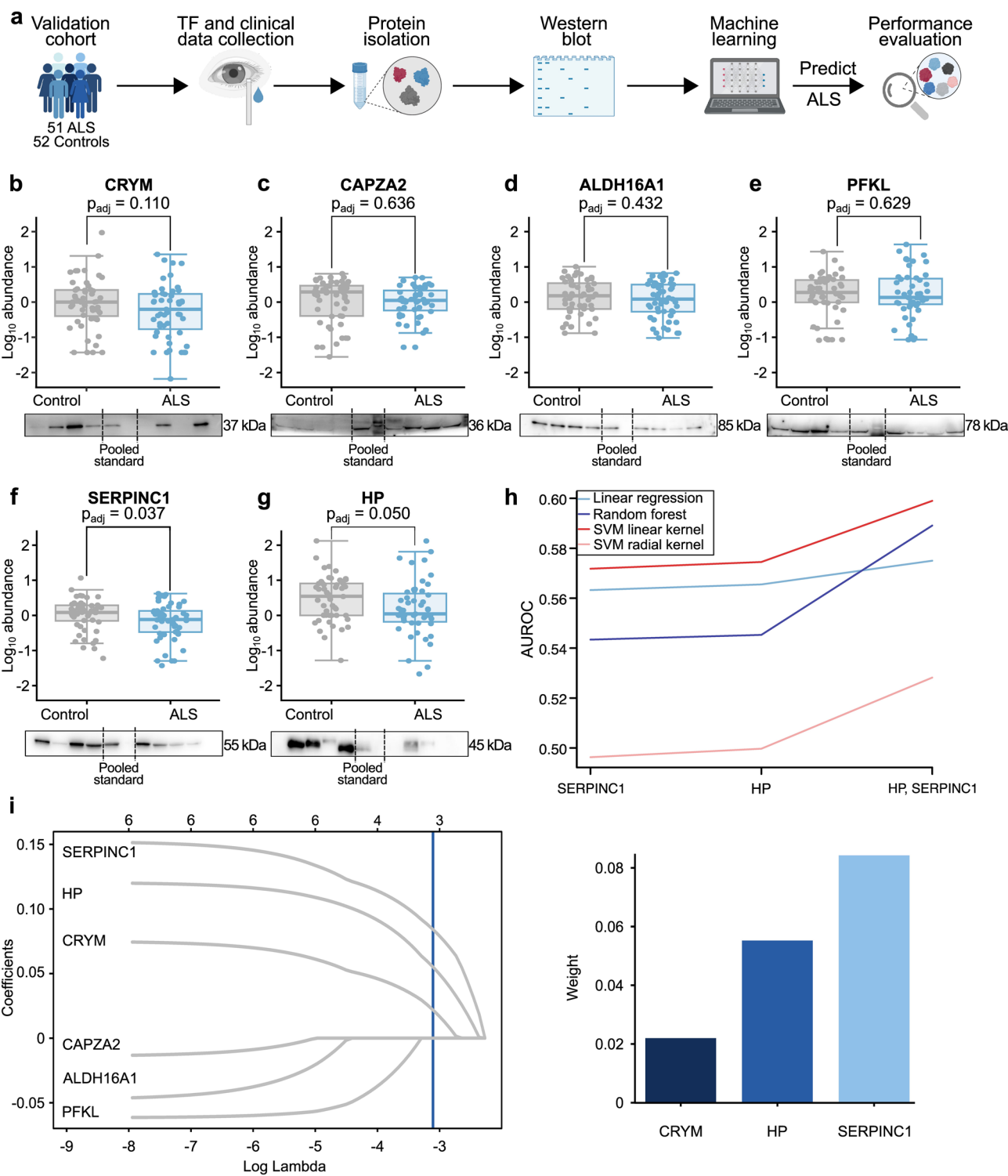


Fig. 3 (See legend on next page.)

(See figure on previous page.)

Fig. 3 Validation of the identified six-protein signature in TF. **a** Experimental workflow for validating the identified TF protein signature. TF and clinical data were collected from an independent cohort of ALS ($n=51$) and control ($n=52$) subjects. TF proteins were analyzed by Western blot, followed by machine learning models to predict ALS/control status and evaluate the diagnostic potential of the proteins. Created with BioRender.com. **b–g** Western blot quantification of CRYM, CAPZA2, ALDH16A1, PFKL, SERPINC1, and HP. Boxplots show $\log_{10}$ -transformed and normalized protein abundance, standardized to a pooled TF sample and the total protein amount of the individual sample. The x-axis represents ALS ($n=51$) and control ($n=52$) groups, and the y-axis displays $\log_{10}$ -transformed normalized abundance. Statistical significance is estimated with Student's t-test and Bonferroni correction. Blots of controls ($n=4$), pooled TF standard ($n=1$), and ALS ($n=4$) are displayed. Full blots in Figure S1d–i and Figure S2. **h** Performance of four machine learning algorithms — SVM (linear+radial kernel), random forest, and linear regression — utilizing Western blot data (SERPINC1, HP and HP and SERPINC1 together) to discriminate ALS from controls. The x-axis represents proteins selected from linear regression models based on the highest weights, while the y-axis shows the corresponding AUROC values, reflecting discriminatory performance. **i** Parameterization of L1 regularization for lasso in one model from the 10-fold cross-validation. The x-axis displays log-transformed lambda values, and the y-axis shows model coefficients. The best-performing lambda is indicated (blue line). **j** The selected lasso model from **i** identifies SERPINC1, HP, and CRYM as key proteins distinguishing ALS from controls. The bar chart shows the weights (y-axis) assigned to each protein (x-axis) in the chosen lasso model. **Abbreviations:** ALS, Amyotrophic Lateral Sclerosis; TF, tear fluid; CRYM, mu-crystallin; CAPZA2, f-actin-capping protein subunit alpha-2; ALDH16A1, aldehyde dehydrogenase family 16 member A1; PFKL, phosphofructokinase, liver type; SERPINC1, antithrombin-III; HP, haptoglobin; p_{adj} , adjusted p -value; SVM, support vector machine; AUROC, area under the receiver operating characteristics curve

patients (ALS: -0.20, IQR: -0.86, 0.24; controls: 0 IQR: -0.45, 0.35; $p_{adj} = 0.110$). Although ALDH16A1 (ALS: 0.08, IQR: -0.27, 0.50; controls: 0.18, IQR: -0.20, 0.56; $p_{adj} = 0.432$), CAPZA2 (ALS: 0.05, IQR: -0.26, 0.38; controls: 0.29, IQR: -0.40, 0.47; $p_{adj} = 0.636$), and PFKL (ALS: 0.13, IQR: -0.08, 0.68; controls: 0.27, IQR: -0.01, 0.64; $p_{adj} = 0.629$) showed numerically lower abundance in ALS, these differences were not statistically significant.

To confirm the performance of SERPINC1 and HP, we trained four models, SVM linear and radial kernel, lasso regression, and random forest, with data including SERPINC1 alone, HP alone, and HP and SERPINC1 combined (Fig. 3h). The three models all showed improved AUROC scores (Table S2) when data included both SERPINC1 and HP (SVM linear kernel: 0.60, SVM radial kernel: 0.53, lasso regression: 0.58, random forest: 0.59) than SERPINC alone (SVM linear kernel: 0.57, SVM radial kernel: 0.50, lasso regression: 0.56, random forest: 0.54) or HP alone (SVM linear kernel: 0.57, SVM radial kernel: 0.50, lasso regression: 0.57, random forest: 0.55).

To evaluate the contribution of all six proteins in an embedded feature selection process, we utilised the lasso regression model with L1 regularization. This approach allows for automatic feature selection while evaluating the importance of each protein in distinguishing ALS patients from controls. SERPINC1 emerged as the most consistently selected protein, followed by HP (Fig. 3i, j; Figure S7). Additionally, CRYM was identified in half of the 10 nested 10-fold cross-validation iterations, while PFKL was selected in three iterations, demonstrating the efficacy of lasso regression in identifying key biomarkers.

Discussion

Improved biomarkers are urgently needed to enhance the accuracy of ALS diagnosis and monitor disease progression. In this context, we investigated the potential of TE, a readily accessible biofluid, as a source of protein biomarkers for ALS. In a discovery approach, we characterized individual TF proteomes in ALS patients and control

subjects without indication of neurodegenerative disease. We identified multiple proteins and pathways with differential regulation between the groups. Six proteins were subsequently evaluated in an independent validation cohort to assess their potential as individual and combined biomarkers for ALS.

Importantly, we did not detect any differences in tear fluid production, in the presence of ocular diseases or medication between the different groups of our cohort, suggesting these did not significantly influence the results in our study. It is important to acknowledge, however, that ocular conditions (e.g. dry eye disease) or environmental factors (e.g. humidity) influence tear fluid production [55–57] and contents such as proteins or miRNA [58, 59].

In our discovery cohort, the higher number of males in the ALS group (Table 1) reflects the higher prevalence of ALS in males [60]. We did not find other statistical differences regarding demographic or clinical features between the groups. Phenotypic heterogeneity poses a considerable challenge in understanding ALS [61]. Clinical phenotypes (e.g., site of onset, propagation pattern, progression rate, or cognitive impairment), but also sex, age, and genetic status can have an impact on molecular composition in postmortem brain tissue or biofluids such as blood and CSF [62–64]. In our study, clustering based on the whole proteome did not allow for separation between the ALS and control cohorts. No molecular subgroup within the entire cohort or within ALS patients only could be attributed to any demographic or clinical parameters.

For sample preparation of our mass spectrometry dataset, we employed the single-pot solid-phase-enhanced sample preparation (SP3) approach. This method involves protein capture on paramagnetic beads, subsequent washing steps, and finally an on-bead tryptic digestion. The SP3 approach was selected as it combines efficient removal of a wide range of buffers and contaminants, is suitable to enrich low microgram amounts of proteins,

and is readily automated and parallelized using magnetic bead handlers for reproducible analysis of medium-to-large sample cohorts [36]. Proteome profiling of digested samples was achieved using data-independent acquisition mass spectrometry (DIA-MS) on a hybrid ion mobility/quadrupole/time-of-flight mass spectrometer. DIA-MS is a widely used approach for the quantitative analysis of proteomes in body fluids, offering high sensitivity with consistent quantitation of peptides in complex matrices [39]. To improve protein detection, we generated and applied a hybrid spectral library containing both DDA (data-dependent acquisition) and DIA mass spectrometry of a pooled and fractionated reference sample [38, 39].

In our protein dataset, we identified a total of 2205 and 876 proteins before and after consistency filtering, respectively, which aligns with the reported range of other TF proteomics studies [65–68]. This relatively high degree of missingness can be readily explained by variations during sampling procedures, such as different Schirmer strip running lengths. Additionally, both inter-individual and diurnal variations in tear fluid composition may contribute, as total protein concentrations have been reported to range between 1 and 11 mg/ml [32, 69]. Nevertheless, approximately 1,000 proteins were consistently detected in at least 60% of the samples, with the majority exhibiting low variability. This included all proteins of our proposed biomarker signature, supporting their suitability for downstream analyses (Figure S8). In accordance to Dor et al. [68], we detected lactotransferrin (LTF), lipocalin-1 (LCN1), serum albumin (ALB), lysozyme C (LYZ), and immunoglobulin heavy constant alpha (IGHA) as the five most abundant TF proteins in our control subjects, as well as in the ALS patients (Figure S9, Data S4). Immunoglobulins, serpins, and 14-3-3 domain proteins were also present as known principal tear protein superfamilies [68].

Differential abundance analysis revealed significant differences in 106 proteins between ALS and control subjects, with 89% of them being less abundant in ALS patients. Notably, several of these significantly deregulated proteins (CST3, CLU, GSN, CALM, and SERPINA3) have been previously studied as potential ALS markers in post-mortem spinal cord [70] or biofluids, such as CSF or plasma [9, 21, 51–54]. The downregulation of immunoglobulins and immune system processes in the TF of ALS patients is one of the key findings of this analysis. Similarly, in another NDD, dysregulation of networks of proteins related to immune dysfunction was also observed in the TF of Parkinson's disease patients [31, 71]. Dysregulation of the immune system has been recognized as a potential driver of disease mechanisms in ALS [72]. Neuroinflammatory proteins, for example CHIT1, serve as non-disease-specific markers for diagnosis and

prognosis [7]. Interestingly, we could also detect chitinase family members (CHI3L1, CHI3L2) in the TF of ALS and control subjects. However, due to limited consistency in all samples (33% missing), CHI3L1 was removed during the filtering step. A suppression of immune system processes in ALS patients compared to controls has already been demonstrated in postmortem brain tissue [62]. In contrast, other studies have shown upregulation of proteins associated with immune response in ALS [73, 74]. Interestingly, immune response pathways seem to contribute to a molecular clustering of ALS patients [62, 64, 75]. Whether the inflammatory changes observed in TF primarily reflect systemic immune alterations or local ocular processes is challenging to determine. However, it is likely that both contribute to the observed protein profiles. Importantly, the prevalence of (inflammatory) eye disease did not differ between groups.

We then took advantage of machine learning models to identify a protein signature in the TF of ALS patients. In a similar approach, Bereman et al. identified proteomic biomarkers for ALS classification in CSF and plasma [54]. In our study, the best classification performance was achieved using an SVM with a linear kernel, which yielded an AUROC of 77% based on all 876 proteins. This model effectively differentiated ALS patients from controls, achieving high sensitivity (81%) and specificity (74%). The strong performance of the linear SVM model is consistent with its suitability for high-dimensional datasets (876 proteins) with limited sample sizes (approximately 100 patients), as commonly encountered in proteomics studies. To facilitate the translation of a diagnostic method into clinical routine, the signature combinations were optimized to a minimal set of six proteins, however achieving the maximal diagnostic performance. We identified a signature of six proteins with an AUROC of 66%, comprising CRYM, PFKL, CAPZA2, ALDH16A1, SERPINC1, and HP. In the discovery cohort, all of them were significantly less abundant in the TF of ALS patients compared to controls. The significantly lower levels of SERPINC1 and HP in ALS patients were validated with Western blot in an independent cohort. While the other proteins failed to reach significance, the direction of regulation was consistent with the data from the discovery cohort as well. Machine learning analysis of the Western blot validation findings resulted in SERPINC1 and HP jointly as the most reliable discriminators, having the highest impact in classifying ALS versus controls, yielding an AUROC of 60%.

SERPINC1 belongs to the serine protease inhibitor (serpin) family and is a crucial inactivator of several pro-coagulant targets in the blood coagulation cascade. Additionally, it has anti-inflammatory effects that can occur through both coagulation-dependent and coagulation-independent mechanisms [76]. In ALS patients' blood

compared to controls, significantly lower serum levels of SERPINC1 have been observed [77]. Another study identified it among the top five proteins reduced in the plasma of ALS patients [54]. In the CSF of ALS patients, SERPINC1 was identified as having a strong impact on the differentiation between fast and slow disease progression, with high SERPINC1 expression levels being associated with faster progression [64].

As a plasma protein, HP binds free hemoglobin (HB) and therefore prevents HB-induced oxidation. As an acute-phase protein, HP also regulates immune responses and exhibits antibacterial activity by limiting iron and heme utilization through microorganisms [78]. The implications of HP in ALS remain largely unexplored. Elevated HP levels have been observed in the serum of ALS patients compared to controls [79, 80]. Interestingly, in line with our findings in TF, HP levels in the CSF of ALS patients were found to be decreased compared to controls [22]. Differential abundance of HP has also been demonstrated in biofluids across multiple NDD, including Alzheimer's disease, Huntington's disease, Parkinson's disease, and frontotemporal dementia [81–85]. Oxidative stress and neuroinflammation are recognized as pivotal hallmarks of NDD pathogenesis. Our findings underscore that the role of HP needs further investigation in ALS, given its involvement in neuroinflammation and various NDD.

CRYM is involved in thyroid hormone metabolism and serves as ketimine reductase in lysine degradation in the brain in the pipecolate pathway [86]. Its protein expression is indeed predominantly observed in the brain, with the highest levels detected in the basal ganglia and cerebral cortex [87, 88], followed by the kidney, prostate, and skin [89]. CRYM has been shown to be specifically expressed in some mice corticospinal motor neurons (CMNS) and subcerebral neurons of layer V in comparison to callosal and corticotectal projection neurons. Its expression, observed during intermediate stages of CMNS development in mice, suggests a potential role in the early specification and differentiation of these neuronal subtypes [90]. Interestingly, the screening of sporadic and familial ALS patients with a multi-gene panel targeting those genes associated with CMNS development identified a missense variant in a highly conserved region of the CRYM gene in cases of sALS [91]. Hommyo et al. used immunohistochemistry on post-mortem brain and spinal cord tissue samples of ALS patients, finding a progressive cranial-to-caudal decline in CRYM expression. They noted a transition from atrophic large pyramidal cells with both CRYM-positive and -negative neurons to a complete loss of CRYM expression in the distal pyramidal tracts. The authors postulated a “dying back” hypothesis with neurodegeneration starting distally, considering

CRYM as a surrogate reflecting axonal degeneration in the corticospinal tract [92].

While the other proteins of the signature did not reach significance in our validation studies, they have all been implicated in neurodegenerative diseases. CAPZA2 is a subunit of an F-actin-capping complex and thus involved in the regulation of elongation of cytoskeletal filaments, playing a crucial role in a range of dynamic cellular processes [93]. Mutations in CAPZA2 can cause intellectual disability and developmental delay [94]. While the role of CAPZA2 in neurodegeneration is not well understood, it has been shown to be down-regulated in CSF of patients with Alzheimer's disease [95]. Its transcript was shown to be up-regulated in the whole blood of ALS patients compared to controls [96]. ALDH16A1 belongs to the Aldehyde dehydrogenase superfamily. Aldehyde dehydrogenases constitute a key enzyme family involved in the detoxification of reactive aldehydes arising from endogenous metabolic processes and exogenous environmental sources, thereby mitigating cellular oxidative and electrophilic stress [97]. In Mast syndrome (SPG21), ALDH16A1 has been associated with masperdin, the protein in which causal mutations have been identified. Loss of ALDH16A1 function was hypothesized to be associated with a particular vulnerability of the upper motor neurons in this setting [98]. Data on the role of ALDH16A1 in ALS is, however, lacking. PFKL is a crucial enzyme in glycolysis. Glucose metabolism defects have been implicated in the pathogenesis of ALS: altered glucose metabolism has been shown in brain tissue of ALS patients as well as the SOD1 mouse model [99, 100]. PFKL is the liver-specific isoform of phosphofructokinase. While systemic glucose metabolism changes have been identified in patients with ALS [101], data on PFKL expression in ALS is lacking. The other isoforms (PFKM in muscle and PFKP in platelets) have been shown to be up-regulated at the transcript level in the spinal cord and induced pluripotent stem cell-derived motor neurons of ALS patients [102].

The initial findings and validation of our protein signature's ability to classify ALS and control subjects are encouraging, though the diagnostic performance in a clinical setting is limited. Larger sample sizes and potentially additional proteins are likely necessary to develop more robust predictive models. We analyzed two independent cohorts from a single center, which limits the generalizability of our findings to other populations. Additionally, our discovery cohort was male-skewed, introducing a potential bias. To mitigate this, we modeled sex as covariate in the proteomics analysis. Importantly, the robustness of our findings is supported by the positive results observed in the independent validation cohort. In our study, we exclusively sampled tears with Schirmer test strips. While tears can also be collected

using microcapillary tubes [103], both techniques are highly applicable for protein analyses [104]. A possible variability in protein composition should be considered [105, 106]. In TF, four proteins - LCN1, LYZ, LTF, and secretory immunoglobulin A - are known to account for 70–85% of total protein concentration [107, 108], thus potentially masking low abundant proteins that could have importance in ALS pathogenesis. This might explain why our proteomic analysis could not detect common biomarkers like neurofilaments in TF, although NfL has been described in the TF of Parkinson's disease patients using the ultrasensitive Single Molecule Array (SIMOA) assay [28]. However, our data revealed an upregulation of cytoskeleton-associated proteins and pathways in the ALS group. The lack of significance of the other signature proteins in our Western blot validation approach might also be attributed to the biological variability within samples, compounded by technical limitations such as limited sample quantities. This might have reduced the robustness and sensitivity of the analyses. Studies in larger cohorts as well as using ultra-sensitive assays such as SIMOA or Nucleic acid Linked Immuno-Sandwich Assay (NULISA) may be needed to detect low abundant proteins reliably.

Conclusions

In summary, TF is a well-accessible and non-invasively obtainable body fluid with substantial promise for biomarker discovery. In our study, we found prominent signals in proteins that have previously been implicated in the pathogenesis of ALS, particularly those related to neuroinflammation. We identified and experimentally validated six differentially abundant proteins - CRYM, PFKL, CAPZA2, ALDH16A1, SERPINC1, and HP. Notably, the combination of SERPINC1, HP, and CRYM demonstrated encouraging performance in disease classification. Future studies on larger and more diverse cohorts, including longitudinal sampling, will inform about the generalizability and value of these proteins as biomarkers for disease state and progression in a clinical setting.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40478-025-02109-6>.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

Supplementary Material 4

Supplementary Material 5

Acknowledgements

The authors thank all patients, their families, and caregivers for their participation in this study. We acknowledge and thank all contributors involved in the study design and in patient recruitment, sample collection, and data processing. A special thanks to Petra Rau and Erica Westenberg for their valuable support in TF collection and the doctors of the motor neuron disease outpatient clinic of TUM University Hospital Munich for their assistance in patient recruitment. We also appreciate the constructive feedback from all members of the Lingor and Menden group, and reviewers, which helped to improve the manuscript. CL acknowledges the Core Facility of Proteomics at the University Medical Center Goettingen (UMG), where mass-spectrometry-based proteome analyses were performed.

Author contributions

PL: conceptualization, formal analysis, funding acquisition, methodology, project administration, resources, supervision, writing - review & editing. LSS: data curation, formal analysis, investigation, methodology, project administration, validation, visualization, writing - original draft. AFD: conceptualization, data curation, formal analysis, methodology, project administration, supervision, visualization, writing - original draft, writing - review & editing. MM: conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, resources, software, supervision, visualization, writing - original draft, writing - review & editing. JR, SA, CM, AG: data curation, formal analysis, investigation, methodology, software, visualization, validation, writing - original draft, writing-review & editing. CL: conceptualization, data curation, investigation, methodology, resources, software, supervision, visualization, writing - original draft, writing - review & editing. LCG: data curation, formal analysis, writing - original draft, writing - review & editing. LT: supervision, visualization, writing - original draft, writing - review & editing. LN: data curation, investigation, methodology, writing - review & editing. IC: Data curation, investigation, writing - review & editing. LSS and AFD contributed equally. PL, MM, and CL contributed equally. All authors read and approved the final manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL. Mass spectrometry instrumentation was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - ID 442069358. PL is supported by the DFG (German Research Foundation) under Germany's Excellence Strategy within the framework of the Munich Cluster for Systems Neurology (EXC 2145, SyNergy - ID 390857198). MM is supported by the European Union's Horizon 2020 Research and Innovation Programme (Grant agreement No. 950293 - COMBAT-RES). PL, MM, LT, and AG are funded by the Bundesministerium für Bildung und Forschung (premodiALS, 01ED2204A, 01ED2204B).

Data availability

The source code for all computational analyses is available for download on GitHub (<https://github.com/MendenLab/Tear-Fluid-Proteomics>). Proteomic data were deposited to the ProteomeXchange Consortium via the PRIDE partner repository [109] with the dataset identifier PXD062971. Deidentified clinical data from this study are provided with a signed data access agreement upon request from the corresponding authors.

Declarations

Ethics approval and consent to participate

Permission was obtained from the Ethics Committee of the Technical University of Munich, School of Medicine (approval number: 9/155). Written informed consent was obtained from all patients prior to participation.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 23 April 2025 / Accepted: 25 August 2025

Published online: 02 September 2025

References

- Hardiman O, Al-Chalabi A, Chio A, Corr EM, Logroscino G, Robberecht W et al (2017) Amyotrophic lateral sclerosis. *Nat Rev Dis Primers* 3:17071
- Brooks BR, Miller RG, Swash M, Munsat TL (2000) El Escorial revisited: revised criteria for the diagnosis of amyotrophic lateral sclerosis. *Amyotroph Lateral Scler Other Motor Neuron Disord* 1(5):293–299
- Shefner JM, Al-Chalabi A, Baker MR, Cui LY, de Carvalho M, Eisen A et al (2020) A proposal for new diagnostic criteria for ALS. *Clin Neurophysiol* 131(8):1975–1978
- Richards D, Morren JA, Pioro EP (2021) Time to Diagnosis and Factors Affecting Diagnostic Delay in Amyotrophic Lateral Sclerosis. *J Neurol Sci* 417:117054
- Masrori P, Van Damme P (2020) Amyotrophic lateral sclerosis: a clinical review. *Eur J Neurol* 27(10):1918–1929
- Steinacker P, Feneberg E, Weishaupt J, Brettschneider J, Tumani H, Andersen PM et al (2016) Neurofilaments in the diagnosis of motoneuron diseases: a prospective study on 455 patients. *J Neurol Neurosurg Psychiatry* 87(1):12–20
- Xu A, Luo Y, Tang Y, Yang F, Gao X, Qiao G et al (2024) Chitinases as a potential diagnostic and prognostic biomarker for amyotrophic lateral sclerosis: a systematic review and meta-analysis. *Neurol Sci* 45(6):2489–2503
- Irwin KE, Sheth U, Wong PC, Gendron TF (2024) Fluid biomarkers for amyotrophic lateral sclerosis: a review. *Mol Neurodegener* 19(1):9
- Zhu Y, Yang M, Li F, Li M, Xu Z, Yang F et al (2018) Aberrant levels of Cystatin C in amyotrophic lateral sclerosis: a systematic review and meta analysis. *Int J Biol Sci* 14(9):1041–1053
- Ceccanti M, Pozzilli V, Cambieri C, Libonati L, Onesti E, Frasca V et al (2020) Creatine kinase and progression rate in amyotrophic lateral sclerosis. *Cells* 9(5):1174
- Agnello L, Colletti T, Lo Sasso B, Vidali M, Spataro R, Gambino CM et al (2021) Tau protein as a diagnostic and prognostic biomarker in amyotrophic lateral sclerosis. *Eur J Neurol* 28(6):1868–1875
- Lunetta C, Lizio A, Gerardi F, Tarlarini C, Filippi M, Riva N et al (2020) Urinary neopterin, a new marker of the neuroinflammatory status in amyotrophic lateral sclerosis. *J Neurol* 267(12):3609–3616
- Shepherd SR, Wu J, Cardoso M, Wiklind L, Dinning PG, Chataway T et al (2017) Urinary p75(ECD): A prognostic, disease progression, and pharmacodynamic biomarker in ALS. *Neurology* 88(12):1137–1143
- Barro C, Chitnis T, Weiner HL (2020) Blood neurofilament light: a critical review of its application to neurologic disease. *Ann Clin Transl Neurol* 7(12):2508–2523
- Verde F, Steinacker P, Weishaupt JH, Kassubek J, Oeckl P, Halbgebauer S et al (2019) Neurofilament light chain in serum for the diagnosis of amyotrophic lateral sclerosis. *J Neurol Neurosurg Psychiatry* 90(2):157–164
- Rossi D, Volanti P, Brambilla L, Colletti T, Spataro R, La Bella V (2018) CSF neurofilament proteins as diagnostic and prognostic biomarkers for amyotrophic lateral sclerosis. *J Neurol* 265(3):510–521
- Benatar M, Macklin EA, Malaspina A, Rogers ML, Hornstein E, Lombardi V et al (2024) Prognostic clinical and biological markers for amyotrophic lateral sclerosis disease progression: validation and implications for clinical trial design and analysis. *EBioMedicine* 108:105323
- Miller TM, Cudkowicz ME, Genge A, Shaw PJ, Sobue G, Bucelli RC et al (2022) Trial of antisense oligonucleotide Tofersen for SOD1 ALS. *N Engl J Med* 387(12):1099–1110
- Di Rosa M, Distefano G, Zorena K, Malaguarnera L (2016) Chitinases and immunity: ancestral molecules with new functions. *Immunobiology* 221(3):399–411
- Thompson AG, Gray E, Bampton A, Raciborska D, Talbot K, Turner MR (2019) CSF chitinase proteins in amyotrophic lateral sclerosis. *J Neurol Neurosurg Psychiatry* 90(11):1215–1220
- Xu Z, Lee A, Nouwens A, Henderson RD, McCombe PA (2018) Mass spectrometry analysis of plasma from amyotrophic lateral sclerosis and control subjects. *Amyotroph Lateral Scler Frontotemporal Degener* 19(5–6):362–376
- Chen Y, Liu XH, Wu JJ, Ren HM, Wang J, Ding ZT et al (2016) Proteomic analysis of cerebrospinal fluid in amyotrophic lateral sclerosis. *Exp Ther Med* 11(6):2095–2106
- Hedl TJ, San Gil R, Cheng F, Rayner SL, Davidson JM, De Luca A et al (2019) Proteomics approaches for biomarker and drug target discovery in ALS and FTD. *Front Neurosci* 13:548
- Börger M, Funke S, Bähr M, Grus F, Lingor P (2015) Biomarker sources for parkinson's disease: time to shed tears? *Basal Ganglia* 5(2):63–69
- Jin K, Imada T, Hisamura R, Ito M, Toriumi H, Tanaka KF et al (2020) Identification of lacrimal gland postganglionic innervation and its regulation of tear secretion. *Am J Pathol* 190(5):1068–1079
- Król-Grzymała A, Sienkiewicz-Szapka E, Fiedorowicz E, Rozmus D, Cieśliska A, Grzybowski A (2022) Tear biomarkers in alzheimer's and parkinson's diseases, and multiple sclerosis: implications for diagnosis (Systematic Review). *Int J Mol Sci* 23(17):10123
- Maass F, Rikker S, Dambeck V, Warth C, Tatenhorst L, Csoti I et al (2020) Increased alpha-synuclein tear fluid levels in patients with parkinson's disease. *Sci Rep* 10(1):8507
- Lin CW, Lai TT, Chen SJ, Lin CH (2022) Elevated α -synuclein and NfL levels in tear fluids and decreased retinal microvascular densities in patients with parkinson's disease. *Geroscience* 44(3):1551–1562
- Gijls M, Ramakers IHGB, Visser PJ, Verhey FRJ, Van De Waarenburg MPH, Schalkwijk CG et al (2021) Association of tear fluid amyloid and Tau levels with disease severity and neurodegeneration. *Sci Rep* 11(1)
- Luib E, Demleitner AF, Cordts I, Westenberg E, Rau P, Pürner D et al (2023) Reduced tear fluid production in neurological diseases: a cohort study in 708 patients. *J Neurol* 217:1824–1836
- Boerger M, Funke S, Leha A, Roser AE, Wuestemann AK, Maass F et al (2019) Proteomic analysis of tear fluid reveals disease-specific patterns in patients with parkinson's disease - A pilot study. *Parkinsonism Relat Disord* 63:3–9
- Gijls M, Arumugam S, Van De Sande N, Webers CAB, Sethu S, Ghosh A et al (2023) Pre-analytical sample handling effects on tear fluid protein levels. *Sci Rep* 13(1)
- Cedarbaum JM, Stambler N, Malta E, Fuller C, Hilt D, Thurmond B et al (1999) The ALSFRS-R: a revised ALS functional rating scale that incorporates assessments of respiratory function. BDNF ALS study group (Phase III). *J Neurol Sci* 169(1–2):13–21
- Requardt MV, Görlich D, Grehl T, Boentert M (2021) Clinical determinants of disease progression in amyotrophic lateral Sclerosis-A retrospective cohort study. *J Clin Med* 10(8)
- Labra J, Menon P, Byth K, Morrison S, Vucic S (2016) Rate of disease progression: a prognostic biomarker in ALS. *J Neurol Neurosurg Psychiatry* 87(6):628–632
- Hughes CS, Moggridge S, Müller T, Sorensen PH, Morin GB, Krijgsveld J (2019) Single-pot, solid-phase-enhanced sample Preparation for proteomics experiments. *Nat Protoc* 14(1):68–85
- Bruderer R, Bernhardt OM, Gandhi T, Reiter L (2016) High-precision iRT prediction in the targeted analysis of data-independent acquisition and its impact on identification and quantitation. *Proteomics* 16(15–16):2246–2256
- Meier F, Brunner AD, Koch S, Koch H, Lubeck M, Krause M et al (2018) Online parallel Accumulation-Serial fragmentation (PASEF) with a novel trapped ion mobility mass spectrometer. *Mol Cell Proteom* 17(12):2534–2545
- Meier F, Brunner AD, Frank M, Ha A, Bludau I, Voytk E et al (2020) DiaPASEF: parallel accumulation-serial fragmentation combined with data-independent acquisition. *Nat Methods* 17(12):1229–1236
- Bruderer R, Bernhardt OM, Gandhi T, Miladinović SM, Cheng LY, Messner S et al (2015) Extending the limits of quantitative proteome profiling with data-independent acquisition and application to acetaminophen-treated three-dimensional liver microtissues. *Mol Cell Proteom* 14(5):1400–1410
- Tyanova S, Cox J, Perseus (2018) A bioinformatics platform for integrative analysis of proteomics data in cancer research. *Methods Mol Biol* 1711:133–148
- Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z et al (2021) ClusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innov (Camb)* 2(3):100141
- Dolgalev msigdb MSigDB Gene Sets for Multiple Organisms in a Tidy Data Format. R package version 7.5.1 2022 [Available from: <<https://CRAN.R-project.org/package=msigdb>>]
- Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA et al (2005) Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences* 102(43):15545–50
- Szklarczyk D, Kirsch R, Koutrouli M, Nastou K, Mehryar F, Hachilif R et al (2023) The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res* 51(D1):D638–d46
- Kuhn M (2008) Building predictive models in R using the caret package. *J Stat Softw.* 28(Issue 5).
- Meyer D, Dimitriadou E, Hornik K, Weingessel A, Leisch F, Chang C, Lin C e1071: Misc Functions of the Department of Statistics, Probability Theory

- Group (Formerly: E1071). R package version 1.7–13. 2023 [Available from: <<https://CRAN.R-project.org/package=e1071>
48. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T et al (2012) Fiji: an open-source platform for biological-image analysis. *Nat Methods* 9(7):676–682
49. Team R, RStudio (2023) Integrated Development for R Boston: RStudio; [Available from: <http://www.rstudio.com/>
50. Team RC (2023) R: A Language and Environment for Statistical Computing Vienna, Austria: R Foundation for Statistical Computing; [Available from: <https://www.R-project.org/>
51. Raghunathan R, Turajane K, Wong LC (2022) Biomarkers in Neurodegenerative Diseases: Proteomics Spotlight on ALS and Parkinson's Disease. *Int J Mol Sci* 23(16)
52. Mohanty L, Henderson RD, McCombe PA, Lee A (2020) Levels of clusterin, CD5L, ficolin-3, and Gelsolin in ALS patients and controls. *Amyotroph Lateral Scler Frontotemporal Degener* 21(7–8):631–634
53. Collins MA, An J, Hood BL, Conrads TP, Bowser RP, Label-Free (2015) LC-MS/MS proteomic analysis of cerebrospinal fluid identifies protein/pathway alterations and candidate biomarkers for amyotrophic lateral sclerosis. *J Proteome Res* 14(11):4486–4501
54. Bereman MS, Beri J, Enders JR, Nash T (2018) Machine learning reveals protein signatures in CSF and plasma fluids of clinical value for ALS. *Sci Rep* 8(1):16334
55. Abusharha AA, Pearce EI, Fagehi R (2016) Effect of Ambient Temperature on the Human Tear Film. *Eye & Contact Lens: Science & Clinical Practice* 42(5):308–312
56. Buckmaster F, Pearce EI (2016) Effects of humidity on tests of tear production. *Cornea* 35(6):754–758
57. Akasaki Y, Inomata T, Sung J, Nakamura M, Kitazawa K, Shih KC et al (2022) Prevalence of comorbidity between dry eye and allergic conjunctivitis: A systematic review and Meta-Analysis. *J Clin Med* 11(13):3643
58. Hsiao Y-T, Huang Y-T, Yu H-J, Fang P-C, Kuo M-T (2022) Tear proteomics approach to distinguishing primary from secondary sjögren's syndrome for dry eye patients with Long-Term instillation of eyedrops. *Int J Mol Sci* 23(23):15239
59. Rassi DM, De Paiva CS, Dias LC, Modulo CM, Adriano L, Fantucci MZ et al (2017) Review: MicroRNAs in ocular surface and dry eye diseases. *Ocul Surf* 15(4):660–669
60. Mehta P, Raymond J, Punjani R, Han M, Larson T, Kaye W et al (2023) Prevalence of amyotrophic lateral sclerosis in the united States using established and novel methodologies, 2017. *Amyotroph Lateral Scler Frontotemporal Degener* 24(1–2):108–116
61. Tzeplaeff L, Jürs AV, Wahnrad C, Demleitner AF (2024) Unraveling the heterogeneity of ALS-A call to redefine patient stratification for better outcomes in clinical trials. *Cells* 13(5)
62. Caldi Gomes L, Hänzelmann S, Hausmann F, Khatri R, Oller S, Parvaz M et al (2024) Multicomic ALS signatures highlight subclusters and sex differences suggesting the MAPK pathway as therapeutic target. *Nat Commun* 15(1):4893
63. Santiago JA, Quinn JP, Potashkin JA (2021) Network analysis identifies Sex-Specific gene expression changes in blood of amyotrophic lateral sclerosis patients. *Int J Mol Sci* 22(13)
64. Vu L, Garcia-Mansfield K, Pompeiano A, An J, David-Dirgo V, Sharma R et al (2023) Proteomics and mathematical modeling of longitudinal CSF differentiates fast versus slow ALS progression. *Ann Clin Transl Neurol* 10(11):2025–2042
65. de Souza GA, Godoy LM, Mann M (2006) Identification of 491 proteins in the tear fluid proteome reveals a large number of proteases and protease inhibitors. *Genome Biol* 7(8):R72
66. Zhou L, Zhao SZ, Koh SK, Chen L, Vaz C, Tanavde V et al (2012) In-depth analysis of the human tear proteome. *J Proteome* 75(13):3877–3885
67. Aass C, Norheim I, Eriksen EF, Thorsby PM, Pepaj M (2015) Single unit filter-aided method for fast proteomic analysis of tear fluid. *Anal Biochem* 480:1–5
68. Dor M, Eperon S, Lalive PH, Guex-Crosier Y, Hamedani M, Salvisberg C et al (2019) Investigation of the global protein content from healthy human tears. *Exp Eye Res* 179:64–74
69. Ng V, Cho P, To C, Ng V, Cho P, To C (2000) Tear proteins of normal young Hong Kong Chinese. *Graefes Archive for Clinical and Experimental Ophthalmology* 238(9)
70. Engelen-Lee J, Blokhuis AM, Spliet WGM, Pasterkamp RJ, Aronica E, Demmers JAA et al (2017) Proteomic profiling of the spinal cord in ALS: decreased ATP5D levels suggest synaptic dysfunction in ALS pathogenesis. *Amyotroph Lateral Scler Frontotemporal Degener* 18(3–4):210–220
71. Acera A, Gómez-Esteban JC, Murueta-Goyena A, Galdos M, Azkargorta M, Elortza F et al (2022) Potential Tear Biomarkers for the Diagnosis of Parkinson's Disease-A Pilot Study. *Proteomes* 10(1)
72. Mimic S, Aru B, Pehlivanoğlu C, Sleiman H, Andjus PR, Yanikkaya Demirel G (2023) Immunology of amyotrophic lateral sclerosis - role of the innate and adaptive immunity. *Front Neurosci* 17:1277399
73. Saris CG, Groen EJ, Koekkoek JA, Veldink JH, van den Berg LH (2013) Meta-analysis of gene expression profiling in amyotrophic lateral sclerosis: a comparison between Transgenic mouse models and human patients. *Amyotroph Lateral Scler Frontotemporal Degener* 14(3):177–189
74. Béland LC, Markovinov A, Jakovac H, De Marchi F, Bilic E, Mazzini L et al (2020) Immunity in amyotrophic lateral sclerosis: blurred lines between excessive inflammation and inefficient immune responses. *Brain Commun* 2(2):fcaa124
75. Eshima J, O'Connor SA, Marschall E, Bowser R, Plaisier CL, Smith BS (2023) Molecular subtypes of ALS are associated with differences in patient prognosis. *Nat Commun* 14(1):95
76. Levy JH, Sniecinski RM, Welsby IJ, Levi M (2016) Antithrombin: anti-inflammatory properties and clinical applications. *Thromb Haemostasis* 115(4):712–728
77. Adachi N, Shoji S (1989) Studies of protease inhibitors in the Sera of patients with amyotrophic lateral sclerosis. *J Neurol Sci* 89(2–3):165–168
78. di Masi A, De Simone G, Ciaccio C, D'Orso S, Coletta M, Ascenzi P (2020) Haptoglobin: from hemoglobin scavenging to human health. *Mol Aspects Med* 73:100851
79. Ikrar T, Steenblock DA, Putranto TA (2019) Atypical levels of heavy metals, creatinine, haptoglobin, and ferritin found in amyotrophic lateral sclerosis outpatients vs. Healthy controls: possible utility as a diagnostic biomarker. *Int J Pathol Clin Res* 5(1)
80. Vassileff N, Leblanc P, Bernard E, Fourier A, Quadro I, Lowe RGT, Spiers JG, Hill AF, Cheng L (2023) Proteomic analysis of human serum Extracellular Vesicles reveals early diagnostic markers for Amyotrophic Lateral Sclerosis. *MedRxiv*
81. Johnson G, Brane D, Block W, van Kammen DP, Gurklis J, Peters JL et al (1992) Cerebrospinal fluid protein variations in common to alzheimer's disease and schizophrenia. *Appl Theor Electrophor* 3(2):47–53
82. Song I-U, Kim Y-D, Chung S-W, Cho H-J (2015) Association between serum haptoglobin and the pathogenesis of alzheimer's disease. *Intern Med* 54(5):453–457
83. Huang YC, Wu YR, Tseng MY, Chen YC, Hsieh SY, Chen CM (2011) Increased prothrombin, Apolipoprotein A-IV, and haptoglobin in the cerebrospinal fluid of patients with huntington's disease. *PLoS ONE* 6(1):e15809
84. Chelliah SS, Bhuvanendran S, Magalingam KB, Kamarudin MNA, Radhakrishnan AK (2022) Identification of blood-based biomarkers for diagnosis and prognosis of parkinson's disease: A systematic review of proteomics studies. *Ageing Res Rev* 73:101514
85. Davidsson P, Sjögren M, Andreasen N, Lindbjör M, Nilsson CL, Westman-Brinkmalm A et al (2002) Studies of the pathophysiological mechanisms in frontotemporal dementia by proteome analysis of CSF proteins. *Brain Res Mol Brain Res* 109(1–2):128–133
86. Hallen A, Cooper AJ, Jamie JF, Karuso P (2015) Insights into enzyme catalysis and thyroid hormone regulation of cerebral ketimine Reductase/μ-Crystallin under physiological conditions. *Neurochem Res* 40(6):1252–1266
87. Sjöstedt E, Zhong W, Fagerberg L, Karlsson M, Mitsios N, Adori C et al (2020) An atlas of the protein-coding genes in the human, pig, and mouse brain. *Science* 367(6482)
88. Kinney CJ, Bloch RJ (2021) μ-Crystallin: A thyroid hormone binding protein. *Endocr Regul* 55(2):89–102
89. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A et al (2015) Proteomics. Tissue-based map of the human proteome. *Science* 347(6220):1260419
90. Arlotta P, Molyneaux BJ, Chen J, Inoue J, Kominami R, Macklis JD (2005) Neuronal subtype-specific genes that control corticospinal motor neuron development in vivo. *Neuron* 45(2):207–221
91. Daoud H, Valdmanis PN, Gros-Louis F, Belzil V, Spiegelman D, Henrion E et al (2011) Resequencing of 29 candidate genes in patients with Familial and sporadic amyotrophic lateral sclerosis. *Arch Neurol* 68(5):587–593
92. Hommyo R, Suzuki SO, Abolhassani N, Hamasaki H, Shijo M, Maeda N et al (2018) Expression of CRYM in different rat organs during development and its decreased expression in degenerating pyramidal tracts in amyotrophic lateral sclerosis. *Neuropathology* 38(3):247–259

93. Wear M, Cooper J (2004) Capping protein: new insights into mechanism and regulation. *Trends Biochem Sci* 29(8):418–428
94. Huang Y, Mao X, Van Jaarsveld RH, Shu L, Terhal PA, Jia Z et al (2020) Variants in CAPZA2, a member of an F-actin capping complex, cause intellectual disability and developmental delay. *Hum Mol Genet* 29(9):1537–1546
95. Liu P, Li L, He F, Meng F, Liu X, Su Y et al (2023) Identification of candidate biomarkers of alzheimer's disease via multiplex cerebrospinal fluid and serum proteomics. *Int J Mol Sci* 24(18):14225
96. Huseby CJ, Delvaux E, Brokaw DL, Coleman PD (2023) Blood RNA transcripts reveal similar and differential alterations in fundamental cellular processes in Alzheimer's disease and other neurodegenerative diseases. *Alzheimer's Dement* 19(6)
97. Lisgara A, Thompson DC, Nebert DW, Vasiliou V (2025) Lessons learned from aldehyde dehydrogenases as non-P450 aldehyde-oxidizing enzymes: implications for the 'exposome' and human health. *Expert Opin Drug Metab Toxicol* 07–04
98. Hanna MC, Blackstone C (2009) Interaction of the SPG21 protein ACP33/ mspardin with the aldehyde dehydrogenase ALDH16A1. *Neurogenetics* 10(3):217–228
99. Ludolph AC, Langen KJ, Regard M, Herzog H, Kemper B, Kuwert T et al (1992) Frontal lobe function in amyotrophic lateral sclerosis: a neuropsychologic and positron emission tomography study. *Acta Neurol Scand* 85(2)
100. McDonald TS, Kumar V, Fung JN, Woodruff TM, Lee JD (2021) Glucose clearance and uptake is increased in the SOD1 G93A mouse model of amyotrophic lateral sclerosis through an insulin-independent mechanism. *FASEB J* 35(7)
101. Pradat P-F, Bruneteau G, Gordon PH, Dupuis L, Bonnefont-Rousselot D, Simon D et al (2010) Impaired glucose tolerance in patients with amyotrophic lateral sclerosis. *Amyotroph Lateral Scler* 11(1–2):166–71
102. Manzo E, Lorenzini I, Barrameda D, O'Conner AG, Barrows JM, Starr A et al (2019) Glycolysis upregulation is neuroprotective as a compensatory mechanism in ALS. *eLife* 8
103. Pieczyński J, Szulc U, Harazna J, Szulc A, Kiewisz J (2021) Tear fluid collection methods: review of current techniques. *Eur J Ophthalmol* 31(5):2245–2251
104. Posa A, Bräuer L, Schicht M, Garreis F, Beileke S, Paulsen F (2013) Schirmer strip vs. capillary tube method: non-invasive methods of obtaining proteins from tear fluid. *Ann Anat* 195(2):137–142
105. Nättinen J, Aapola U, Jylhä A, Vaajanen A, Uusitalo H (2020) Comparison of capillary and schirmer strip tear fluid sampling methods using SWATH-MS proteomics approach. *Transl Vis Sci Technol* 9(3):16
106. Green-Church KB, Nichols KK, Kleinholz NM, Zhang L, Nichols JJ (2008) Investigation of the human tear film proteome using multiple proteomic approaches. *Mol Vis* 14:456–470
107. Azzarolo AM, Brew K, Kota S, Ponomareva O, Schwartz J, Zylberberg C (2004) Presence of tear Lipocalin and other major proteins in lacrimal fluid of rabbits. *Comp Biochem Physiol B Biochem Mol Biol* 138(2):111–117
108. Tiffany JM (2008) The normal tear film. *Dev Ophthalmol* 41:1–20
109. Perez-Riverol Y, Bai J, Bandla C, García-Seisdedos D, Hewapathirana S, Kamatchinathan S et al (2022) The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res* 50(D1):D543–d52

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.