miRNA biomarkers for diagnosis of ALS and FTD, developed by a nonlinear machine learning approach.

Short title: miRNA biomarkers for diagnosis of ALS and FTD

Iddo Magen[1,2, $], Nancy-Sarah Yacovzada[1,2, $], Jason D. Warren[3], Carolin Heller[3,4], Imogen Swift[3,4], Yoana Bobeva[5], Andrea Malaspina[5,6], Jonathan D. Rohrer[3*] Pietro Fratta[6*] and Eran Hornstein[1,2*]

1 Department of Molecular Genetics, Weizmann Institute of Science, Rehovot, Israel

2 Department of Molecular Neuroscience, Weizmann Institute of Science, Rehovot, Israel

3 Dementia Research Centre, and 4 UK Dementia Research Institute, Department of Neurodegenerative Disease, UCL Queen Square Institute of Neurology, London, UK.

5 Centre for Neuroscience and Trauma, Blizard Institute, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, London, UK.

6 Department of Neuromuscular Diseases, UCL Queen Square Institute of Neurology, London, UK

$ Equal contribution

* To whom correspondence should be addressed: a.malaspina@ucl.ac.uk; j.rohrer@ucl.ac.uk; p.fratta@ucl.ac.uk; eran.hornstein@weizmann.ac.il;

Abstract word count: 153

Text word count (excluding title page, abstract, references and figure legends): 5279

No. of items (Figures, Tables, Supplementary items): 8, 2, 6

References: 102

Abstract

The neurodegenerative disorders amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD) reside on a clinical and pathological continuum. Heterogeneity in clinical presentation too often delays clinical diagnosis and calls for molecular biomarkers to assist diagnosis, including cell free microRNAs (miRNA). However, nonlinearity in the relationship of miRNAs to clinical states and underpowered cohorts has limited research in this domain. Here, we prospectively enrolled a large cohort of 495 subjects with ALS (n=202) and FTD (n=168), or non-neurodegenerative controls (n=125). Based on cell-free plasma miRNA profiling by next generation sequencing and machine learning approaches, we develop nonlinear prediction models that accurately distinguish ALS and FTD from non-neurodegenerative controls in ~90% of cases. Among the miRNAs that contribute to classifying disease, we identified miRNAs shared between conditions as well as disease-specific miRNAs. The fascinating potential of diagnostic miRNA biomarkers might enable early-stage detection and a cost-effective screening approach for clinical trials that can facilitate drug development.

Introduction

Amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD) are two neurodegenerative diseases that reside along a clinico-pathological continuum (*1-6*). ALS is primarily a disease of the human motor neuron system, resulting in muscle atrophy, paralysis and death (*7-9*), whereas FTD is characterized by frontal and temporal lobe atrophy, and typically manifests with behavioral or language deficits (*6*).

Shared neuropathological inclusions of the RNA-binding proteins TAR DNA-binding protein 43 (TDP-43) (*3*) and shared causative genes, chromosome 9 open reading frame 72 (*C9ORF72*), *TARDBP* (TDP-43), *VCP* and *TBK1* (*10, 11*), further resonate with comorbidity and clinical symptoms of both diseases in 10-20% of patients. Thus, one in five ALS patients meets the clinical criteria for a concomitant diagnosis of FTD, and one in eight FTD patients is also diagnosed with ALS. However, there are also unique mutations to each disease entity. For example, mutations in heterogeneous nuclear ribonucleoprotein A1 (hnRNPA1) cause ALS, but not FTD (*4, 12-15*), and mutations in progranulin (GRN), and charged multivesicular body protein 2B (CHMP2B) may cause FTD but not ALS.

The diagnosis of ALS is currently based on medical history, clinical examination and exclusion of mimicking disorders. Full-blown ALS is easily diagnosed by trained practitioners, however the diagnostic process often takes 1-1.5 years, since progressive deterioration, indicating both upper and lower motor neuron involvement, is required to define ALS. FTD can also be difficult to diagnose, due to heterogeneity in clinical presentation (*16*). Three main phenotypes of FTD are described: behavioral variant frontotemporal dementia (bvFTD), characterized by changes in social behaviour and conduct, semantic dementia (SD), characterized by the loss of semantic knowledge, leading to impaired word comprehension, and progressive non-fluent aphasia (PNFA), characterized by progressive difficulties in speech production (*16, 17*).

Brain imaging and several biofluid proteins have been proposed as biomarkers for both ALS and FTD (*18-33*). Among the proteins, noticeable are phospho-neurofilament heavy chain, neurofilament light chain (NfL), TDP-43, Tau and pro-inflammatory cytokines. In this context, a recent study concluded that high NfL blood levels are indicative of the intensity of neurodegeneration or the extent of the degenerated axons in both ALS and FTD (*32*) but cannot contribute to the distinction between these two diseases.

miRNAs are a class of small, non-coding RNAs, that can be quantified in biofluids in a massively parallel fashion, yielding fine-grained profiles (*34*). miRNAs play important roles in ALS

pathogenesis and in motor neuron survival (*35-38*). Accordingly, plasma miRNAs were suggested as diagnostic biomarkers for ALS (*39-51*) or FTD (*42, 52-57*)

However, definitive markers have not been reported so far and initial biomarker studies were confounded by cohort size, sample heterogeneity biased or pre-selection of candidate miRNA. Furthermore, diagnosis of comorbid ALS-FTD patients with miRNAs has not been pursued.

In this study, we profiled blood plasma miRNAs (*34*), in a cohort of 533 subjects and developed miRNA-based classifiers for diagnosing ALS and FTD. We implemented ensemble machine learning approach, to address biomarker nonlinearity and were able to expose unrevealed disease-associated signals. The diagnostic power of the study roots from unbiased miRNA signature in a large and heterogeneous meta-cohort, and is validated in an independent held-out cohort according to the TRIPOD guidelines (*58*). Therefore, circulating miRNAs hold a fascinating potential as diagnostic biomarkers and as means for patient stratification in clinical trials.

Results

We sought to determine the overall diagnostic capability of miRNA measurements in ALS and FTD. To this end, we based our study on analysis of plasma miRNA expression and the development of computational diagnostic models. The initial cohorts included a total of 533 participants, enrolled between 2009 and 2019, with clinical follow up time that ranged between 3 and 66 months. Summary of participants' basic characteristics are described in Table 1.

Quality control of data

We have performed next generation RNA sequencing for all samples and subjected miRNA data to quality control steps prior to prediction model development. We observed a younger mean age in both ALS controls (mean age 49.9±13, 95% CI [47.2- 52.5]) and in FTD controls (mean age of 61±13, 95% CI [57-64]), whereas in other groups the mean age was around 65 (Figure S1). Since both ALS and FTD symptoms manifest late in life, with median disease onset at 55 and 58 years of age, respectively, we reduced age-variance by excluding 38 participants younger than 40, which reduced differences in mean age across the remaining meta-cohort of 495 subjects by 33%. Thus, 168 out of 169 FTD patients and 202 out of 206 patients with ALS were included in the analysis along with 125 out of 158 non-neurodegenerative control samples (Table S1). We then verified that a merged dataset of ALS and FTD controls, collected in two different clinical centers does not introduce biases, using Kullback–Leibler divergence measure and a t-distributed stochastic neighbor embedding (t-SNE) analysis (Figure S2A). This enabled analysis of all control data in a unified manner. ALS and FTD cases did not differ from the merged control group, or from each other (Figure S2B).

Out of the >2000 miRNA species that were aligned to the genome, only 132 fulfilled QC criteria. Next, we quantified the differential miRNA levels in ALS, FTD and control samples (Figure 1A-C), noting miRNAs that may distinguish ALS and FTD from control and from one another.

Development of binary classifiers for the diagnosis of ALS or FTD

We first established a baseline single-disease predictor for either ALS or FTD. A diagnostic prediction model for FTD was developed on a randomly selected training set of 135 FTD cases and 84 controls, comprising 75% of the total cohort (168 cases, 125 controls). As a replication cohort, the remaining 25% of the data were held out for model validation (33 FTD, 41 control samples).

The 132 miRNAs were tested as potential predictors, using an ensemble machine learning approach. miRNAs were ranked by their predictive value in the diagnosis of FTD, vs. individuals that did not suffer from neurodegeneration and considered healthy. For selecting the most predictive features during model development, we used *Recursive Feature Elimination* (RFE), an efficient approach that removes miRNAs (features) iteratively to identify those that contribute the most to prediction accuracy (Figure 2A). RFE obtained a set of 13 miRNAs with highest feature importance. The 13-miRNA-based predictor presented a mean receiver operating characteristic area under the curve (ROC AUC) of 0.9, by internal cross validation (Figure 2B). Furthermore, the 13-miRNA-based predictor was capable of predicting the diagnosis of FTD in an independent held-out (replication) cohort of 74 individuals with an ROC AUC of >0.9 (Figure 2C) and an average precision of 0.9. At a defined probability threshold (0.6), the classifier exhibits a 0.8 precision (PPV) with a 0.85 recall (sensitivity), a 0.87 specificity and a 0.2 false discovery rate values. The model is well calibrated and satisfactorily performs on held-out data (Figure 2D, E). Furthermore, a family of models trained with only a subset of the 13 most predictive features, displays a stable AUC ROC performance and reassures that the selected final features are suitable (Figure 2F). In summary, we determined 13 miRNAs that are able to call the diagnosis of FTD with high accuracy.

We then sought to better understand the prediction model by investigating the relative effect of each individual miRNA. Therefore, we utilized post-hoc SHapley Additive exPlanations (SHAP) feature importance analysis to uncover the contribution of individual miRNAs to the FTD diagnostic predictor (Figure 3A, B). The key predictor revealed by SHAP is the astrocyte-derived exosomal miR-361 (*59*). Little is known about other miRNAs contributing to prediction.

We further tested individual miRNAs as single predictors of FTD diagnosis. The ROC analysis revealed maximal AUC values of ~0.69 for miR-423-5p and miR-125b-5p (Table S2). In conclusion, 13 miRNAs are able to call the diagnosis of FTD with high accuracy.

Similarly, an ALS-prediction model was developed by ensemble machine learning with RFE, using data from 149 ALS cases and 96 controls (training set). 53 ALS and 29 controls were held-out as an independent replication (validation) cohort. Following a computational RFE approach comparable to that developed for FTD (in Figure 2A), we depicted 12 miRNAs with the highest feature importance (Figure 4A). The predictor developed with these 12 miRNAs presented a mean ROC AUC of 0.86, by internal cross validation (Figure 4B). Furthermore, that 12-miRNA-based predictor was capable of predicting the diagnosis of FTD in an independent held-out (replication) cohort of 82 individuals with an ROC AUC of 0.78 (Figure 4C). The ALS model is sufficiently

calibrated and a family of models, trained with only a subset of the 12 most predictive features, displays a stable AUC ROC performance (Figure 4 D-F). With a defined probability threshold set to 0.45, the classifier obtained a 0.81 precision (PPV) with a recall (sensitivity) of 0.98 over 0.94 specificity and a 0.19 false discovery rate.

SHAP feature importance analysis uncovered the contribution of miR-206 and miR-1-3p, two homologous muscle-expressed miRNAs, that were reported to increase in the blood of ALS patients (*39, 60-63*). Likewise, miR-103a and miR-181b-5p are enriched in brain tissue and in spinal cord tissues (*64*), and miR-181b-5p has been proposed as a prognostic marker for ALS (*65*). In addition, we tested the diagnostic ability of each individual miRNAs as single predictors in terms of ROC AUC, which revealed predictive value at the range of 0.75 for miR-186-5p (maximal value) and down to 0.39 for miR-320a (minimal value, Table S3). We conclude that some individual miRNAs can be used on their own, however miRNA cooperative predictive value provides a superior performance (>0.78 vs. ~0.75 AUC ROC).

Development of multi-class predictor of ALS and FTD Diagnosis

Multi-disease classifiers may display improved performance by solving a single optimization problem, while exploiting a larger meta-cohort. Such classifiers can take into account the varying levels of relationships that may exist among the different disease entities. Additionally, a predictor that is potentially able to compare ALS and FTD may be informative about the molecular biomarker interpretation of the FTD-ALS continuum (ALS-FTD)(*6*).

We established a meta-cohort composed of 495 subjects, with measurements of 132 miRNAs that passed QC. Then, we applied RFE on to the meta-cohort to reveal 16 miRNAs useful for creating a ternary classifier (ALS vs. FTD vs. healthy, Figure 6A). The multi-disease classifier achieved a weighted mean "one-vs-all" ROC AUC of 0.9 across all folds for all classes during training (Figure 6B) and a weighted "one-vs-all" AUC-ROC of 0.88 on untouched replication cohort (AUC ROC 0.76, 0.86, 0.97 for controls, ALS, FTD, respectively, Figure 6C). The model obtained sensitivity of 0.92/0.83, specificity of 0.95/0.79 and an NPV of 0.96/0.88 for FTD/ALS, respectively (Table 2). We also computed AUC ROC metric for one-vs-one classification, according to the Hand & Till algorithm (*66*): FTD vs controls 0.88, ALS vs controls 0.74, FTD vs ALS 0.97.

In addition, we tested the performance of each of the 16 most predictive miRNAs as single predictors (Table S4) and present the confusion matrices (Figure 6D, E). One miRNA, miR-185-5p, is used to predict both ALS and FTD (Figure 6F). Three other miRNAs, not selected in any of

the binary single-disease models, are exclusively selected in the multi-disease model (miR-24-3p, miR-148a-3p and miR-339-5p, Figure 6F).

SHAP feature importance analysis uncovered the contribution of individual miRNAs to the multi disease predictor (Figure 7A, B, C). The most predictive miRNA, miR-1-3p, is shown to equally influence predicting FTD and ALS, while having only a minor contribution in predicting 'controls' class (Figure 7B,C). miR-206, which is a muscular miRNA like miR-1-3p, is the most predictive in ALS, with almost no predictive capacity in the FTD model. Similarly, mir-361 mostly contribute to risk of FTD. This suggests shared miRNAs that are robust features that may be biologically relevant to the ALS-FTD continuum.

Finally, we tested whether miRNA plasma miRNA classifiers reported in previous works (*42, 56*) perform equally to, better or worse than our classifiers in the training and held out datasets. Combination of miR-34a-5p, miR-345-5p, miR-200c-3p and miR-10a-3p, selected previously (*56*), was consistently inferior to our classifiers in distinguishing between FTD and controls, ALS and controls, and FTD and ALS, with lower AUC values (Figure 8). Similarly, the classifiers selected in Sheinerman et al., 2017 (42) were inferior to our panels of miRNA classifiers. Thus, when tested in comparison to reported miRNA classifiers from the literature, our panels of miRNA classifiers perform better, with AUC values ≥0.78 and in some cases exceeding 0.9.

In summary, we have shown that a panel of <16 miRNAs can distinguish between patients with ALS or FTD and non-degeneration controls, as well as between ALS, FTD and controls, in a multiclass model.

Discussion

The need to facilitate the diagnosis of neurodegeneration, in early disease stages and in the face of clinical heterogeneity, raises the hope for new effective biomarkers for ALS and FTD. Biomarkers may further inform about the comorbidity risk within the ALS-FTD continuum (*6*)

Circulating miRNAs hold a fascinating potential as diagnostic biomarkers that were not fully addressed to date. By using a large cohort of 495 subjects, an unbiased, next generation sequencing approach and advanced computational methods, we overcome limitations of past works in developing biomarkers for neurodegeneration. First, we overcome the specificity problem recently reported for NfL. Since NfL levels are high in many neurodegenerative states, it cannot distinguish ALS from FTD (32). In addition, within miRNA biomarker development, our binary classifiers are more precise than published predictors (*42, 56*). In addition, we develop a multi-disease miRNA predictor model for ALS and FTD.

The use of gradient boosting trees, an ensemble learning approach, allows discovering miRNAs that are in a nonlinear relationship with the disease status. Machine-learning classifiers gain affirmation by cross validation in the training dataset. Furthermore, our models are externally validated on held-out data, which was not used during feature selection and model development, according to the TRIPOD guidelines (*58*).

The multi-disease model learns about related (but different) diagnostic tasks, and jointly selects features for all tasks. This way sample size is increased, features-to-sample ratio of the data is reduced and yet the particularities of each data set are kept (67). In the selection of features, the multi-disease model is intentionally optimized to separate between ALS and FTD by penalizing for features that enhance disease misclassification and prioritizing miRNAs that are differentially measured. This is deciphered when the relative contribution of individual features to the predictive model unfolds by SHAP analysis (*68*).

We revealed that 16 miRNAs are able to classify FTD, ALS or healthy controls, most of which are disease-specific (e.g., miR-206 for ALS and miR-361 for FTD). Interestingly, the most predictive miRNAs are abundant in the neuromuscular and immune systems, which is conceivable, as neuronal loss, muscle atrophy and neuro-inflammation, come into play in these pathologies.

According to the human miRNA tissue atlas (*64*), most of the selected miRNAs are CNS-enriched (let-7c-5p/7i-5p, miR-23b-3p, miR-26a-5p, miR-103a-3p, miR-107, miR-125b-5p, miR-148a-3p, miR-181b-5p, miR-185-5p, miR-320a, miR-339-5p, miR-379 and miR-484), or muscle-enriched (miR-1-3p, miR-26a-5p, and miR-206).

miR-26a, miR-326, miR-484 and miR-361 were associated with FTD diagnosis in our data and with cognitive deficits or Alzheimer's disease (AD) in other analyses (*69-75*). Interestingly, miR-326, miR-484 and miR-361 contributed to FTD diagnosis more than to ALS diagnosis. miR-23b improved cognitive impairments in a rat model of traumatic brain injury (*76*) and alleviated neuronal apoptosis (*77*). Brain-enriched miR-107, was selected as predictor in the FTD model and in the multi-disease model, and was also implicated in AD (*78-83*).

miR-206 and miR-1-3p are muscle-derived miRNAs, that were shown to increase in the blood of ALS patients (*39, 60-63*), presumably as a result of muscle tissue breakdown. Accordingly, both miRNAs were selected as predictive features in the ALS and multiclass models but not the FTD model. However, when breaking out the contribution of these two miRNAs to the diagnosis of ALS, FTD and controls, miR-206 specifically contributes to the diagnosis of ALS, but not of FTD, while the plasma levels of miR-1-3p contribute to prediction of both diseases in the multiclass model. According to the SHAP analysis, it seems that low levels of miR-1-3p contribute to the diagnosis of FTD plausibly by distinguishing it from high levels of miR-1-3p in ALS.

Let-7c and miR-326 are associated with suppression of neuroinflammation (*84, 85*). Changes in blood levels of miR-326, miR-26a, and miR-629 are associated with multiple sclerosis (MS), a disease that results from autoimmune neuroinflammation, (*86-93*). Moreover, serum miR-629 was negatively correlated with MS patient brain volume and lesion severity, respectively (*94*). Thus, some of our miRNA predictors are associated directly or indirectly with neuroinflammation, a process which occurs in both FTD and ALS (*95*).

We would like to put forward a few notable limitations: as is the case with machine learning, the study is limited in providing mechanistic insight for the measured changes in cell-free miRNA levels and why defined nonlinear patterns are strongly predictive of disease states. Discrepancies from conclusions of past studies (*42, 52-54, 56*) are plausibly related to the progress we present in power (larger cohort) and to our unbiased analysis that contrasts past biased choice of miRNA candidates. While we kept held-out cohort for external validation, towards clinical diagnostic usage, it is warranted to test the predictor on an independent cohort of different ethnicity. In addition, the documentation of FTD patients with neuromuscular symptoms and of ALS patients with behavioral symptoms is limited and reduced our ability to investigate the full range of information at the ALS-FTD continuum.

In summary, we have found specific molecular miRNA patterns that can contribute to diagnosis of FTD and ALS. These patterns are further alluding to differences and commonalities in the underlying molecular changes in the ALS - FTD continuum. Therefore, the work encourages

testing if circulating miRNAs biomarkers can be used as a cost-effective screening approach to increase speed or precision in the diagnosis of suspected neurodegenerative states, thereby allowing early-stage detection of high-risk individuals prior to inclusion in prospective clinical trials. More broadly, the findings demonstrate the importance of integrating machine learning into clinical biomarker studies, addressing nonlinearity and exposing otherwise cryptic disease-associated signals.

Methods

Standard protocol approvals, registrations, and patient consents

Approvals were obtained from the local research ethics committee and all participants provided written consent (or gave verbal permission for a carer to sign on their behalf). For ALS samples, recruitment, sampling procedures and data collection have been performed according to Protocol (Protocol number 001, version 5.0 Final – 30th November 2015).

Participants and Sampling

Demographic data of study participants are detailed in Table 1. ALS samples and their respective healthy controls (N = 206 and 102, respectively) were obtained from the ALS biomarker study. (*96*) ALS patients were diagnosed according to standard criteria by experienced ALS neurologists (*96*). Healthy controls were typically spouses or relatives of patients.

FTD subjects and their respective controls were enrolled in the longitudinal FTD cohort studies at UCL. Study cohort included 169 FTD patients and 56 healthy controls.

Blood was collected by venipuncture in EDTA tubes, and plasma was recovered from the whole blood sample by centrifugation and stored at −80°C until RNA extraction subsequent small RNA next generation sequencing. Frozen plasma samples of ALS, FTD and controls from the UCL Biobanks were shipped to the Weizmann Institute of Science for molecular analysis.

Study Design

We determined the sample size by the following power analysis: 150 controls and 150 cases (either ALS or FTD) are required to obtain an ROC AUC of 0.7 with a power of 99% and a p-value of 0.0001. Phenotypic data on de-identified patients was separated and blinded during steps of the molecular analysis.

Small RNA Next Generation Sequencing

Total RNA was extracted from plasma using the miRNeasy micro kit (Qiagen, Hilden, Germany) and quantified with Qubit fluorometer using RNA broad range (BR) assay kit (Thermo Fisher Scientific, Waltham, MA). For small RNA next generation sequencing (RNA-seq), libraries were prepared from 7.5 ng of total RNA using the QIAseq miRNA Library Kit and QIAseq miRNA NGS 48 Index IL (Qiagen), by an experimenter who was blinded to the identity of samples. Samples were randomly allocated to library preparation and sequencing in batches. Precise linear quantification of miRNA is achieved by using unique molecular identifiers (UMIs), of random 12-nucleotide after 3' and 5' adapter ligation, within the reverse transcription primers (*34*). cDNA

libraries were amplified by PCR for 22 cycles, with a 3' primer that includes a 6-nucleotide unique index, followed by on-bead size selection and cleaning. Library concentration was determined with Qubit fluorometer (dsDNA high sensitivity assay kit; Thermo Fisher Scientific, Waltham, MA) and library size with Tapestation D1000 (Agilent). Libraries with different indices were multiplexed and sequenced on NextSeq 500/550 v2 flow cell or Novaseq SP100 (Illumina), with 75bp single read and 6bp index read. Fastq files were de-multiplexed using the user-friendly transcriptome analysis pipeline (UTAP) (*97*). Human miRNAs, as defined by miRBase (*98*), were mapped using the GeneGlobe pipeline (https://geneglobe.qiagen.com/us/analyze). Sequencing data were normalized with DESeq2 package (*99*) under the assumption that miRNA counts followed negative binomial distribution and data were corrected for the library preparation batch in order to reduce its potential bias.

Constructing Cohorts and Biases Associated with the Data-gathering Process

Each disease cohort has been utilized for building and validating a single-disease (binary) classifier that can be used as an applicative predictor for clinicians. Then, a meta-cohort, a unified case-control cohort for both ALS and FTD disease, was constructed to be used for development and assessment of a multi-disease risk prediction model. As our dataset is observational and generated by fusion of few study cohorts, we particularly considered bias associated with the data-gathering process. First, we verified that the control cohorts are not coming from an underlying different population using t-SNE analysis following batch correction and DESeq2 normalization. We then performed QC on the age distribution of subjects and thereby excluded 38 participants younger than 40, to reduce differences in mean age across the different cohorts. A higher prevalence of males among ALS patients (59%, vs. 31 % in controls) was observed. This reflects the higher prevalence of ALS in males than in females, as reported previously (*100*). Therefore, to eliminate the confounding effect, resulting from statistical relationship between sex and ALS status, the sex variable was removed from the list of candidate predictors. A higher prevalence of males among FTD patients (65%) than among the controls (31%) was observed as well. Hence, sex was removed from the list of candidate predictors to eliminate confounding biases.

Verifying de-separation of non-degeneration control cohorts

We established a cohort of 158 non-neurodegenerative controls. Since they originated from two repositories, Queen Mary hospital (N=102) and University College London from (N=56), we first verified that the two sets (distributions) can be considered together and that they do not impose

biases, based on the biorepository from which they were collected. Non-neurodegenerative controls from Queen Mary hospital are presented in light blue (ALS), while controls from University College London are presented in red. The t-distributed stochastic neighbor embedding (t-SNE) analysis reveals that both control distributions could have been sampled from a single population (Figure S1). The Kullback–Leibler divergence, a measure of difference between probability distributions, indicating a small difference between the two sets in question (KL = 0.283). Therefore, control samples were unified and considered as a single meta-cohort for analysis of miRNA abundance, without introducing additional sampling biases.

Gradient Boosted Trees for the Development of Disease Binary Classifiers

The ALS-disease binary classifier and FTD-disease binary classifier were developed using Gradient Boosting Classifier, a machine learning algorithm that uses a gradient boosting framework.

Diagnostic models were developed, validated and reported according to the TRIPOD guidelines (*58*) (https://www.tripod-statement.org/). Gradient Boosting trees (*101, 102*), a decision-tree-based ensemble model, differ fundamentally from conventional statistical techniques that aim to fit a single model using the entire dataset. Such ensemble approach improves performance by combining strengths of models that learn the data by recursive binary splits, such as trees, and of "boosting", an adaptive method for combining several simple (base) models. At each iteration of the gradient boosting algorithm, a subsample of the training data is selected at random (without replacement) from the entire training data set, and then a simple base learner is fitted on each subsample. The final boosted trees model is an additive tree model, constructed by sequentially fitting such base learners on different subsamples. This procedure incorporates randomization, which is known to substantially improve the predictor accuracy and also increase robustness. Additionally, boosted trees can fit complex nonlinear relationships, and automatically handle interaction effects between predictors as addition to other advantages of tree-based methods, such as handling features of different types and accommodating missing data. Hence, in many cases their predictive performance is superior to most traditional modelling methods.

Additional gain of these algorithms is the various loss functions that can be applied, both for binary and multi-class problems. Using the softmax loss function, we explicitly estimated the class conditional probabilities, which allow us to demonstrate the performance of each of the classifiers both as "soft-classifiers" (i.e., predicting class probabilities) and "hard-classifiers" (i.e., setting a

probability threshold and predicting a class). The former approximates a continuous number as output - the class conditional probabilities - and then performs classification based on these estimated probabilities. In contrast, hard classifiers output a discrete number as the decision - directly targeting the classification decision boundary, without producing the probability estimation.

A gradient boosting classifier was developed with a feature set of 132 miRNA predictors, age and sex. Dataset was partitioned to training-set (75%) and validation-set (25%) which was used as held-out data. The training-set was cross-validated during training with stratified 3-fold cross validation. An ROC was generated for each of the folds and individual and mean AUCs were calculated along with 95% confidence intervals.

The chosen hyperparams in ALS-disease classifier: gamma=0.1, learning_rate=0.1, max_depth=3, n_estimators=120, subsample=0.9 and tol=0.01.

The chosen hyperparams in FTD-disease classifier: ccp_alpha=0.0, learning_rate=0.5, max_depth=8, max_features=0.45, min_samples_leaf=14, min_samples_split=8,, n_estimators=100, subsample=0.45 and tol=0.0001.

Gradient Boosted Trees for Multi-class Prediction with "one-vs-the-rest" Strategy

A Gradient Boosting multi-class classifier was implemented with "one-vs-all" (or one-vs-the-rest) strategy for diagnostic prediction between ALS diseased, FTD diseased and control subjects. Performance of a multi-class model is visualized by plotting the performance of 3 binary classifiers. "One-vs-rest" strategy involves fitting a single classifier per class, with the samples of that class considered as positive samples, and all other samples as negatives. A soft-classifier with "one-vs-rest" strategy outputs 3 probabilities for each subject in the data, where a good classifier should assign a high probability to the correct class, while assigning low probabilities to the other classes. Hyperparams for XGBClassifier: objective='multi:softprob', booster='gbtree', gamma=0.01, learning_rate=0.5, max_depth=3, n_estimators=120, subsample=0.9 and tol=0.01.

Dataset was partitioned to training-set (75%, stratified 3-fold cross validation) and validation-set (25%) which was used as held-out data. An ROC was generated for each of the folds and individual and mean AUCs were calculated along with 95% confidence intervals. The area under the ROC curve is a natural measure of separability between the 3 estimated probability distributions and possesses the singular merit that it is independent of costs, priors, or (consequently) any classification threshold(*66*).

Predictor Selection by Recursive Feature Elimination (RFE)

For selecting the most predictive features during prediction model development, we used *Recursive Feature Elimination* (RFE) algorithm, an efficient recursive approach for eliminating features from a training dataset with K-fold cross validation. RFE works by iteratively removing features and using model accuracy to identify which features contribute the most to prediction. Tree-based importance scores of 132 miRNAs were used here in order to rank features, and thus reduce the dimension of miRNA measurements needed for prediction by ~90% (12-16 features in a model in total).

RFE was implemented with Extra Trees Classifier and the following hyperparms: criterion="entropy", max_features=0.9, n_estimators=10. The performance scorer used to optimize selection: ROC AUC score in single disease models and weighted ROC AUC score in the multi-disease model.

Feature Importance and SHAP Analysis

Although gradient boosting tree models are complex models, they can automatically provide an approximation of feature importance from the trained boosted trees. A miRNA predictor is assigned with an importance score in every single tree, where the Gini purity index is used to assess split points in the tree. The score of a feature is calculated based on the amount of improvement in the Gini index achieved by split points that include the feature, weighted by the number of observations in that node. The final importance score of a feature is calculated by an average across all decision trees within the final model.

For local interpretability of the predictive model, we used SHapley Additive exPlanations (SHAP) (*68*), the current state of the art in Machine Learning explainability tools. Giving the smallest set of features that reduced the loss function the most (by XGBoost feature importance), SHAP provides estimates and visualizations to infer what decisions the model is making. This is achieved by quantifying the contribution that each feature brings to each prediction made by the model.

Appendix

IDs of 38 subjects removed due to age-based QC:

'CTRL_rep9', 'CTRL_rep10', 'CTRL_rep13', 'CTRL_rep19', 'CTRL_rep16', 'CTRL_rep18', 'CTRL_rep20', 'FTD30', 'CTRL_rep17', 'CTRL12', 'CTRL13', 'CTRL2', 'CTRL5', 'CTRL56', 'CTRL57', 'CTRL58', 'CTRL68', 'CTRL100', 'CTRL111', 'CTRL20', 'CTRL22', 'CTRL23', 'CTRL24', 'CTRL27', 'CTRL29', 'CTRL33', 'CTRL48', 'CTRL50', 'CTRL54', 'CTRL60', 'CTRL62', 'CTRL73', 'CTRL82', 'CTRL83', 'FA42', 'SA22', 'BUH00025', 'UCH00019'

Acknowledgments

We thank Vittoria Lombardi (UCL) for technical assistance. We acknowledge patients with FTD, ALS, MS and healthy volunteers for their contribution and ALS biomarkers study co-workers for biobanking, which has made this study possible (REC 09/H0703/27). We also thank the North Thames Local Research Network (LCRN) for its support. EH is the Mondry Family Professorial Chair and Head of the Nella and Leon Benoziyo Center for Neurological Diseases.

Funding

EH was supported by the ISF Legacy 828/17 grant, Target ALS 118945 grant, European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013)

/ ERC grant agreement n° 617351. Israel Science Foundation, the ALS-Therapy Alliance, AFM Telethon 20576 grant, Motor Neuron Disease Association (UK), The Thierry Latran Foundation for ALS research, ERA-Net for Research Programmes on Rare Diseases (FP7), A. Alfred Taubman through IsrALS, Yeda-Sela, Yeda-CEO, Israel Ministry of Trade and Industry, Y. Leon Benoziyo Institute for Molecular Medicine, Benoziyo Center Neurological Disease, Kekst Family Institute for Medical Genetics, David and Fela Shapell Family Center for Genetic Disorders Research, Crown Human Genome Center, Nathan, Shirley, Philip and Charlene Vener New Scientist Fund, Julius and Ray Charlestein Foundation, Fraida Foundation, Wolfson Family Charitable Trust, Adelis Foundation, MERCK (UK), Maria Halphen, Estates of Fannie Sherr, Lola Asseof, Lilly Fulop, E. and J. Moravitz. Teva Pharmaceutical Industries Ltd. as part of the Israeli National Network of Excellence in Neuroscience (NNE) postdoc Fellowship to IM 117941. The Dementia Research Centre is supported by Alzheimer's Research UK, Brain Research Trust, and The Wolfson Foundation. This work was supported by the NIHR Queen Square Dementia Biomedical Research Unit, the NIHR UCL/H Biomedical Research Centre and the Leonard Wolfson Experimental Neurology Centre (LWENC) Clinical Research Facility as well as an Alzheimer's Society grant (AS-PG-16-007). JDR is supported by an MRC Clinician Scientist Fellowship (MR/M008525/1) and has received funding from the NIHR Rare Disease Translational Research Collaboration (BRC149/NS/MH). PF is supported by an MRC/MND LEW Fellowship and by the NIHR UCLH BRC. This work was also supported by the Motor Neuron Disease Association (MNDA) 839-791.

The CReATe Consortium (U54 NS090291) is part of the Rare Diseases Clinical Research Network (RDCRN), an initiative of the Office of Rare Diseases Research (ORDR), National Center for Advancing Translational Sciences (NCATS). CReATe is funded through collaboration between NCATS and the National Institute of Neurological Disorders and Stroke (NINDS). Supplementary support for the CReATe biorepository was provided by the ALS Association (Grant ID 16-TACL-242).

NSY was supported by the Israeli Council for Higher Education (CHE) via the Weizmann Data Science Research Center.

Competing interests

The authors state that they have no competing interests.

References

1. S. C. Ling, M. Polymenidou, D. W. Cleveland, Converging mechanisms in ALS and FTD: disrupted RNA and protein homeostasis. *Neuron* **79**, 416-438 (2013).
2. E. Majounie, A. E. Renton, K. Mok, E. G. Dopper, A. Waite, S. Rollinson, A. Chio, G. Restagno, N. Nicolaou, J. Simon-Sanchez, J. C. van Swieten, Y. Abramzon, J. O. Johnson, M. Sendtner, R. Pamphlett, R. W. Orrell, S. Mead, K. C. Sidle, H. Houlden, J. D. Rohrer, K. E. Morrison, H. Pall, K. Talbot, O. Ansorge, A. L. S. F. T. D. C. Chromosome, F. F. A. French research network on, I. Consortium, D. G. Hernandez, S. Arepalli, M. Sabatelli, G. Mora, M. Corbo, F. Giannini, A. Calvo, E. Englund, G. Borghero, G. L. Floris, A. M. Remes, H. Laaksovirta, L. McCluskey, J. Q. Trojanowski, V. M. Van Deerlin, G. D. Schellenberg, M. A. Nalls, V. E. Drory, C. S. Lu, T. H. Yeh, H. Ishiura, Y. Takahashi, S. Tsuji, I. Le Ber, A. Brice, C. Drepper, N. Williams, J. Kirby, P. Shaw, J. Hardy, P. J. Tienari, P. Heutink, H. R. Morris, S. Pickering-Brown, B. J. Traynor, Frequency of the C9orf72 hexanucleotide repeat expansion in patients with amyotrophic lateral sclerosis and frontotemporal dementia: a cross-sectional study. *Lancet Neurol* **11**, 323-330 (2012).
3. M. Neumann, D. M. Sampathu, L. K. Kwong, A. C. Truax, M. C. Micsenyi, T. T. Chou, J. Bruce, T. Schuck, M. Grossman, C. M. Clark, L. F. McCluskey, B. L. Miller, E. Masliah, I. R. Mackenzie, H. Feldman, W. Feiden, H. A. Kretzschmar, J. Q. Trojanowski, V. M. Lee, Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science* **314**, 130-133 (2006).
4. A. E. Renton, E. Majounie, A. Waite, J. Simon-Sanchez, S. Rollinson, J. R. Gibbs, J. C. Schymick, H. Laaksovirta, J. C. van Swieten, L. Myllykangas, H. Kalimo, A. Paetau, Y. Abramzon, A. M. Remes, A. Kaganovich, S. W. Scholz, J. Duckworth, J. Ding, D. W. Harmer, D. G. Hernandez, J. O. Johnson, K. Mok, M. Ryten, D. Trabzuni, R. J. Guerreiro, R. W. Orrell, J. Neal, A. Murray, J. Pearson, I. E. Jansen, D. Sondervan, H. Seelaar, D. Blake, K. Young, N. Halliwell, J. B. Callister, G. Toulson, A. Richardson, A. Gerhard, J. Snowden, D. Mann, D. Neary, M. A. Nalls, T. Peuralinna, L. Jansson, V. M. Isoviita, A. L. Kaivorinne, M. Holtta-Vuori, E. Ikonen, R. Sulkava, M. Benatar, J. Wuu, A. Chio, G. Restagno, G. Borghero, M. Sabatelli, I. Consortium, D. Heckerman, E. Rogaeva, L. Zinman, J. D. Rothstein, M. Sendtner, C. Drepper, E. E. Eichler, C. Alkan, Z. Abdullaev, S. D. Pack, A. Dutra, E. Pak, J. Hardy, A. Singleton, N. M. Williams, P. Heutink, S. Pickering-Brown, H. R. Morris, P. J. Tienari, B. J. Traynor, A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron* **72**, 257-268 (2011).
5. M. A. van Es, O. Hardiman, A. Chio, A. Al-Chalabi, R. J. Pasterkamp, J. H. Veldink, L. H. van den Berg, Amyotrophic lateral sclerosis. *Lancet* **390**, 2084-2098 (2017).
6. M. J. Strong, S. Abrahams, L. H. Goldstein, S. Woolley, P. McLaughlin, J. Snowden, E. Mioshi, A. Roberts-South, M. Benatar, T. HortobaGyi, J. Rosenfeld, V. Silani, P. G. Ince, M. R. Turner, Amyotrophic lateral sclerosis - frontotemporal spectrum disorder (ALS-FTSD): Revised diagnostic criteria. *Amyotroph Lateral Scler Frontotemporal Degener* **18**, 153-174 (2017).
7. A. Al-Chalabi, O. Hardiman, The epidemiology of ALS: a conspiracy of genes, environment and time. *Nat Rev Neurol* **9**, 617-628 (2013).
8. O. Hardiman, L. H. van den Berg, M. C. Kiernan, Clinical diagnosis and management of amyotrophic lateral sclerosis. *Nat Rev Neurol* **7**, 639-649 (2011).
9. W. Robberecht, T. Philips, The changing scene of amyotrophic lateral sclerosis. *Nat Rev Neurosci* **14**, 248-264 (2013).
10. Y. A. Abramzon, P. Fratta, B. J. Traynor, R. Chia, The Overlapping Genetics of Amyotrophic Lateral Sclerosis and Frontotemporal Dementia. *Front Neurosci* **14**, 42 (2020).

11. F. L. Conforti, A. E. Renton, H. Houlden, Editorial: Multifaceted Genes in Amyotrophic Lateral Sclerosis-Frontotemporal Dementia. *Front Neurosci* **15**, 680185 (2021).
12. P. M. Andersen, A. Al-Chalabi, Clinical genetics of amyotrophic lateral sclerosis: what do we really know? *Nat Rev Neurol* **7**, 603-615 (2011).
13. S. Chen, P. Sayana, X. Zhang, W. Le, Genetics of amyotrophic lateral sclerosis: an update. *Mol Neurodegener* **8**, 28 (2013).
14. E. T. Cirulli, B. N. Lasseigne, S. Petrovski, P. C. Sapp, P. A. Dion, C. S. Leblond, J. Couthouis, Y. F. Lu, Q. Wang, B. J. Krueger, Z. Ren, J. Keebler, Y. Han, S. E. Levy, B. E. Boone, J. R. Wimbish, L. L. Waite, A. L. Jones, J. P. Carulli, A. G. Day-Williams, J. F. Staropoli, W. W. Xin, A. Chesi, A. R. Raphael, D. McKenna-Yasek, J. Cady, J. M. Vianney de Jong, K. P. Kenna, B. N. Smith, S. Topp, J. Miller, A. Gkazi, F. S. Consortium, A. Al-Chalabi, L. H. van den Berg, J. Veldink, V. Silani, N. Ticozzi, C. E. Shaw, R. H. Baloh, S. Appel, E. Simpson, C. Lagier-Tourenne, S. M. Pulst, S. Gibson, J. Q. Trojanowski, L. Elman, L. McCluskey, M. Grossman, N. A. Shneider, W. K. Chung, J. M. Ravits, J. D. Glass, K. B. Sims, V. M. Van Deerlin, T. Maniatis, S. D. Hayes, A. Ordureau, S. Swarup, J. Landers, F. Baas, A. S. Allen, R. S. Bedlack, J. W. Harper, A. D. Gitler, G. A. Rouleau, R. Brown, M. B. Harms, G. M. Cooper, T. Harris, R. M. Myers, D. B. Goldstein, Exome sequencing in amyotrophic lateral sclerosis identifies risk genes and pathways. *Science* **347**, 1436-1441 (2015).
15. O. M. Peters, M. Ghasemi, R. H. Brown, Jr., Emerging mechanisms of molecular pathology in ALS. *J Clin Invest* **125**, 1767-1779 (2015).
16. G. M. McKhann, M. S. Albert, M. Grossman, B. Miller, D. Dickson, J. Q. Trojanowski, D. Work Group on Frontotemporal, D. Pick's, Clinical and pathological diagnosis of frontotemporal dementia: report of the Work Group on Frontotemporal Dementia and Pick's Disease. *Arch Neurol* **58**, 1803-1809 (2001).
17. J. S. Snowden, D. Neary, D. M. Mann, Frontotemporal dementia. *Br J Psychiatry* **180**, 140-143 (2002).
18. M. R. Turner, M. C. Kiernan, P. N. Leigh, K. Talbot, Biomarkers in amyotrophic lateral sclerosis. *The Lancet Neurology* **8**, 94-109 (2009).
19. Z. Xu, R. D. Henderson, M. David, P. A. McCombe, Neurofilaments as Biomarkers for Amyotrophic Lateral Sclerosis: A Systematic Review and Meta-Analysis. *PLoS One* **11**, e0164625 (2016).
20. C. E. Teunissen, N. Elias, M. J. Koel-Simmelink, S. Durieux-Lu, A. Malekzadeh, T. V. Pham, S. R. Piersma, T. Beccari, L. H. Meeter, E. G. Dopper, J. C. van Swieten, C. R. Jimenez, Y. A. Pijnenburg, Novel diagnostic cerebrospinal fluid biomarkers for pathologic subtypes of frontotemporal dementia identified by proteomics. *Alzheimers Dement (Amst)* **2**, 86-94 (2016).
21. N. Sheikh-Bahaei, S. A. Sajjadi, A. L. Pierce, Current role for biomarkers in clinical diagnosis of Alzheimer disease and frontotemporal dementia. *Current treatment options in neurology* **19**, 46 (2017).
22. M. Goncalves, M. De Carvalho, C. Peixoto, P. Alves, C. Barreto, A. Oliva, S. Pinto, A. Laborinho-Pronto, M. Gromicho, J. Costa, Phosphoneurofilament heavy chain and vascular endothelial growth factor as cerebrospinal fluid biomarkers for ALS. *Amyotroph Lateral Scler Frontotemporal Degener* **18**, 134-136 (2017).
23. M. T. Gonzalez-Garza, H. R. Martinez, D. E. Cruz-Vega, M. Hernandez-Torre, J. E. Moreno-Cuevas, Adipsin, MIP-1b, and IL-8 as CSF Biomarker Panels for ALS Diagnosis. *Dis Markers* **2018**, 3023826 (2018).
24. L. G. R. Prado, N. P. Rocha, L. C. de Souza, I. C. S. Bicalho, R. S. Gomez, M. Vidigal-Lopes, N. F. T. Braz, E. L. M. Vieira, A. L. Teixeira, Longitudinal assessment of clinical and inflammatory markers in patients with amyotrophic lateral sclerosis. *J Neurol Sci* **394**, 69-74 (2018).

25. D. Rossi, P. Volanti, L. Brambilla, T. Colletti, R. Spataro, V. La Bella, CSF neurofilament proteins as diagnostic and prognostic biomarkers for amyotrophic lateral sclerosis. *J Neurol* **265**, 510-521 (2018).
26. H. Bian, J. C. Van Swieten, S. Leight, L. Massimo, E. Wood, M. Forman, P. Moore, I. de Koning, C. M. Clark, S. Rosso, J. Trojanowski, V. M.-Y. Lee, M. Grossman, CSF biomarkers in frontotemporal lobar degeneration with known pathology. *Neurology* **70**, 1827-1835 (2008).
27. F. Bright, E. L. Werry, C. Dobson-Stone, O. Piguet, L. M. Ittner, G. M. Halliday, J. R. Hodges, M. C. Kiernan, C. T. Loy, M. Kassiou, J. J. Kril, Neuroinflammation in frontotemporal dementia. *Nature Reviews Neurology* **15**, 540-555 (2019).
28. L. M. Forgrave, M. Ma, J. R. Best, M. L. DeMarco, The diagnostic performance of neurofilament light chain in CSF and blood for Alzheimer's disease, frontotemporal dementia, and amyotrophic lateral sclerosis: A systematic review and meta-analysis. *Alzheimers Dement (Amst)* **11**, 730-743 (2019).
29. L. H. Meeter, L. D. Kaat, J. D. Rohrer, J. C. Van Swieten, Imaging and fluid biomarkers in frontotemporal dementia. *Nature Reviews Neurology* **13**, 406 (2017).
30. L. H. H. Meeter, E. G. Vijverberg, M. Del Campo, A. J. M. Rozemuller, L. Donker Kaat, F. J. de Jong, W. M. van der Flier, C. E. Teunissen, J. C. van Swieten, Y. A. L. Pijnenburg, Clinical value of neurofilament and phospho-tau/tau ratio in the frontotemporal dementia spectrum. *Neurology* **90**, e1231-e1239 (2018).
31. E. Feneberg, E. Gray, O. Ansorge, K. Talbot, M. R. Turner, Towards a TDP-43-Based Biomarker for ALS and FTLD. *Mol Neurobiol* **55**, 7789-7801 (2018).
32. N. J. Ashton, S. Janelidze, A. Al Khleifat, A. Leuzy, E. L. van der Ende, T. K. Karikari, A. L. Benedet, T. A. Pascoal, A. Lleo, L. Parnetti, D. Galimberti, L. Bonanni, A. Pilotto, A. Padovani, J. Lycke, L. Novakova, M. Axelsson, L. Velayudhan, G. D. Rabinovici, B. Miller, C. Pariante, N. Nikkheslat, S. M. Resnick, M. Thambisetty, M. Scholl, G. Fernandez-Eulate, F. J. Gil-Bea, A. Lopez de Munain, A. Al-Chalabi, P. Rosa-Neto, A. Strydom, P. Svenningsson, E. Stomrud, A. Santillo, D. Aarsland, J. C. van Swieten, S. Palmqvist, H. Zetterberg, K. Blennow, A. Hye, O. Hansson, A multicentre validation study of the diagnostic value of plasma neurofilament light. *Nat Commun* **12**, 3400 (2021).
33. C. Delaby, D. Alcolea, M. Carmona-Iragui, I. Illan-Gala, E. Morenas-Rodriguez, I. Barroeta, M. Altuna, T. Estelles, M. Santos-Santos, J. Turon-Sans, L. Munoz, R. Ribosa-Nogue, I. Sala-Matavera, B. Sanchez-Saudinos, A. Subirana, L. Videla, B. Benejam, S. Sirisi, S. Lehmann, O. Belbin, J. Clarimon, R. Blesa, J. Pagonabarraga, R. Rojas-Garcia, J. Fortea, A. Lleo, Differential levels of Neurofilament Light protein in cerebrospinal fluid in patients with a wide range of neurodegenerative disorders. *Sci Rep* **10**, 9161 (2020).
34. A. M. L. Coenen-Stass, I. Magen, T. Brooks, I. Z. Ben-Dov, L. Greensmith, E. Hornstein, P. Fratta, Evaluation of methodologies for microRNA biomarker detection by next generation sequencing. *RNA Biol* **15**, 1133-1145 (2018).
35. C. Eitan, E. Hornstein, Vulnerability of microRNA biogenesis in FTD-ALS. *Brain Res*, (2016).
36. A. Emde, C. Eitan, L. L. Liou, R. T. Libby, N. Rivkin, I. Magen, I. Reichenstein, H. Oppenheim, R. Eilam, A. Silvestroni, B. Alajajian, I. Z. Ben-Dov, J. Aebischer, A. Savidor, Y. Levin, R. Sons, S. M. Hammond, J. M. Ravits, T. Moller, E. Hornstein, Dysregulated miRNA biogenesis downstream of cellular stress and ALS-causing mutations: a new mechanism for ALS. *EMBO J* **34**, 2633-2651 (2015).
37. S. Haramati, E. Chapnik, Y. Sztainberg, R. Eilam, R. Zwang, N. Gershoni, E. McGlinn, P. W. Heiser, A. M. Wills, I. Wirguin, L. L. Rubin, H. Misawa, C. J. Tabin, R. Brown, Jr., A. Chen, E. Hornstein, miRNA malfunction causes spinal motor neuron disease. *Proc Natl Acad Sci U S A* **107**, 13111-13116 (2010).

38. I. Reichenstein, C. Eitan, S. Diaz-Garcia, G. Haim, I. Magen, A. Siany, M. L. Hoye, N. Rivkin, T. Olender, B. Toth, R. Ravid, A. D. Mandelbaum, E. Yanowski, J. Liang, J. K. Rymer, R. Levy, G. Beck, E. Ainbinder, S. M. K. Farhan, K. A. Lennox, N. M. Bode, M. A. Behlke, T. Möller, S. Saxena, C. A. M. Moreno, G. Costaguta, K. R. van Eijk, H. Phatnani, A. Al-Chalabi, A. N. Başak, L. H. van den Berg, O. Hardiman, J. E. Landers, J. S. Mora, K. E. Morrison, P. J. Shaw, J. H. Veldink, S. L. Pfaff, O. Yizhar, C. Gross, R. H. Brown, J. M. Ravits, M. B. Harms, T. M. Miller, E. Hornstein, Human genetics and neuropathology suggest a link between miR-218 and amyotrophic lateral sclerosis pathophysiology. *Science Translational Medicine* **11**, eaav5264 (2019).
39. H. M. de Andrade, M. de Albuquerque, S. H. Avansini, S. R. C. de, D. B. Dogini, A. Nucci, B. Carvalho, I. Lopes-Cendes, M. C. Franca, Jr., MicroRNAs-424 and 206 are potential prognostic markers in spinal onset amyotrophic lateral sclerosis. *J Neurol Sci* **368**, 19-24 (2016).
40. I. Takahashi, Y. Hama, M. Matsushima, M. Hirotani, T. Kano, H. Hohzen, I. Yabe, J. Utsumi, H. Sasaki, Identification of plasma microRNAs as a biomarker of sporadic Amyotrophic Lateral Sclerosis. *Mol Brain* **8**, 67 (2015).
41. J. M. Toivonen, R. Manzano, S. Olivan, P. Zaragoza, A. Garcia-Redondo, R. Osta, MicroRNA-206: a potential circulating biomarker candidate for amyotrophic lateral sclerosis. *PLoS One* **9**, e89065 (2014).
42. K. S. Sheinerman, J. B. Toledo, V. G. Tsivinsky, D. Irwin, M. Grossman, D. Weintraub, H. I. Hurtig, A. Chen-Plotkin, D. A. Wolk, L. F. McCluskey, L. B. Elman, J. Q. Trojanowski, S. R. Umansky, Circulating brain-enriched microRNAs as novel biomarkers for detection and differentiation of neurodegenerative diseases. *Alzheimers Res Ther* **9**, 89 (2017).
43. S. A. Banack, R. A. Dunlop, P. A. Cox, An miRNA fingerprint using neural-enriched extracellular vesicles from blood plasma: towards a biomarker for amyotrophic lateral sclerosis/motor neuron disease. *Open Biol* **10**, 200116 (2020).
44. D. Saucier, G. Wajnberg, J. Roy, A. P. Beauregard, S. Chacko, N. Crapoulet, S. Fournier, A. Ghosh, S. M. Lewis, A. Marrero, C. O'Connell, R. J. Ouellette, P. J. Morin, Identification of a circulating miRNA signature in extracellular vesicles collected from amyotrophic lateral sclerosis patients. *Brain Res* **1708**, 100-108 (2019).
45. M. Benigni, C. Ricci, A. R. Jones, F. Giannini, A. Al-Chalabi, S. Battistini, Identification of miRNAs as potential biomarkers in cerebrospinal fluid from amyotrophic lateral sclerosis patients. *Neuromolecular medicine* **18**, 551-560 (2016).
46. F. Cloutier, A. Marrero, C. O'Connell, MicroRNAs as potential circulating biomarkers for amyotrophic lateral sclerosis. *Journal of Molecular Neuroscience* **56**, 102-112 (2015).
47. G. Joilin, P. N. Leigh, S. F. Newbury, M. Hafezparast, An overview of microRNAs as biomarkers of ALS. *Frontiers in neurology* **10**, 186 (2019).
48. M. Ravnik-Glavač, D. Glavač, Circulating RNAs as potential biomarkers in amyotrophic lateral sclerosis. *International journal of molecular sciences* **21**, 1714 (2020).
49. C. Ricci, C. Marzocchi, S. Battistini, MicroRNAs as biomarkers in amyotrophic lateral sclerosis. *Cells* **7**, 219 (2018).
50. R. Waller, E. F. Goodall, M. Milo, J. Cooper-Knock, M. Da Costa, E. Hobson, M. Kazoka, H. Wollff, P. R. Heath, P. J. Shaw, Serum miRNAs miR-206, 143-3p and 374b-5p as potential biomarkers for amyotrophic lateral sclerosis (ALS). *Neurobiology of aging* **55**, 123-131 (2017).
51. L. Wang, L. Zhang, Circulating microRNAs as diagnostic biomarkers for motor neuron disease. *Frontiers in neuroscience* **14**, (2020).
52. J. Denk, F. Oberhauser, J. Kornhuber, J. Wiltfang, K. Fassbender, M. L. Schroeter, A. E. Volk, J. Diehl-Schmid, J. Prudlo, A. Danek, B. Landwehrmeyer, M. Lauer, M. Otto, H. Jahn, F. T. s. group, Specific serum and CSF microRNA profiles distinguish sporadic

behavioural variant of frontotemporal dementia compared with Alzheimer patients and cognitively healthy controls. *PLoS One* **13**, e0197329 (2018).

53. M. Grasso, P. Piscopo, G. Talarico, L. Ricci, A. Crestini, G. Tosto, M. Gasparini, G. Bruno, M. A. Denti, A. Confaloni, Plasma microRNA profiling distinguishes patients with frontotemporal dementia from healthy subjects. *Neurobiol Aging*, (2019).

54. P. Piscopo, M. Grasso, M. Puopolo, E. D'Acunto, G. Talarico, A. Crestini, M. Gasparini, R. Campopiano, S. Gambardella, A. E. Castellano, G. Bruno, M. A. Denti, A. Confaloni, Circulating miR-127-3p as a Potential Biomarker for Differential Diagnosis in Frontotemporal Dementia. *J Alzheimers Dis* **65**, 455-464 (2018).

55. R. Schneider, P. McKeever, T. Kim, C. Graff, J. C. van Swieten, A. Karydas, A. Boxer, H. Rosen, B. L. Miller, R. Laforce, Jr., D. Galimberti, M. Masellis, B. Borroni, Z. Zhang, L. Zinman, J. D. Rohrer, M. C. Tartaglia, J. Robertson, F. T. D. I. Genetic, Downregulation of exosomal miR-204-5p and miR-632 as a biomarker for FTD: a GENFI study. *J Neurol Neurosurg Psychiatry* **89**, 851-858 (2018).

56. V. Kmetzsch, V. Anquetil, D. Saracino, D. Rinaldi, A. Camuzat, T. Gareau, L. Jornea, S. Forlani, P. Couratier, D. Wallon, Plasma microRNA signature in presymptomatic and symptomatic subjects with C9orf72-associated frontotemporal dementia and amyotrophic lateral sclerosis. *Journal of Neurology, Neurosurgery & Psychiatry* **92**, 485-493 (2021).

57. D. Sproviero, S. Gagliardi, S. Zucca, M. Arigoni, M. Giannini, M. Garofalo, M. Olivero, M. Dell'Orco, O. Pansarasa, S. Bernuzzi, M. Avenali, M. Cotta Ramusino, L. Diamanti, B. Minafra, G. Perini, R. Zangaglia, A. Costa, M. Ceroni, N. I. Perrone-Bizzozero, R. A. Calogero, C. Cereda, Different miRNA Profiles in Plasma Derived Small and Large Extracellular Vesicles from Patients with Neurodegenerative Diseases. *Int J Mol Sci* **22**, (2021).

58. G. S. Collins, J. B. Reitsma, D. G. Altman, K. G. Moons, Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med* **162**, 55-63 (2015).

59. X. Bu, D. Li, F. Wang, Q. Sun, Z. Zhang, Protective Role of Astrocyte-Derived Exosomal microRNA-361 in Cerebral Ischemic-Reperfusion Injury by Regulating the AMPK/mTOR Signaling Pathway and Targeting CTSB. *Neuropsychiatr Dis Treat* **16**, 1863-1877 (2020).

60. R. Waller, E. F. Goodall, M. Milo, J. Cooper-Knock, M. Da Costa, E. Hobson, M. Kazoka, H. Wollff, P. R. Heath, P. J. Shaw, J. Kirby, Serum miRNAs miR-206, 143-3p and 374b-5p as potential biomarkers for amyotrophic lateral sclerosis (ALS). *Neurobiol Aging* **55**, 123-131 (2017).

61. R. Raheja, K. Regev, B. C. Healy, M. A. Mazzola, V. Beynon, F. Von Glehn, A. Paul, C. Diaz-Cruz, T. Gholipour, B. I. Glanz, P. Kivisakk, T. Chitnis, H. L. Weiner, J. D. Berry, R. Gandhi, Correlating serum micrornas and clinical parameters in amyotrophic lateral sclerosis. *Muscle Nerve* **58**, 261-269 (2018).

62. E. Tasca, V. Pegoraro, A. Merico, C. Angelini, Circulating microRNAs as biomarkers of muscle differentiation and atrophy in ALS. *Clin Neuropathol* **35**, 22-30 (2016).

63. C. Malacarne, M. Galbiati, E. Giagnorio, P. Cavalcante, F. Salerno, F. Andreetta, C. Cagnoli, M. Taiana, M. Nizzardo, S. Corti, V. Pensato, A. Venerando, C. Gellera, S. Fenu, D. Pareyson, R. Masson, L. Maggi, E. Dalla Bella, G. Lauria, R. Mantegazza, P. Bernasconi, A. Poletti, S. Bonanno, S. Marcuzzo, Dysregulation of Muscle-Specific MicroRNAs as Common Pathogenic Feature Associated with Muscle Atrophy in ALS, SMA and SBMA: Evidence from Animal Models and Human Patients. *Int J Mol Sci* **22**, (2021).

64. N. Ludwig, P. Leidinger, K. Becker, C. Backes, T. Fehlmann, C. Pallasch, S. Rheinheimer, B. Meder, C. Stähler, E. Meese, A. Keller, Distribution of miRNA expression across human tissues. *Nucleic Acids Research* **44**, 3865-3877 (2016).

65. I. Magen, N. S. Yacovzada, E. Yanowski, A. Coenen-Stass, J. Grosskreutz, C.-H. Lu, L. Greensmith, A. Malaspina, P. Fratta, E. Hornstein, Circulating miR-181 is a prognostic biomarker for amyotrophic lateral sclerosis. *bioRxiv*, 833079 (2021).
66. D. J. Hand, R. J. Till, A simple generalisation of the area under the ROC curve for multiple class classification problems. *Machine learning* **45**, 171-186 (2001).
67. C.-A. Azencott, Sorbonne Université , UPMC, (2020).
68. S. M. Lundberg, G. Erion, H. Chen, A. DeGrave, J. M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, S.-I. Lee, From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence* **2**, 56-67 (2020).
69. W. S. Chang, Y. H. Wang, X. T. Zhu, C. J. Wu, Genome-Wide Profiling of miRNA and mRNA Expression in Alzheimer's Disease. *Med Sci Monit* **23**, 2721-2731 (2017).
70. B. He, W. Chen, J. Zeng, W. Tong, P. Zheng, MicroRNA-326 decreases tau phosphorylation and neuron apoptosis through inhibition of the JNK signaling pathway by targeting VAV1 in Alzheimer's disease. *J Cell Physiol* **235**, 480-493 (2020).
71. N. Ludwig, T. Fehlmann, F. Kern, M. Gogol, W. Maetzler, S. Deutscher, S. Gurlit, C. Schulte, A. K. von Thaler, C. Deuschle, F. Metzger, D. Berg, U. Suenkel, V. Keller, C. Backes, H. P. Lenhof, E. Meese, A. Keller, Machine Learning to Detect Alzheimer's Disease from Circulating Non-coding RNAs. *Genomics Proteomics Bioinformatics* **17**, 430-440 (2019).
72. L. Allach El Khattabi, S. Heide, J.-H. Caberg, J. Andrieux, M. Doco Fenzy, C. Vincent-Delorme, P. Callier, S. Chantot-Bastaraud, A. Afenjar, O. Boute-Benejean, M. P. Cordier, L. Faivre, C. Francannet, M. Gerard, A. Goldenberg, A. Masurel-Paulet, A.-L. Mosca-Boidron, N. Marle, A. Moncla, N. Le Meur, M. Mathieu-Dramard, G. Plessis, G. Lesca, M. Rossi, P. Edery, A. Delahaye-Duriez, L. De Pontual, A. C. Tabet, A. Lebbar, L. Suiro, C. Ioos, A. Natiq, S. Chafai Elalaoui, C. Missirian, A. Receveur, C. François-Fiquet, P. Garnier, C. Yardin, C. Laroche, P. Vago, D. Sanlaville, J. M. Dupont, B. Benzacken, E. Pipiras, 16p13.11 microduplication in 45 new patients: refined clinical significance and genotype–phenotype correlations. *Journal of Medical Genetics* **57**, 301-307 (2020).
73. T. F. Goncalves, R. M. Piergiorge, J. M. Dos Santos, J. Gusmao, M. M. G. Pimentel, C. B. Santos-Reboucas, Network Profiling of Brain-Expressed X-Chromosomal MicroRNA Genes Implicates Shared Key MicroRNAs in Intellectual Disability. *J Mol Neurosci* **67**, 295-304 (2019).
74. T. S. Wingo, J. Yang, W. Fan, S. Min Canon, E. S. Gerasimov, A. Lori, B. Logsdon, B. Yao, N. T. Seyfried, J. J. Lah, A. I. Levey, P. A. Boyle, J. A. Schneider, P. L. De Jager, D. A. Bennett, A. P. Wingo, Brain microRNAs associated with late-life depressive symptoms are also associated with cognitive trajectory and dementia. *NPJ Genom Med* **5**, 6 (2020).
75. J. P. Cogswell, J. Ward, I. A. Taylor, M. Waters, Y. Shi, B. Cannon, K. Kelnar, J. Kemppainen, D. Brown, C. Chen, R. K. Prinjha, J. C. Richardson, A. M. Saunders, A. D. Roses, C. A. Richards, Identification of miRNA changes in Alzheimer's disease brain and CSF yields putative biomarkers and insights into disease pathways. *J Alzheimers Dis* **14**, 27-41 (2008).
76. L. Sun, A. Liu, J. Zhang, W. Ji, Y. Li, X. Yang, Z. Wu, J. Guo, miR-23b improves cognitive impairments in traumatic brain injury by targeting ATG12-mediated neuronal autophagy. *Behav Brain Res* **340**, 126-136 (2018).
77. Q. Chen, J. Xu, L. Li, H. Li, S. Mao, F. Zhang, K. Zen, C. Y. Zhang, Q. Zhang, MicroRNA-23a/b and microRNA-27a/b suppress Apaf-1 protein and alleviate hypoxia-induced neuronal apoptosis. *Cell Death Dis* **5**, e1132 (2014).
78. M. Muller, H. B. Kuiperij, J. A. Claassen, B. Kusters, M. M. Verbeek, MicroRNAs in Alzheimer's disease: differential expression in hippocampus and cell-free cerebrospinal fluid. *Neurobiol Aging* **35**, 152-158 (2014).

79. P. T. Nelson, W. X. Wang, MiR-107 is reduced in Alzheimer's disease brain neocortex: validation study. *J Alzheimers Dis* **21**, 75-79 (2010).
80. M. Prendecki, J. Florczak-Wyspianska, M. Kowalska, J. Ilkowski, T. Grzelak, K. Bialas, W. Kozubski, J. Dorszewska, APOE genetic variants and apoE, miR-107 and miR-650 levels in Alzheimer's disease. *Folia Neuropathol* **57**, 106-116 (2019).
81. J. Wang, C. Chen, Y. Zhang, An investigation of microRNA-103 and microRNA-107 as potential blood-based biomarkers for disease risk and progression of Alzheimer's disease. *J Clin Lab Anal* **34**, e23006 (2020).
82. S. G. Yilmaz, M. E. Erdal, A. A. Ozge, M. A. Sungur, Can Peripheral MicroRNA Expression Data Serve as Epigenomic (Upstream) Biomarkers of Alzheimer's Disease? *OMICS* **20**, 456-461 (2016).
83. W. X. Wang, B. W. Rajeev, A. J. Stromberg, N. Ren, G. Tang, Q. Huang, I. Rigoutsos, P. T. Nelson, The expression of microRNA miR-107 decreases early in Alzheimer's disease and may accelerate disease progression through regulation of beta-site amyloid precursor protein-cleaving enzyme 1. *J Neurosci* **28**, 1213-1223 (2008).
84. J. Ni, X. Wang, S. Chen, H. Liu, Y. Wang, X. Xu, J. Cheng, J. Jia, X. Zhen, MicroRNA let-7c-5p protects against cerebral ischemia injury via mechanisms involving the inhibition of microglia activation. *Brain Behav Immun* **49**, 75-85 (2015).
85. X. H. Zhao, Y. B. Wang, J. Yang, H. Q. Liu, L. L. Wang, MicroRNA-326 suppresses iNOS expression and promotes autophagy of dopaminergic neurons through the JNK signaling by targeting XBP1 in a mouse model of Parkinson's disease. *J Cell Biochem* **120**, 14995-15006 (2019).
86. R. L. Lindberg, F. Hoffmann, M. Mehling, J. Kuhle, L. Kappos, Altered expression of miR-17-5p in CD4+ lymphocytes of relapsing-remitting multiple sclerosis patients. *Eur J Immunol* **40**, 888-898 (2010).
87. D. Yang, W. Z. Wang, X. M. Zhang, H. Yue, B. Li, L. Lin, J. Fu, MicroRNA expression aberration in Chinese patients with relapsing remitting multiple sclerosis. *J Mol Neurosci* **52**, 131-137 (2014).
88. M. A. Honardoost, A. Kiani-Esfahani, K. Ghaedi, M. Etemadifar, M. Salehi, miR-326 and miR-26a, two potential markers for diagnosis of relapse and remission phases in patient with relapsing-remitting multiple sclerosis. *Gene* **544**, 128-133 (2014).
89. M. Niwald, M. Migdalska-Sek, E. Brzezianska-Lasota, E. Miller, Evaluation of Selected MicroRNAs Expression in Remission Phase of Multiple Sclerosis and Their Potential Link to Cognition, Depression, and Disability. *J Mol Neurosci* **63**, 275-282 (2017).
90. K. Regev, B. C. Healy, A. Paul, C. Diaz-Cruz, M. A. Mazzola, R. Raheja, B. I. Glanz, P. Kivisakk, T. Chitnis, M. Jagodic, F. Piehl, T. Olsson, M. Khademi, S. Hauser, J. Oksenberg, S. J. Khoury, H. L. Weiner, R. Gandhi, Identification of MS-specific serum miRNAs in an international multicenter study. *Neurol Neuroimmunol Neuroinflamm* **5**, e491 (2018).
91. W. E. Sharaf-Eldin, N. A. Kishk, Y. Z. Gad, H. Hassan, M. A. M. Ali, M. S. Zaki, M. R. Mohamed, M. L. Essawi, Extracellular miR-145, miR-223 and miR-326 expression signature allow for differential diagnosis of immune-mediated neuroinflammatory diseases. *J Neurol Sci* **383**, 188-198 (2017).
92. N. Baulina, O. Kulakova, I. Kiselev, G. Osmak, E. Popova, A. Boyko, O. Favorova, Immune-related miRNA expression patterns in peripheral blood mononuclear cells differ in multiple sclerosis relapse and remission. *J Neuroimmunol* **317**, 67-76 (2018).
93. S. Dolati, L. Aghebati-Maleki, M. Ahmadi, F. Marofi, Z. Babaloo, H. Ayramloo, Z. Jafarisavari, H. Oskouei, A. Afkham, V. Younesi, M. Nouri, M. Yousefi, Nanocurcumin restores aberrant miRNA expression profile in multiple sclerosis, randomized, double-blind, placebo-controlled trial. *J Cell Physiol* **233**, 5222-5230 (2018).

94. K. Regev, B. C. Healy, F. Khalid, A. Paul, R. Chu, S. Tauhid, S. Tummala, C. Diaz-Cruz, R. Raheja, M. A. Mazzola, F. von Glehn, P. Kivisakk, S. L. Dupuy, G. Kim, T. Chitnis, H. L. Weiner, R. Gandhi, R. Bakshi, Association Between Serum MicroRNAs and Magnetic Resonance Imaging Measures of Multiple Sclerosis Severity. *JAMA Neurol* **74**, 275-285 (2017).
95. M. E. McCauley, R. H. Baloh, Inflammation in ALS/FTD pathogenesis. *Acta Neuropathol* **137**, 715-730 (2019).
96. B. R. Brooks, R. G. Miller, M. Swash, T. L. Munsat, D. World Federation of Neurology Research Group on Motor Neuron, El Escorial revisited: revised criteria for the diagnosis of amyotrophic lateral sclerosis. *Amyotroph Lateral Scler Other Motor Neuron Disord* **1**, 293-299 (2000).
97. R. Kohen, J. Barlev, G. Hornung, G. Stelzer, E. Feldmesser, K. Kogan, M. Safran, D. Leshkowitz, UTAP: User-friendly Transcriptome Analysis Pipeline. *BMC bioinformatics* **24**, 154 (2019).
98. A. Kozomara, S. Griffiths-Jones, miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res* **42**, D68-73 (2014).
99. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).
100. G. B. D. M. N. D. Collaborators, Global, regional, and national burden of motor neuron diseases 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Neurol* **17**, 1083-1097 (2018).
101. J. Elith, J. R. Leathwick, T. Hastie, A working guide to boosted regression trees. *J Anim Ecol* **77**, 802-813 (2008).
102. I. H. Witten, E. Frank, M. A. Hall, C. J. Pal, in *Data Mining (Fourth Edition),* I. H. Witten, E. Frank, M. A. Hall, C. J. Pal, Eds. (Morgan Kaufmann, 2017), chap. 12, pp. 479-501.

Main tables and Figures

	Control	ALS	Control	FTD
Number of subjects (% males)	102 (28%)	206 (59%)	56 (46%)	169 (65%)
Age at enrolment	49.9±13.7 yr.	64.4±11.0 yr.	61.0±13.3 yr.	65.6±8.4 yr.
Age of onset (1st reported symptoms)		62.1±11.3 yr.		60.1±8.4 yr.
Disease duration at enrolment		28.2±30.7 m		67.0±41.4 m
ALSFRS-R at enrolment		36.0±8.1		
El-Escorial (Definite/probable/lab-supported/possible/suspected/other)		57/90/27/29/1/2		
Bulbar onset/total		64/206		
Riluzole treated / total		139/206		
ALS Family history / total		9/206		
C9ORF72 genetics/total		14/206		
FTD clinical subtype (bvFTD/PNFA/SD/FTD-ALS/others)				82/40/28/5/14
FTD Mutation carriers (C9ORF72/MAPT/ GRN/TBK1)				18/14/13/2
Likely FTD pathology (TDP-43/Tau)				63/19

Table 1. Summary of demographic and clinical characteristics of ALS and FTD cases and their respective control samples. ALSFRS-R: ALS functional rating scale. bvFTD: behavioural FTD; PNFA: progressive nonfluent aphasia; SD: semantic dementia. Mean±SD.

Class/ Metric	Precision/ PPV	Sensitivity/ Recall	Specificity/ TNR	NPV	f1	# samples
Controls	66%	50%	88%	80%	57%	38
FTD	90%	92%	95%	96%	91%	39
ALS	71%	83%	79%	88%	76%	47
macro avg	75%	75%			75%	124
weighted avg	75%	76%			75%	124

Table 2. *Performance summary of "one-vs-rest" multi-disease classifier*

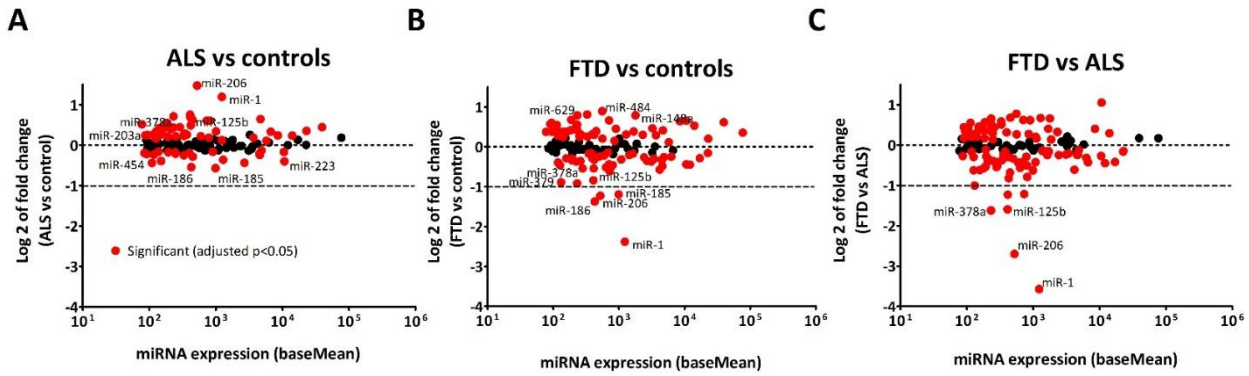


Figure 1. **miRNA signature associated with ALS and FTD**. MA plot of differential miRNA expression in (A) ALS patients (N=202) versus non-neurodegenerative controls (N=125); (B) FTD patients (N=168) versus non-neurodegenerative controls (N=125); (C) FTD versus ALS. Log 2 transformed fold change (y-axis), against mean miRNA abundance (x-axis). Red: significantly changed miRNAs ($p<0.05$, Wald test).

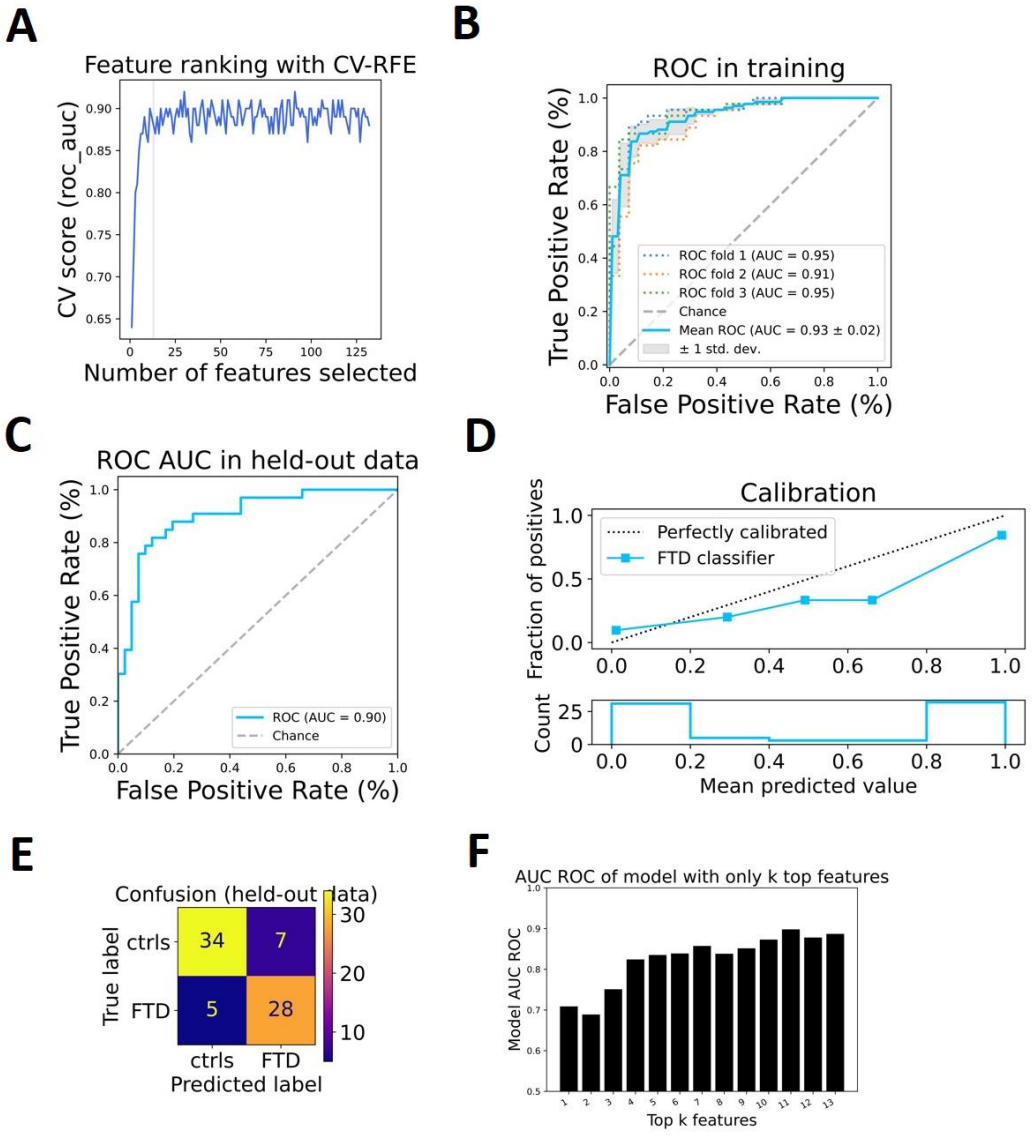


Figure 2: Diagnosis of FTD by a distinctive cell free miRNA signature.

(A) Accuracy (AUC ROC, y-axis) as a function of the number of plasma miRNAs used in the FTD prediction model. Increasing numbers of miRNA features (x-axis) successively-elected in the recursive feature elimination process reveal an optimum at 13 features (blue line).

(B) ROC curves in the training set: true positive rate (y-axis) vs. false positive rate (x-axis). Mean values and variance when data from 219 samples with 3-fold cross validation. Mean AUC 0.93 ± SD. 95% CI is 91%-95%.

(C) Performance and generalizability on held-out data revealed by a ROC curve with AUC of 0.9 and average precision of 0.89.

(D) Reliability diagram, plotting truly observed vs. predicted average probability in five bins reveals a sufficient calibrated model on the held-out set.

(E) Confusion matrix showing the prediction errors on held out dataset. At a probability threshold of 60%, and 0.83 Negative predictive value (NPV), over 0.85 True Positive Rate (also, recall/ sensitivity) and 0.17 False Positive Rate. Specificity is 0.87 and 0.2 false discovery rate.

(F) AUC ROC of a model trained with only a subset of top 13 most predictive features, reveals a stable performance with the selected final features.

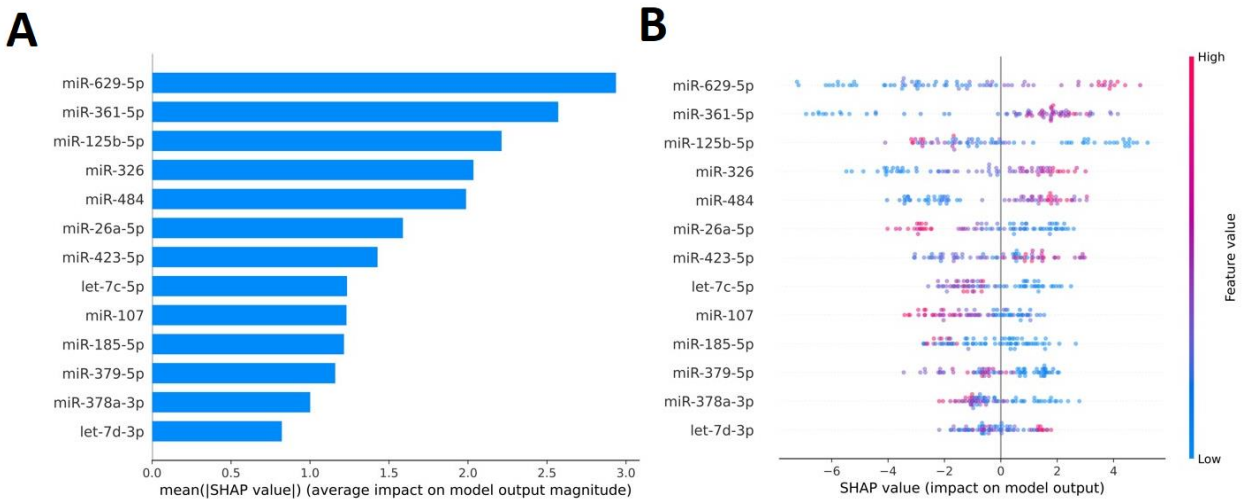


Figure 3: The contribution of individual miRNAs to the predictor of FTD diagnosis.
(A) Mean absolute SHapley Additive exPlanations (SHAP) values break down the impact of specific miRNAs on FTD disease predictor output in the held-out cohort. (B) Illustration of the relationship between the miRNA levels (low - blue to high - red), SHAP values and the impact on the prediction in the held-out cohort

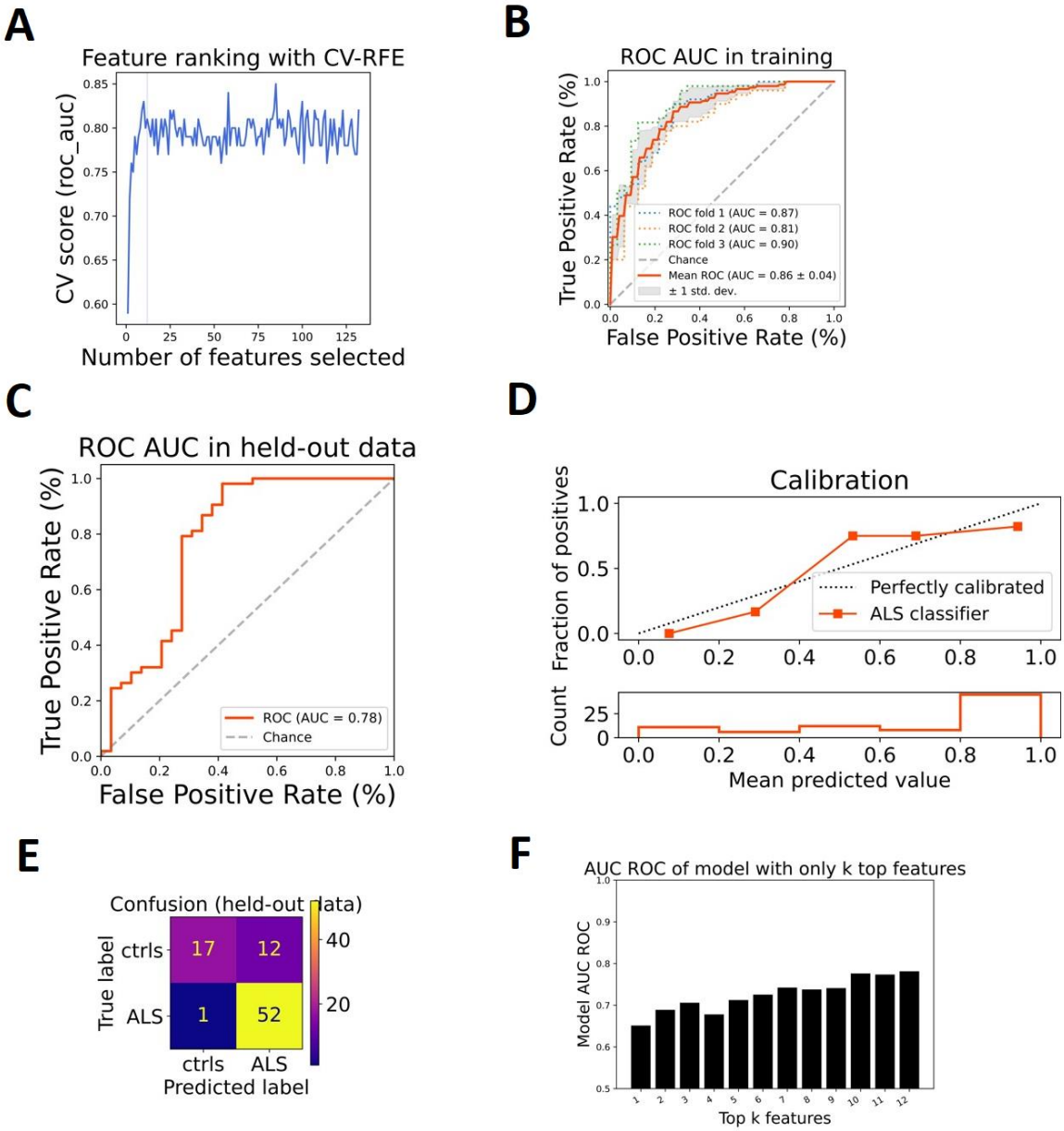


Figure 4: Diagnosis of ALS by a distinctive cell free miRNA signature.

(A) Accuracy (AUC ROC, y-axis) as a function of the number of plasma miRNAs used in the ALS prediction model. Increasing numbers of miRNA features (x-axis) successively elected in the recursive feature elimination process reveal an optimum at 12 features (blue line).

(B) ROC curves in the training set: true positive rate (y-axis) vs. false positive rate (x-axis). Mean values and variance when data from 245 samples with 3-fold cross validation. Mean AUC 0.86 ± SD. 95% CI is 82%-90%.

(C) Performance and generalizability on held-out data revealed by a ROC curve with AUC of 0.78 and average precision of 0.82.

(D) Reliability diagram, plotting truly observed vs. predicted average probability in five bins reveals a calibrated model on the held-out set.

(E) Confusion matrix showing the prediction errors on held out dataset. At a probability threshold of 45%, 0.81 precision (PPV), 0.98 True Positive Rate (also, “recall”/”sensitivity”) and 0.19 False Discovery Rate are demonstrated. Specificity is 0.94.

(F) AUC ROC of a model trained with only a subset of top 12 most predictive features reveals a stable performance with the selected final features.

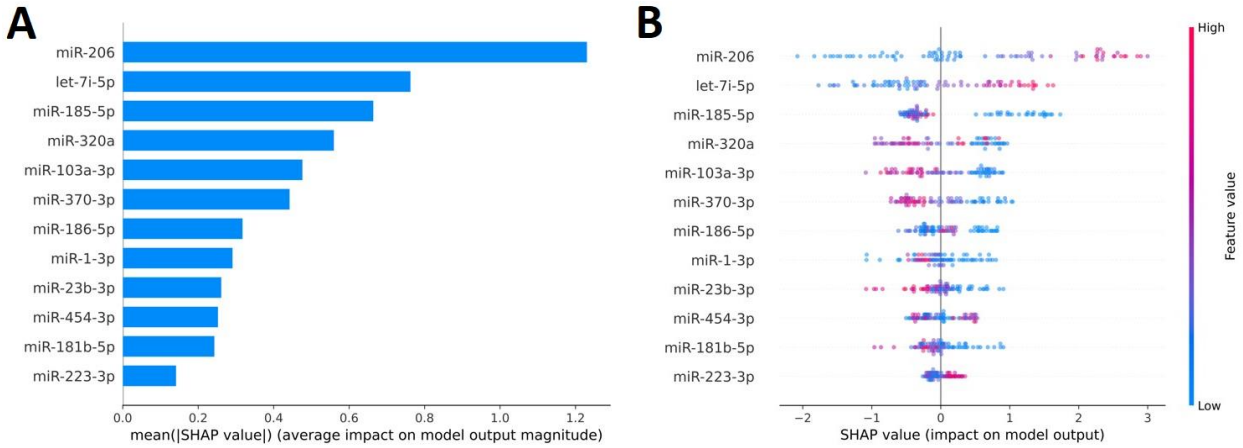


Figure 5: The contribution of individual miRNAs to the predictor of ALS diagnosis.
(A) Mean absolute SHapley Additive exPlanations (SHAP) values break down the impact of specific miRNAs on ALS disease predictor output in the held-out cohort. (B) Illustration of the relationship between the miRNA levels (low-blue to high-red), SHAP values and the impact on the prediction in the held-out cohort.

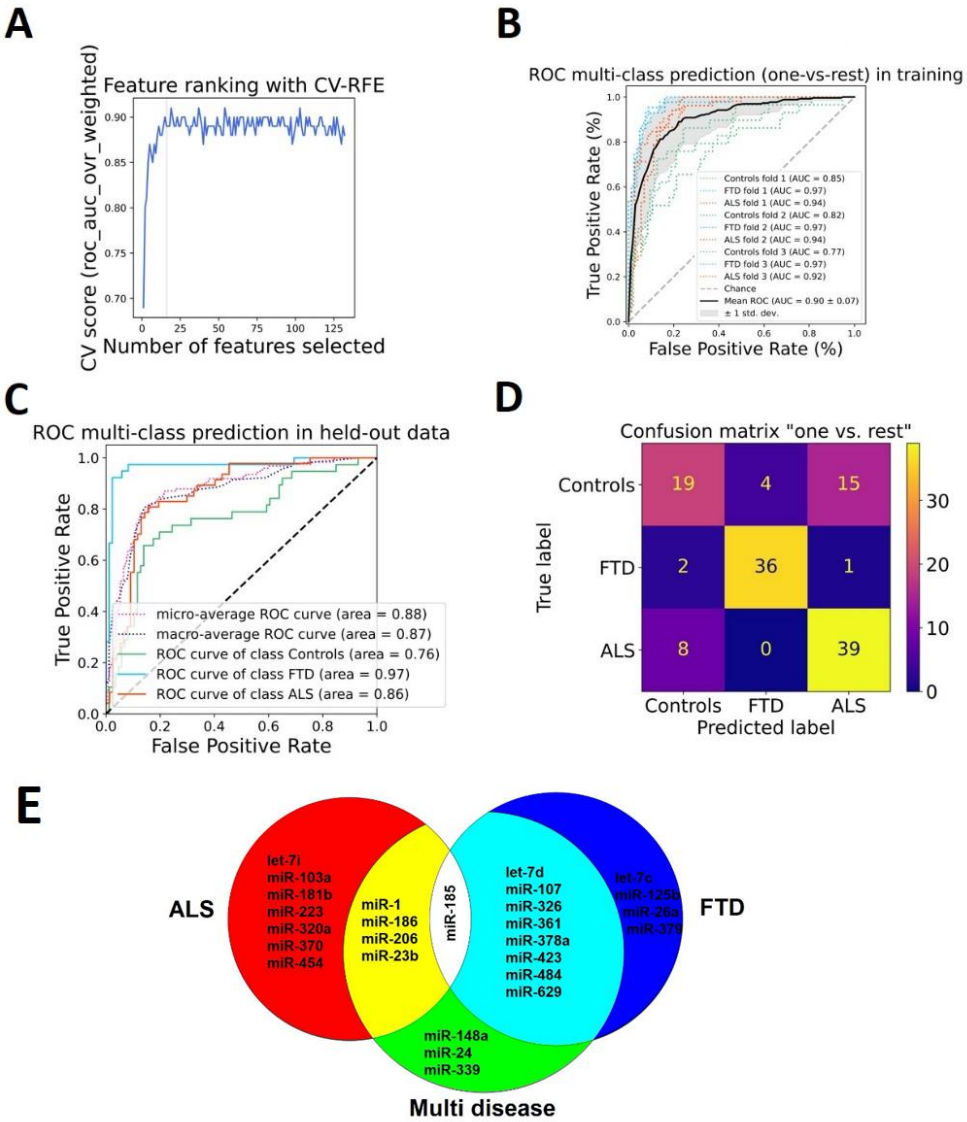


Figure 6: A multi-disease miRNA-based prediction model displays a superior diagnostic power over the binary models. A 3-fold cross-validated RFE with Extra Tree Classifier reveals that 16 features out of the original 132 approximate the same weighted mean one-vs-rest ROC AUC. (B) Fitting a Gradient Boosting multi-class classifier ("one-vs-all" strategy) on the training set. The multi-disease classifier achieved a weighted "one-vs-all" ROC AUC of 0.92, 0.91, 0.89 for each fold, respectively. Mean AUC of 90%±7% is obtained, indicating an overall good fit of the soft-classifier to all classes. Variance of the curve is shown in gray. (C) Model generalizability by ROC AUC per class in the held-out set (AUC of 0.76 for controls, 0.97 for FTD and 0.86 for ALS class), reassuring good accuracy for the soft-classifier over all three classes (micro-average AUC of 0.88 and macro-average AUC of 0.87), ensuring the model is not overfitting. (D) Confusion matrix (with a one-vs-rest strategy) in the held-out data. Interestingly, 15/19 control subjects that were falsely classified were classified as "ALS". It can be observed that of the 16 false positives in 'ALS' class, 93.75% were actually healthy control. (E) a Venn diagram of miRNAs shared in the three classification models.

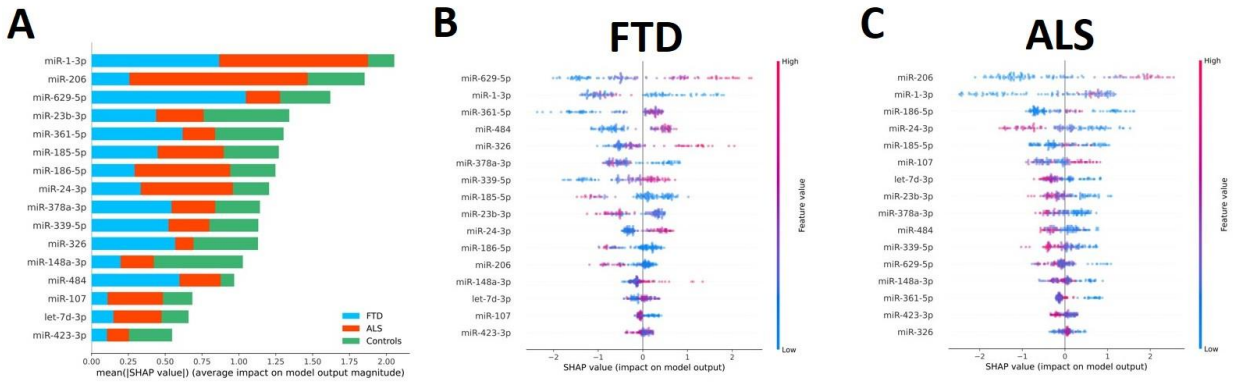


Figure 7: The contribution of individual features miRNAs shared and disease-specific miRNA predictors. (A) SHAP analysis explains how features impact the multi-disease predictor output, namely, class probabilities. Average impact (mean absolute SHAP values) of miRNAs on model output for features, ordered according to their importance over all classes. (B-C) Combining feature importance with feature effects to illustrate the relationship between the value of the feature and the impact on the prediction, for a specific class; each point is a SHAP value for a feature and a specific subject. The position on the y-axis is determined by the feature and scattered on the x-axis by the SHAP value. The color represents the value of the feature from low (blue) to high (red). For example, the levels of mir-1-3p equally influence predictions of FTD and ALS, with negligible contribution to predicting 'controls'. miR-326 and mir-629-5p contributes to the FTD predictor more than ALS, whilst high expression of mir-206 increases the predicted ALS risk (while having no effect on the risk of FTD). High values of miR-326, a miRNA inflammatory regulator, increases the predicted FTD risk while miR-185-5p has the reversed impact with higher values reducing the risk of FTD.

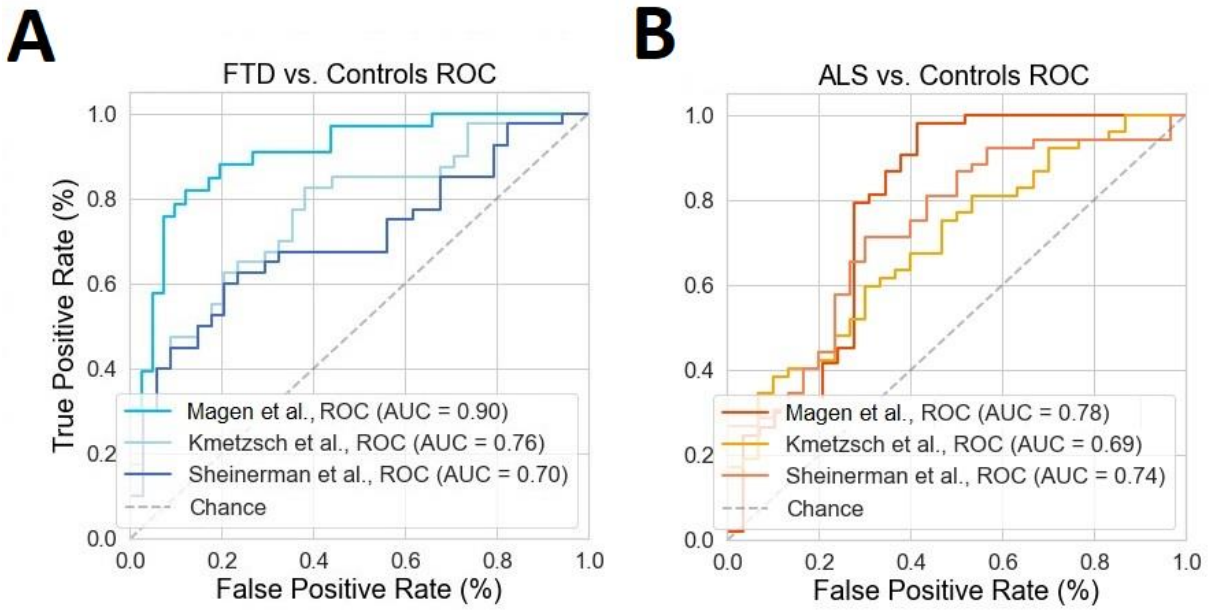


Figure 8. AUC values for different miRNA classifiers for discriminating FTD vs control (A) and ALS vs controls (B) in the held-out data. The classifiers used in (A) were either the 13 selected in the FTD model in our data, a combination of miR-34a-5p, miR-345-5p, miR-200c-3p and miR-10a-3p selected in Kmetzsch et al., 2020 (56), or a combination of miR-335/let-7e, miR-99b/let-7e and miR-9-3p/miR-181a selected in Sheinerman et al., 2017 (42); the classifiers used in (B) were either the 12 selected in the ALS model in our data, a combination of miR-34a-5p, miR-345-5p, miR-200c-3p and miR-10a-3p selected in Kmetzsch et al., 2020 or a combination of miR-206/miR-31, miR-206/miR-125b and miR-99/miR-338-3p, selected in Sheinerman et al., 2017.

Supplementary tables and figures

	Before Exclusion				*After Exclusion*		
Group Name	***Age***	***Sex***	***Mean Age***	***Group Size***	***Mean Age***	***Group Size***	***Age***
ALS cases (n=206)	64.4±11.01, 95% CI [62.9, 65.9]	**F**	65.9	84	66.6	82	64.9±10.4, 95% CI [63.47, 66.35]
		M	63.3	122	63.8	120	
ALS controls (n=102)	49.9±13.68, 95% CI [47.25, 52.55]	**F**	51.1	73	55.3	59	55.7±10, 95% CI [53.46, 57.94]
		M	46.8	29	57.0	18	
FTD cases (n=169)	65.58± 8.37, 95% CI [64.32, 66.84]	**F**	66.8	59	67.3	58	65.75±8.1, 95% CI [64.53, 66.97]
		M	64.9	110	64.9	110	
FTD controls (n= 56)	60.96± 13.28, 95% CI [57.48, 64.44]	**F**	61.1	30	65.6	26	65.55± 7.3, 95% CI [63.49, 67.61]
		M	60.8	26	65.6	22	

Supplementary Table 1. *Age and sex characteristics - before and after exclusion by age (age-based QC)*

Predictor	Single Feature AUC
miR-423-5p	0.69
miR-125b-5p	0.69
miR-26a-5p	0.68
miR-107	0.68
miR-185-5p	0.64
miR-629-5p	0.62
let-7d-3p	0.61
miR-379-5p	0.61
miR-361-5p	0.6
let-7c-5p	0.6
miR-326	0.58
miR-378a-3p	0.56
miR-484	0.55

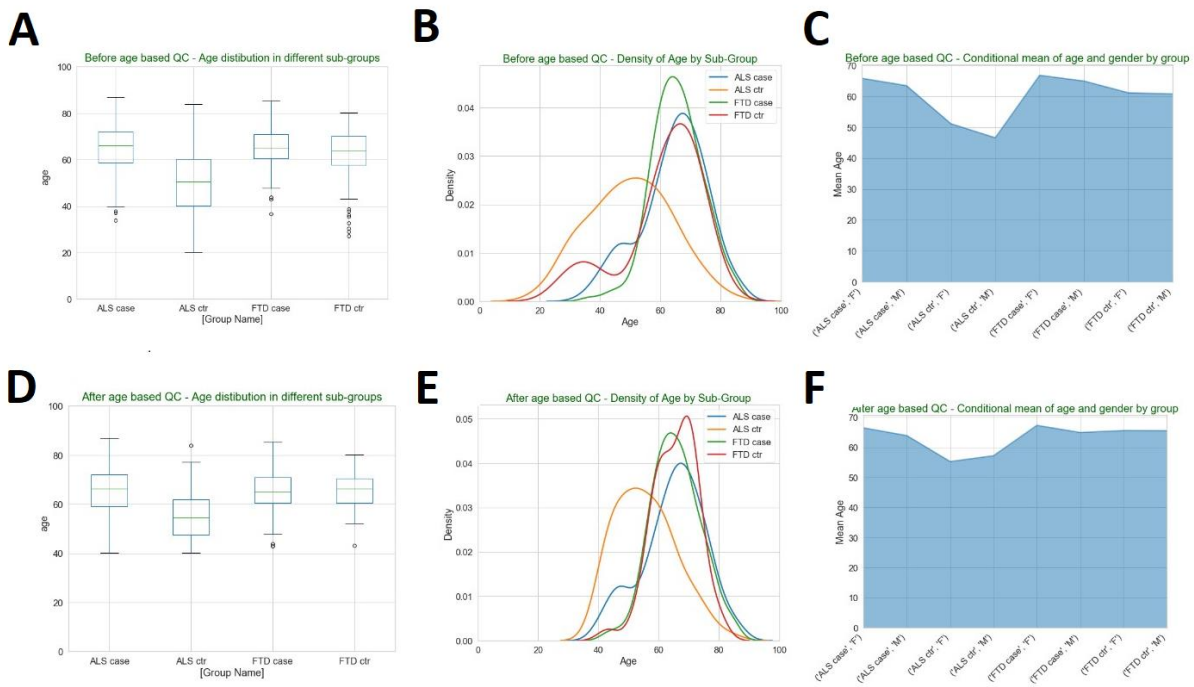
Supplementary Table 2. *Predictive power of selected features when used as a single predictor for FTD on held-out data. Among the most predictive features are miRNAs expressed in the brain, such as miR-26a-5p, miR-125b-5p and let-7c-5p.*

Predictor	Single Feature AUC
miR-186-5p	0.75
miR-223-3p	0.73
miR-23b-3p	0.72
miR-454-3p	0.72
miR-1-3p	0.71
miR-206	0.65
miR-103a-3p	0.62
miR-185-5p	0.61
miR-370-3p	0.56
miR-181b-5p	0.54
let-7i-5p	0.44
miR-320a	0.39

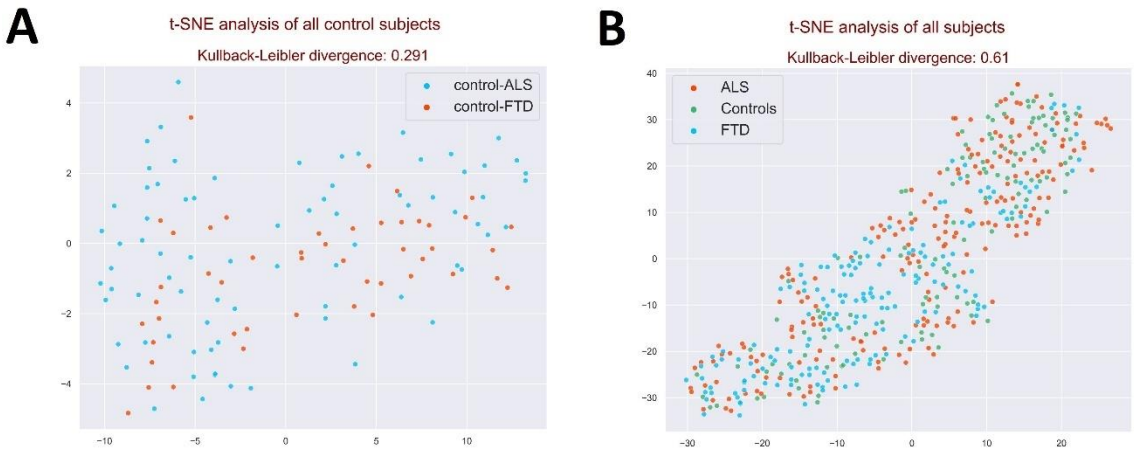
Supplementary Table 3. Predictive power of selected features, when used as a single predictor for ALS diagnosis on held-out data.

Predictor	Single Feature weighted one-vs.-all AUC
miR-1-3p	0.71
miR-185-5p	0.7
miR-186-5p	0.69
miR-629-5p	0.65
miR-206	0.65
miR-423-3p	0.65
miR-378a-3p	0.62
let-7d-3p	0.62
miR-339-5p	0.61
miR-23b-3p	0.6
miR-484	0.58
miR-148a-3p	0.57
miR-107	0.57
miR-361-5p	0.56
miR-326	0.55
miR-24-3p	0.45

Supplementary Table 4. *List of most predictive features and their predictive power (on held-out data) when used as a single predictor for ALS-FTD*



Supplementary Figure 1: QC of prognostic factors in the meta-cohort. Depicting heterogeneity of prognostic factors in the four different groups in the study, before (upper) and after (lower) exclusion of 38 outlier subjects under the age of 40. While the univariate distribution of age in some sub-groups overlap, the difference in the shape of the age distributions may reflect a possible bias induced by age (B). The total standardized differences in mean age across sub-groups was reduced by 68% after exclusion. (A), (D) Boxplot age quartiles and variation of subjects by group - indicating skewness and variability outside the upper and lower quartiles of age in some groups. After outlier exclusion, the groups "ALS-ctr" and "FTD ctr" are more comparable and the high variability in "FTD ctr" is reduced. (B), (E) Distribution of age conditioned by group: estimating the shape of the Probability Density Function (PDF) of age by non-parametric Gaussian kernel density estimate. Outlier removal has increased the overlap of the different PDFs. (C), (F) Mean age conditioned by gender and group. The different groups in the study are more balanced in terms of age and gender after age-outlier removal.



Supplementary Figure 2. *t-SNE analysis of control subjects (A) and all subjects (B) presenting de-separation between different control groups and between disease groups and merged control group.*