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], 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

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

Abstract word count: 153

Text word count (excluding references and figure legends): 5548

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

References: 86

Abstract (153 words)

The neurodegenerative syndromes 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-neurodegeneration 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-degeneration 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 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 continuum (1-5). ALS is primarily a disease of the human motor neuron system, resulting in muscle atrophy, paralysis and death (6-8), whereas FTD is characterized by frontal and temporal lobe atrophy, and typically manifests with behavioral or language deficits (9).

ALS and FTD reside on a genetic and clinical continuum (10). 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 (11, 12) , which 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, 13-16), 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 confounding 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 (17). 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 (17, 18).

Brain imaging and several biofluid proteins have been proposed as biomarkers for both ALS and FTD (19-32). Among the proteins, noticeable are phospho-neurofilament heavy chain, neurofilament light chain (NfL), TDP-43, Tau and pro-inflammatory cytokines.

miRNAs are a class of small, non-coding RNAs, that can be quantified in biofluids in a massively parallel fashion, yielding fine-grained profiles (33). miRNAs play important roles in ALS pathogenesis and in motor neuron survival (34-37). Accordingly, plasma miRNAs were suggested as diagnostic biomarkers for ALS (38-50) or FTD (41, 51-55)

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 (33), in a cohort of 533 subjects and generated miRNA-based classifiers for diagnosing ALS and FTD. We implemented machine learning, to address biomarker nonlinearity and were able to expose unrevealed disease-associated signals that are validated and reported according to the TRIPOD guidelines (56). 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. 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 predictive capability of miRNA measurements in assisting the diagnosis ALS and FTD diseases. 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.

Out of 533 individuals that were initially enrolled, 38 were excluded from the dataset because their age was outlier to cohort's age distribution, to reduce biases induced by age (Methods). 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-neurodegeneration control samples (Table S1).

We have performed next generation RNA sequencing for all samples and subjected miRNA data to quality control steps prior to bioinformatic analysis. We verified that a merged dataset of 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 S1). This enabled analysis of all control data in a unified manner. Out of the >2000 miRNA species that were aligned to the genome, only 132 fulfilled QC criterions and additional 12 miRNAs were excluded since they displayed high correlation to other features (Pearson correlation > 0.85). 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 ALS was developed on a training set of 151 ALS cases and 94 controls, comprising 75% of the total cohort (202 cases, 125 controls). As a replication cohort, the remaining 25% of the data were held out for model validation (52 ALS, 30 control samples).

The remaining 120 miRNAs were tested as potential predictors, using an ensemble machine learning approach. miRNAs were ranked by their predictive value in the diagnosis of ALS, 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). Recursive feature elimination (RFE)

obtained a set of 14 miRNAs with highest feature importance. The 14-miRNA-based predictor presented a mean receiver operating characteristic area under the curve (ROC AUC) of 0.85, by internal cross validation (Figure 2B). Furthermore, that 14-miRNA-based predictor was capable of predicting the diagnosis of ALS in an independent held-out (replication) cohort of 85 individuals with an ROC AUC of 0.85 (Figure 2C) and an average precision of 0.87. At a defined probability threshold (50%), the classifier exhibits a 0.84 precision (PPV) with a 0.92 recall (sensitivity), a 0.84 specificity and a 0.16 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 14 most predictive features, displays a stable AUC ROC performance and reassures that the selected final features are suitable (Figure 2F). In summary, we determined 14 miRNAs that are able to call the diagnosis of ALS 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 ALS diagnostic predictor (Figure 3A, B). SHAP revealed miR-206 and miR-1-3p, two homologous muscle-expressed miRNAs that were reported to increase in the blood of ALS patients (38, 57-59). Likewise, miR-103a and miR-181b-5p are enriched in brain tissue and in spinal cord tissues (60), and miR-181b-5p was suggested as a prognostic marker for ALS (61). In addition, we tested the ability of each individual miRNAs as single predictors. Therefore, we have performed an ROC AUC analysis of the 14 miRNAs, which revealed predictive value at the range of 0.76 for miR-206 (maximal value) and down to 0.43 for miR-20 (minimal value, Table S2). We conclude that some individual miRNAs can be used on their own, however miRNA cooperative predictive value provides a superior performance (>0.85 vs. ~0.76 AUC ROC).

Similarly, an FTD-prediction model was developed by ensemble machine learning with RFE, using data from 168 FTD cases and 125 controls (training set). 41 FTD and 33 controls were held-out as an independent replication (validation) cohort. Following a computational RFE approach comparable to that developed for ALS (in Figure 2), we depicted 18 miRNAs with the highest feature importance (Figure 4A). The predictor developed with these 18 miRNA presented a mean ROC AUC of >0.9 in both training and replication cohorts (Figure 4B, C). The FTD model is well-calibrated and a family of models, trained with only a subset of the 18 most predictive features, displays a stable AUC ROC performance (Figure 4 D-F). With a defined probability threshold set to 0.6, the classifier obtained a 0.82 precision (PPV) with a recall (sensitivity) of 0.85 over 0.875 specificity and 0.85 NPV.

SHAP feature importance analysis uncovered the contribution of individual miRNAs to the predictor of FTD diagnosis (Figure 5 A,B), including of the astrocyte-derived exosomal miR-361 (62). 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 ~0.75 for miR-26a, miR-185 and miR-629 (Table S3). In conclusion, 18 miRNAs are able to call the diagnosis of FTD with high accuracy.

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) (63).

We established a meta-cohort, composed of 495 subjects (ALS: 202, FTD: 168, controls: 125). We excluded 38 individuals whose young age was outlier to the whole cohort and further corrected for batch effects and additional confounders (see Methods). Nine miRNAs out of the 132 candidate predictors were excluded, since they displayed high correlation to other features (Pearson correlation > 0.85). Then, we applied RFE on to the meta-cohort to reveal 18 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 0.88 across all folds for all classes (Figure 6B) and a weighted “one-vs-all” AUC-ROC of 0.9 on untouched replication cohort (AUC ROC 0.81, 0.89, 0.95 for controls, ALS, FTD, respectively, Figure 6C). The model obtained sensitivity of 0.85/0.91, specificity of 0.95/0.73 and an NPV of 0.93/0.93 for FTD/ALS, respectively (Table 2). In addition, we tested the performance of each of the 18 most predictive miRNAs as single predictors (Table S4) and present the confusion matrices (Figure 6D, E). Three miRNAs, miR-1-3p, miR-185-5p and miR-23b-3p, are used to predict both ALS and FTD (Figure 6F). Four other miRNAs, not selected in the binary single-disease models, are exclusively selected in the multi-disease model (miR-142-5p, miR-146b-5p, miR-148a-3p and miR-197-3p, 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 second most

predictive in the ALS model, with almost no predictive capacity in the FTD model. This suggests shared miRNAs that are robust features that may be biologically relevant to the ALS-FTD continuum.

In summary, we have shown that a panel of <20 miRNA 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 (63).

Circulating miRNAs hold a fascinating potential as diagnostic biomarkers that were not fully addressed to date. By using a large cohort, an unbiased, next generation sequencing approach and advanced computational methods, we overcome limitations of past works in developing biomarkers for neurodegeneration according to the TRIPOD guidelines (56). The novelty in our findings is in development of a multi-disease model for ALS and FTD, based on a cell-free miRNA profiling of 495 subjects. We reveal that 18 miRNAs are able to classify FTD, ALS or healthy controls. Therefore, the work encourages testing if miRNAs can be used in a prospective clinical trial to increase speed or precision in the diagnosis of suspected neurodegenerative states.

The use of tree ensemble prediction models allows discovering nonlinear signatures of miRNAs and is superior in this sense to developing biomarkers based on standard expression analysis. In addition, machine-learning classifiers gain affirmation by cross validation in the training dataset and by rigorous testing of findings on an independent replication cohort, which was not used during feature selection and model development. Additionally, increased sample size in gained by learning about multiple related (but different) diagnostic tasks, and jointly selecting features for all tasks. This complementary approach reduces the features-to-sample ratio of the data, while keeping the particularities of each data set (64).

To decipher the relative contribution of an individual feature to the predictive model, we utilized SHapley Additive exPlanations algorithm (65), which quantifies feature contribution relative to the model predictions (Figure 7B,C,D). Interestingly, the most predictive miRNAs are abundant in the neuromuscular and immune systems. This is conceivable as neuron loss, neuro-inflammation, and muscle atrophy come into play in these pathologies. Among the selected miRNAs, brain miRNAs (miR-103a-3p, miR-125b-5p, miR-26a-5p, miR-326 and miR-484) and muscle miRNAs (miR-1-3p, miR-26a-5p, and miR-206) (60). miR-206 greatly contributes to ALS diagnosis, as previously reported (38, 57, 59), but does not contribute to the prediction of FTD, a disease in which muscular tissue is mostly spared. Low blood miR-103a-3p and let-7d-3p levels contribute to ALS prediction in accordance with (58, 66). Interestingly, miR-26a, miR-146b and miR-125b-5p are more predictive of FTD than ALS and miR-629-5p is increased in the serum of multiple sclerosis (MS) patients and negatively correlated with brain volume (67-69). Expression of miR-

326 may be involved in Alzheimer's disease (70) and autoimmune diseases (71, 72). Finally, lower levels of miR-484 and miR-197 were associated with a cognitive decline (73).

Intriguingly, the plasma levels of some miRNAs do contribute to both the prediction of ALS and FTD, which might be related to shared mechanisms within the ALS-FTD continuum. For example, miR-1-3p is a strong predictor of ALS as expected (74), but also of FTD disease.

Limitations: 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 (41, 51-54), 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. However, 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. Circulating miRNAs biomarkers may be further developed to allow early-stage detection of high-risk individuals contribute to ALS-FTD patient subtyping and to cost-effective screening approach for 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. ALS patients were diagnosed according to standard criteria by experienced ALS neurologists (75). Healthy controls were typically spouses or relatives of patients. Plasma samples were stored in -80^{0}C until RNA extraction and subsequent small RNA next generation sequencing.

Participants were enrolled in the longitudinal FTD cohort studies at UCL. Frozen plasma samples from the UCL FTD Biobank were shipped to the Weizmann Institute of Science for molecular analysis. Study cohort included 169 FTD patients and 56 healthy controls.

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. 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 performing RNA-seq.

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 (33). 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) (76). Human miRNAs, as defined by miRBase (77), were mapped using the GeneGlobe pipeline (https://geneglobe.qiagen.com/us/analyze). Sequencing data were normalized with DESeq2 package (78) 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. Baseline characteristics of case and control subjects that choose to volunteer to study might differ from the true population. We therefore handled self-selection bias by standardizing subjects with Inverse Probability Weighting (IPW) in a Propensity Score Modeling. Below, we describe how each bias was examined and handled. 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 (79). 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-neurodegeneration 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-neurodegeneration 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.

Age based QC

ALS and FTD are typically developed after the age of 40. 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 above 60 (Figure S2). 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%.

Meta-cohort: fusion of multiple study cohorts and adjusting for self-selection bias

Baseline characteristics of ALS subjects often differ systematically from those of FTD subjects, and let alone from healthy controls. Simple comparison of subjects may be biased by any factors that predict disease status (e.g., age). In observational studies, it is assumed that some volunteer subjects present an inherent self-selection bias, assuming that some subjects are more likely than others to participate in the study (e.g., the patient's spouse tends to volunteer as a control, control subjects are younger on average, ALS is more prevalent in man, etc.). This problem of prognostic confounding and self-selection bias arises when data is not from a Randomized Control Trial (RCT) and subjects were not randomly sampled from the underlying population in interest(80). Moreover, in a case where several study cohorts are merged, balancing between characteristics of the groups should be induced to prevent potential biases.

Here, we consider a meta-cohort of 495 subjects with four sub-groups:

- 202 ALS cases
- 77 ALS controls
- 168 FTD cases
- 48 FTD controls

A heterogeneity of age and sex was observed between sub-groups (Figure S2). An adjustment to these baseline covariates was made by equating the sub-groups based on age and sex using Propensity Score (PS) Modeling with Random Forest algorithm(81). Age and sex variables were included in the PS model, since they both are related to the disease outcome and therefore are prognostically important covariates (82). PS estimates the probability of a subject in a sub-group given his age and sex (each subject originally volunteered only to one sub-group). The estimated PS score is used to inflate/deflate the weight for subjects who are under/over-represented in the final meta-cohort. We used the inverse PS and applied *Inverse Probability Weighting (IPW)* to create a weighted meta-cohort in which the distribution of measured baseline covariates (age and sex) is independent of what sub-group the subject belongs to. In particular, the propensity score is used as a "balancing score": subjects are weighted by the inverse propensity score, so the distribution of observed baseline covariates will be similar between the different groups of subjects. By making the groups comparable, we thereby ensure that the effect of age is not incorporated in the multi-disease classifier.

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 (56) (https://www.tripod-statement.org/). Gradient Boosting trees (83, 84), 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(85). However, in such a strategy, the FPR metric might be underestimated due to the large number of negative examples resulting from the dichotomization. We therefore show, in addition to conventional ROC curves (TPR vs. FPR), the Precision-Recall curves and the average "one-vs-one" AUC score as established by Hand and Till in 2001(85). As data is treated as a collection of binary problems, we chose to evaluate the "one-vs-rest" multi-class model with a "one-vs-one" generalization as well. Hand and Till's AUC generalization for a multi-class setting consider all distinct pairs of classes i and j (taking different orderings into account, a total of 6 pairs).

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.

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: weighted ROC AUC score.

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. Importance scores of 132 miRNAs were used here in order to rank features in the multi-disease classifier, and thus reduce the dimension of miRNA measurements needed for prediction by 69% (41 features in total).

For local interpretability of the predictive model, we used SHapley Additive exPlanations (SHAP) (65), 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. Ling SC, Polymenidou M, and Cleveland DW. Converging mechanisms in ALS and FTD: disrupted RNA and protein homeostasis. *Neuron.* 2013;79(3):416-38.
2. Majounie E, Renton AE, Mok K, Dopper EG, Waite A, Rollinson S, Chio A, Restagno G, Nicolaou N, Simon-Sanchez J, et al. Frequency of the C9orf72 hexanucleotide repeat expansion in patients with amyotrophic lateral sclerosis and frontotemporal dementia: a cross-sectional study. *Lancet Neurol.* 2012;11(4):323-30.
3. Neumann M, Sampathu DM, Kwong LK, Truax AC, Micsenyi MC, Chou TT, Bruce J, Schuck T, Grossman M, Clark CM, et al. Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science.* 2006;314(5796):130-3.
4. Renton AE, Majounie E, Waite A, Simon-Sanchez J, Rollinson S, Gibbs JR, Schymick JC, Laaksovirta H, van Swieten JC, Myllykangas L, et al. A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron.* 2011;72(2):257-68.
5. van Es MA, Hardiman O, Chio A, Al-Chalabi A, Pasterkamp RJ, Veldink JH, and van den Berg LH. Amyotrophic lateral sclerosis. *Lancet.* 2017;390(10107):2084-98.
6. Al-Chalabi A, and Hardiman O. The epidemiology of ALS: a conspiracy of genes, environment and time. *Nat Rev Neurol.* 2013;9(11):617-28.
7. Hardiman O, van den Berg LH, and Kiernan MC. Clinical diagnosis and management of amyotrophic lateral sclerosis. *Nat Rev Neurol.* 2011;7(11):639-49.
8. Robberecht W, and Philips T. The changing scene of amyotrophic lateral sclerosis. *Nat Rev Neurosci.* 2013;14(4):248-64.
9. Harvey RJ, Skelton-Robinson M, and Rossor MN. The prevalence and causes of dementia in people under the age of 65 years. *Journal of neurology, neurosurgery, and psychiatry.* 2003;74(9):1206-9.
10. Strong MJ, Abrahams S, Goldstein LH, Woolley S, McLaughlin P, Snowden J, Mioshi E, Roberts-South A, Benatar M, HortobaGyi T, et al. Amyotrophic lateral sclerosis - frontotemporal spectrum disorder (ALS-FTSD): Revised diagnostic criteria. *Amyotroph Lateral Scler Frontotemporal Degener.* 2017;18(3-4):153-74.
11. Abramzon YA, Fratta P, Traynor BJ, and Chia R. The Overlapping Genetics of Amyotrophic Lateral Sclerosis and Frontotemporal Dementia. *Front Neurosci.* 2020;14(42.
12. Conforti FL, Renton AE, and Houlden H. Editorial: Multifaceted Genes in Amyotrophic Lateral Sclerosis-Frontotemporal Dementia. *Front Neurosci.* 2021;15(680185.
13. Andersen PM, and Al-Chalabi A. Clinical genetics of amyotrophic lateral sclerosis: what do we really know? *Nat Rev Neurol.* 2011;7(11):603-15.
14. Chen S, Sayana P, Zhang X, and Le W. Genetics of amyotrophic lateral sclerosis: an update. *Mol Neurodegener.* 2013;8(28.
15. Cirulli ET, Lasseigne BN, Petrovski S, Sapp PC, Dion PA, Leblond CS, Couthouis J, Lu YF, Wang Q, Krueger BJ, et al. Exome sequencing in amyotrophic lateral sclerosis identifies risk genes and pathways. *Science.* 2015;347(6229):1436-41.
16. Peters OM, Ghasemi M, and Brown RH, Jr. Emerging mechanisms of molecular pathology in ALS. *J Clin Invest.* 2015;125(5):1767-79.
17. McKhann GM, Albert MS, Grossman M, Miller B, Dickson D, Trojanowski JQ, Work Group on Frontotemporal D, and Pick's D. Clinical and pathological diagnosis of frontotemporal dementia: report of the Work Group on Frontotemporal Dementia and Pick's Disease. *Arch Neurol.* 2001;58(11):1803-9.
18. Snowden JS, Neary D, and Mann DM. Frontotemporal dementia. *Br J Psychiatry.* 2002;180(140-3.

19. Turner MR, Kiernan MC, Leigh PN, and Talbot K. Biomarkers in amyotrophic lateral sclerosis. *The Lancet Neurology.* 2009;8(1):94-109.
20. Xu Z, Henderson RD, David M, and McCombe PA. Neurofilaments as Biomarkers for Amyotrophic Lateral Sclerosis: A Systematic Review and Meta-Analysis. *PLoS One.* 2016;11(10):e0164625.
21. Teunissen CE, Elias N, Koel-Simmelink MJ, Durieux-Lu S, Malekzadeh A, Pham TV, Piersma SR, Beccari T, Meeter LH, Dopper EG, et al. Novel diagnostic cerebrospinal fluid biomarkers for pathologic subtypes of frontotemporal dementia identified by proteomics. *Alzheimer's & dementia (Amsterdam, Netherlands).* 2016;2(86-94.
22. Sheikh-Bahaei N, Sajjadi SA, and Pierce AL. Current role for biomarkers in clinical diagnosis of Alzheimer disease and frontotemporal dementia. *Current treatment options in neurology.* 2017;19(12):46.
23. Goncalves M, De Carvalho M, Peixoto C, Alves P, Barreto C, Oliva A, Pinto S, Laborinho-Pronto A, Gromicho M, and Costa J. Phosphoneurofilament heavy chain and vascular endothelial growth factor as cerebrospinal fluid biomarkers for ALS. *Amyotroph Lateral Scler Frontotemporal Degener.* 2017;18(1-2):134-6.
24. Gonzalez-Garza MT, Martinez HR, Cruz-Vega DE, Hernandez-Torre M, and Moreno-Cuevas JE. Adipsin, MIP-1b, and IL-8 as CSF Biomarker Panels for ALS Diagnosis. *Dis Markers.* 2018;2018(3023826.
25. Prado LGR, Rocha NP, de Souza LC, Bicalho ICS, Gomez RS, Vidigal-Lopes M, Braz NFT, Vieira ELM, and Teixeira AL. Longitudinal assessment of clinical and inflammatory markers in patients with amyotrophic lateral sclerosis. *J Neurol Sci.* 2018;394(69-74.
26. Rossi D, Volanti P, Brambilla L, Colletti T, Spataro R, and La Bella V. CSF neurofilament proteins as diagnostic and prognostic biomarkers for amyotrophic lateral sclerosis. *J Neurol.* 2018;265(3):510-21.
27. Bian H, Van Swieten JC, Leight S, Massimo L, Wood E, Forman M, Moore P, de Koning I, Clark CM, Rosso S, et al. CSF biomarkers in frontotemporal lobar degeneration with known pathology. *Neurology.* 2008;70(19 Part 2):1827-35.
28. Bright F, Werry EL, Dobson-Stone C, Piguet O, Ittner LM, Halliday GM, Hodges JR, Kiernan MC, Loy CT, Kassiou M, et al. Neuroinflammation in frontotemporal dementia. *Nature Reviews Neurology.* 2019;15(9):540-55.
29. Forgrave LM, Ma M, Best JR, and DeMarco ML. 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. *Alzheimer's & dementia (Amsterdam, Netherlands).* 2019;11(730-43.
30. Meeter LH, Kaat LD, Rohrer JD, and Van Swieten JC. Imaging and fluid biomarkers in frontotemporal dementia. *Nature Reviews Neurology.* 2017;13(7):406.
31. Meeter LHH, Vijverberg EG, Del Campo M, Rozemuller AJM, Donker Kaat L, de Jong FJ, van der Flier WM, Teunissen CE, van Swieten JC, and Pijnenburg YAL. Clinical value of neurofilament and phospho-tau/tau ratio in the frontotemporal dementia spectrum. *Neurology.* 2018;90(14):e1231-e9.
32. Feneberg E, Gray E, Ansorge O, Talbot K, and Turner MR. Towards a TDP-43-Based Biomarker for ALS and FTLD. *Mol Neurobiol.* 2018;55(10):7789-801.
33. Coenen-Stass AML, Magen I, Brooks T, Ben-Dov IZ, Greensmith L, Hornstein E, and Fratta P. Evaluation of methodologies for microRNA biomarker detection by next generation sequencing. *RNA Biol.* 2018;15(8):1133-45.
34. Eitan C, and Hornstein E. Vulnerability of microRNA biogenesis in FTD-ALS. *Brain Res.* 2016.
35. Emde A, Eitan C, Liou LL, Libby RT, Rivkin N, Magen I, Reichenstein I, Oppenheim H, Eilam R, Silvestroni A, et al. Dysregulated miRNA biogenesis downstream of cellular

stress and ALS-causing mutations: a new mechanism for ALS. *EMBO J.* 2015;34(21):2633-51.
36. Haramati S, Chapnik E, Sztainberg Y, Eilam R, Zwang R, Gershoni N, McGlinn E, Heiser PW, Wills AM, Wirguin I, et al. miRNA malfunction causes spinal motor neuron disease. *Proc Natl Acad Sci U S A.* 2010;107(29):13111-6.
37. Reichenstein I, Eitan C, Diaz-Garcia S, Haim G, Magen I, Siany A, Hoye ML, Rivkin N, Olender T, Toth B, et al. Human genetics and neuropathology suggest a link between miR-218 and amyotrophic lateral sclerosis pathophysiology. *Science Translational Medicine.* 2019;11(523):eaav5264.
38. de Andrade HM, de Albuquerque M, Avansini SH, de SRC, Dogini DB, Nucci A, Carvalho B, Lopes-Cendes I, and Franca MC, Jr. MicroRNAs-424 and 206 are potential prognostic markers in spinal onset amyotrophic lateral sclerosis. *J Neurol Sci.* 2016;368(19-24.
39. Takahashi I, Hama Y, Matsushima M, Hirotani M, Kano T, Hohzen H, Yabe I, Utsumi J, and Sasaki H. Identification of plasma microRNAs as a biomarker of sporadic Amyotrophic Lateral Sclerosis. *Mol Brain.* 2015;8(1):67.
40. Toivonen JM, Manzano R, Olivan S, Zaragoza P, Garcia-Redondo A, and Osta R. MicroRNA-206: a potential circulating biomarker candidate for amyotrophic lateral sclerosis. *PLoS One.* 2014;9(2):e89065.
41. Sheinerman KS, Toledo JB, Tsivinsky VG, Irwin D, Grossman M, Weintraub D, Hurtig HI, Chen-Plotkin A, Wolk DA, McCluskey LF, et al. Circulating brain-enriched microRNAs as novel biomarkers for detection and differentiation of neurodegenerative diseases. *Alzheimers Res Ther.* 2017;9(1):89.
42. Banack SA, Dunlop RA, and Cox PA. An miRNA fingerprint using neural-enriched extracellular vesicles from blood plasma: towards a biomarker for amyotrophic lateral sclerosis/motor neuron disease. *Open Biol.* 2020;10(6):200116.
43. Saucier D, Wajnberg G, Roy J, Beauregard AP, Chacko S, Crapoulet N, Fournier S, Ghosh A, Lewis SM, Marrero A, et al. Identification of a circulating miRNA signature in extracellular vesicles collected from amyotrophic lateral sclerosis patients. *Brain Res.* 2019;1708(100-8.
44. Benigni M, Ricci C, Jones AR, Giannini F, Al-Chalabi A, and Battistini S. Identification of miRNAs as potential biomarkers in cerebrospinal fluid from amyotrophic lateral sclerosis patients. *Neuromolecular medicine.* 2016;18(4):551-60.
45. Cloutier F, Marrero A, and O'Connell C. MicroRNAs as potential circulating biomarkers for amyotrophic lateral sclerosis. *Journal of Molecular Neuroscience.* 2015;56(1):102-12.
46. Joilin G, Leigh PN, Newbury SF, and Hafezparast M. An overview of microRNAs as biomarkers of ALS. *Frontiers in neurology.* 2019;10(186.
47. Ravnik-Glavač M, and Glavač D. Circulating RNAs as potential biomarkers in amyotrophic lateral sclerosis. *International journal of molecular sciences.* 2020;21(5):1714.
48. Ricci C, Marzocchi C, and Battistini S. MicroRNAs as biomarkers in amyotrophic lateral sclerosis. *Cells.* 2018;7(11):219.
49. Waller R, Goodall EF, Milo M, Cooper-Knock J, Da Costa M, Hobson E, Kazoka M, Wollff H, Heath PR, and Shaw PJ. Serum miRNAs miR-206, 143-3p and 374b-5p as potential biomarkers for amyotrophic lateral sclerosis (ALS). *Neurobiology of aging.* 2017;55(123-31.
50. Wang L, and Zhang L. Circulating microRNAs as diagnostic biomarkers for motor neuron disease. *Frontiers in neuroscience.* 2020;14(
51. Denk J, Oberhauser F, Kornhuber J, Wiltfang J, Fassbender K, Schroeter ML, Volk AE, Diehl-Schmid J, Prudlo J, Danek A, et al. Specific serum and CSF microRNA profiles distinguish sporadic behavioural variant of frontotemporal dementia compared with Alzheimer patients and cognitively healthy controls. *PLoS One.* 2018;13(5):e0197329.

52. Grasso M, Piscopo P, Talarico G, Ricci L, Crestini A, Tosto G, Gasparini M, Bruno G, Denti MA, and Confaloni A. Plasma microRNA profiling distinguishes patients with frontotemporal dementia from healthy subjects. *Neurobiol Aging.* 2019.
53. Piscopo P, Grasso M, Puopolo M, D'Acunto E, Talarico G, Crestini A, Gasparini M, Campopiano R, Gambardella S, Castellano AE, et al. Circulating miR-127-3p as a Potential Biomarker for Differential Diagnosis in Frontotemporal Dementia. *Journal of Alzheimer's disease : JAD.* 2018;65(2):455-64.
54. Schneider R, McKeever P, Kim T, Graff C, van Swieten JC, Karydas A, Boxer A, Rosen H, Miller BL, Laforce R, Jr., et al. Downregulation of exosomal miR-204-5p and miR-632 as a biomarker for FTD: a GENFI study. *Journal of neurology, neurosurgery, and psychiatry.* 2018;89(8):851-8.
55. Kmetzsch V, Anquetil V, Saracino D, Rinaldi D, Camuzat A, Gareau T, Jornea L, Forlani S, Couratier P, and Wallon D. Plasma microRNA signature in presymptomatic and symptomatic subjects with C9orf72-associated frontotemporal dementia and amyotrophic lateral sclerosis. *Journal of Neurology, Neurosurgery & Psychiatry.* 2021;92(5):485-93.
56. Collins GS, Reitsma JB, Altman DG, and Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med.* 2015;162(1):55-63.
57. Waller R, Goodall EF, Milo M, Cooper-Knock J, Da Costa M, Hobson E, Kazoka M, Wollff H, Heath PR, Shaw PJ, et al. Serum miRNAs miR-206, 143-3p and 374b-5p as potential biomarkers for amyotrophic lateral sclerosis (ALS). *Neurobiol Aging.* 2017;55(123-31.
58. Raheja R, Regev K, Healy BC, Mazzola MA, Beynon V, Von Glehn F, Paul A, Diaz-Cruz C, Gholipour T, Glanz BI, et al. Correlating serum micrornas and clinical parameters in amyotrophic lateral sclerosis. *Muscle Nerve.* 2018;58(2):261-9.
59. Tasca E, Pegoraro V, Merico A, and Angelini C. Circulating microRNAs as biomarkers of muscle differentiation and atrophy in ALS. *Clin Neuropathol.* 2016;35(1):22-30.
60. Ludwig N, Leidinger P, Becker K, Backes C, Fehlmann T, Pallasch C, Rheinheimer S, Meder B, Stähler C, Meese E, et al. Distribution of miRNA expression across human tissues. *Nucleic Acids Research.* 2016;44(8):3865-77.
61. Magen I, Yacovzada NS, Yanowski E, Coenen-Stass A, Grosskreutz J, Lu C-H, Greensmith L, Malaspina A, Fratta P, and Hornstein E. Circulating miR-181 is a prognostic biomarker for amyotrophic lateral sclerosis. *bioRxiv.* 2021:833079.
62. Bu X, Li D, Wang F, Sun Q, and Zhang Z. 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.* 2020;16(1863-77.
63. Vural V, Fung G, Rosales R, and Dy JG. *Proceedings of the 21st international jont conference on Artifical intelligence*. Pasadena, California, USA: Morgan Kaufmann Publishers Inc.; 2009:1267–72.
64. Azencott C-A. Sorbonne Université , UPMC; 2020.
65. Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, Katz R, Himmelfarb J, Bansal N, and Lee S-I. From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence.* 2020;2(56-67.
66. Liguori M, Nuzziello N, Introna A, Consiglio A, Licciulli F, D'Errico E, Scarafino A, Distaso E, and Simone IL. Dysregulation of MicroRNAs and Target Genes Networks in Peripheral Blood of Patients With Sporadic Amyotrophic Lateral Sclerosis. *Front Mol Neurosci.* 2018;11(288.
67. Lindberg RL, Hoffmann F, Mehling M, Kuhle J, and Kappos L. Altered expression of miR-17-5p in CD4+ lymphocytes of relapsing-remitting multiple sclerosis patients. *Eur J Immunol.* 2010;40(3):888-98.

68. Yang D, Wang WZ, Zhang XM, Yue H, Li B, Lin L, and Fu J. MicroRNA expression aberration in Chinese patients with relapsing remitting multiple sclerosis. *J Mol Neurosci.* 2014;52(1):131-7.
69. Regev K, Healy BC, Khalid F, Paul A, Chu R, Tauhid S, Tummala S, Diaz-Cruz C, Raheja R, Mazzola MA, et al. Association Between Serum MicroRNAs and Magnetic Resonance Imaging Measures of Multiple Sclerosis Severity. *JAMA neurology.* 2017;74(3):275-85.
70. He B, Chen W, Zeng J, Tong W, and Zheng P. 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.* 2020;235(1):480-93.
71. Choi I, Woo JH, Jou I, and Joe EH. PINK1 Deficiency Decreases Expression Levels of mir-326, mir-330, and mir-3099 during Brain Development and Neural Stem Cell Differentiation. *Exp Neurobiol.* 2016;25(1):14-23.
72. Jadideslam G, Ansarin K, Sakhinia E, Alipour S, Pouremamali F, and Khabbazi A. The MicroRNA-326: Autoimmune diseases, diagnostic biomarker, and therapeutic target. *J Cell Physiol.* 2018;233(12):9209-22.
73. Wingo TS, Yang J, Fan W, Min Canon S, Gerasimov ES, Lori A, Logsdon B, Yao B, Seyfried NT, Lah JJ, et al. Brain microRNAs associated with late-life depressive symptoms are also associated with cognitive trajectory and dementia. *NPJ Genom Med.* 2020;5(6.
74. Chen JF, Tao Y, Li J, Deng Z, Yan Z, Xiao X, and Wang DZ. microRNA-1 and microRNA-206 regulate skeletal muscle satellite cell proliferation and differentiation by repressing Pax7. *J Cell Biol.* 2010;190(5):867-79.
75. Brooks BR, Miller RG, Swash M, Munsat TL, and World Federation of Neurology Research Group on Motor Neuron D. El Escorial revisited: revised criteria for the diagnosis of amyotrophic lateral sclerosis. *Amyotroph Lateral Scler Other Motor Neuron Disord.* 2000;1(5):293-9.
76. Kohen R, Barlev J, Hornung G, Stelzer G, Feldmesser E, Kogan K, Safran M, and Leshkowitz D. UTAP: User-friendly Transcriptome Analysis Pipeline. *BMC bioinformatics.* 2019;24(154.
77. Kozomara A, and Griffiths-Jones S. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.* 2014;42(Database issue):D68-73.
78. Love MI, Huber W, and Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15(12):550.
79. Collaborators GBDMND. Global, regional, and national burden of motor neuron diseases 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Neurol.* 2018;17(12):1083-97.
80. Bareinboim E, and Pearl J. Causal inference and the data-fusion problem. *Proc Natl Acad Sci U S A.* 2016;113(27):7345-52.
81. Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behav Res.* 2011;46(3):399-424.
82. Austin PC, Grootendorst P, and Anderson GM. A comparison of the ability of different propensity score models to balance measured variables between treated and untreated subjects: a Monte Carlo study. *Stat Med.* 2007;26(4):734-53.
83. Elith J, Leathwick JR, and Hastie T. A working guide to boosted regression trees. *J Anim Ecol.* 2008;77(4):802-13.
84. Witten IH, Frank E, Hall MA, and Pal CJ. In: Witten IH, Frank E, Hall MA, and Pal CJ eds. *Data Mining (Fourth Edition)*. Morgan Kaufmann; 2017:479-501.
85. Hand DJ, and Till RJ. A simple generalisation of the area under the ROC curve for multiple class classification problems. *Machine learning.* 2001;45(2):171-86.

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	76%	47%	93%	79%	58%	40
FTD	89%	85%	96%	93%	87%	39
ALS	66%	91%	73%	94%	77%	45
macro avg	77%	74%			74%	124
weighted avg	77%	75%			74%	124

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

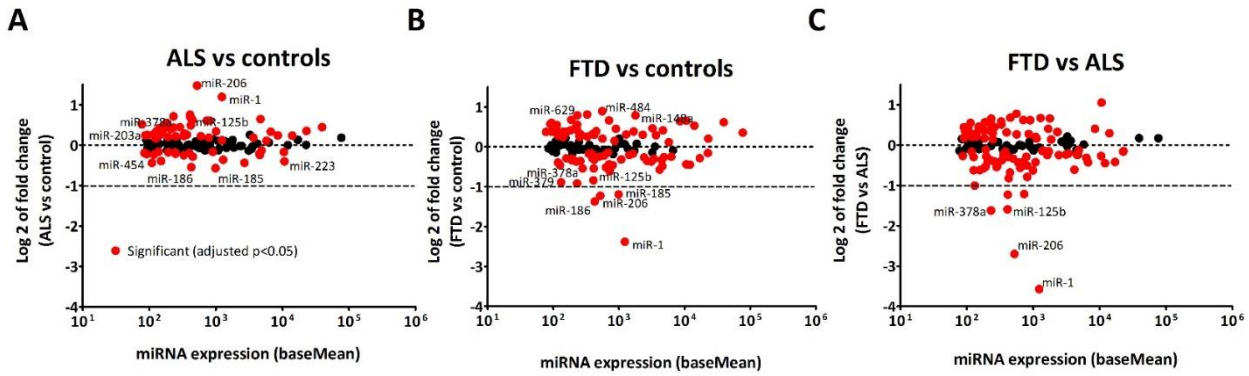


Figure 1. **miRNA signature associated with ALS and FTD**. MA plot of differential miRNA expression in (A) ALS patients (N=202) versus non-neurodegeneration controls (N=125); (B) FTD patients (N=168) versus non-neurodegeneration controls (N=125); (C) FTD versus ALS. Log 2 transformed fold change (y-axis),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

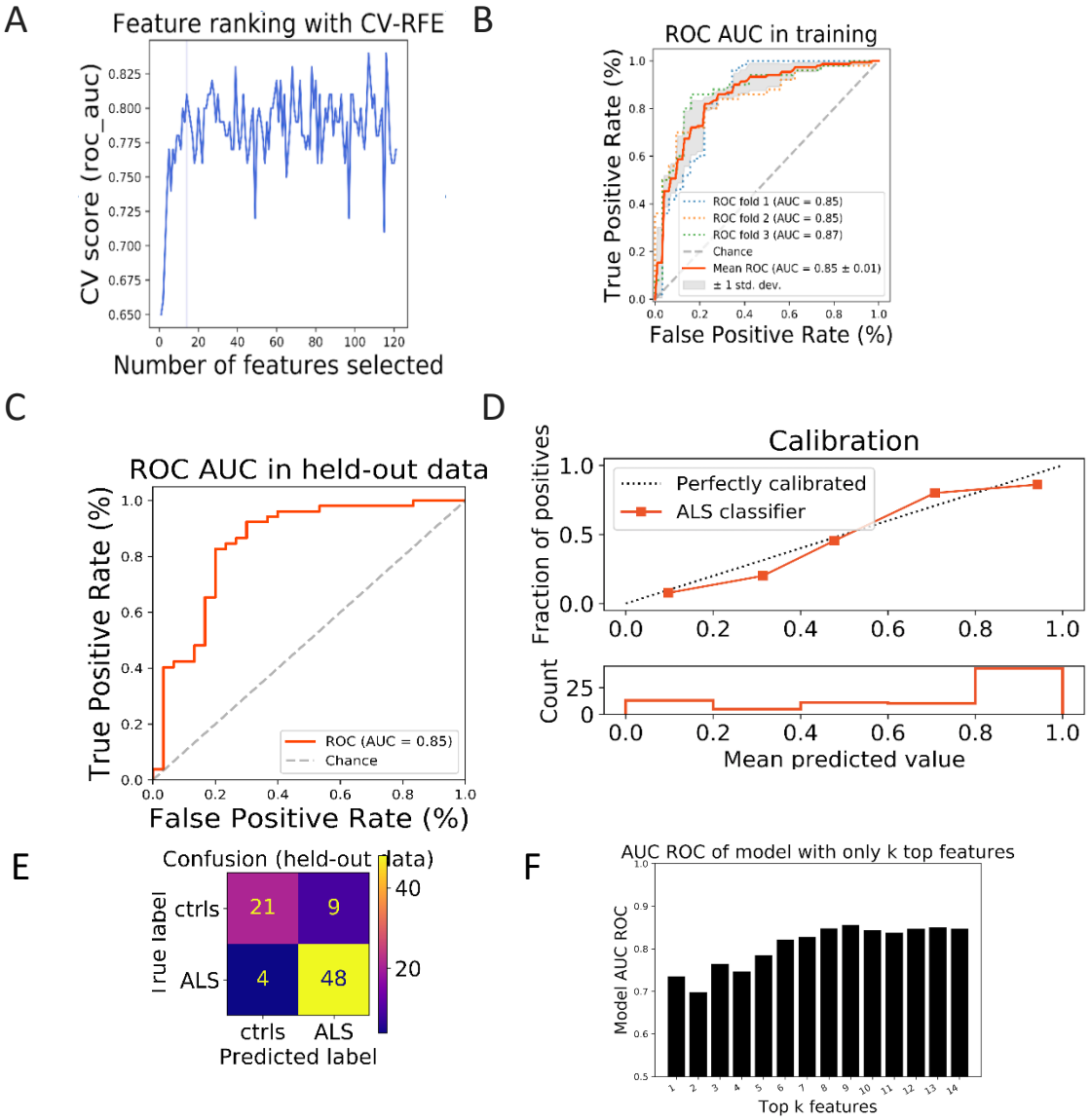


Figure 2: Diagnosis of ALS by a distinctive cell free miRNA signature.

(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 14 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 0.85 ± SD. 95% CI is 84%-86%.

(C) Performance and generalizability on held-out data revealed by a ROC curve with AUC of 0.85 and average precision of 0.87.

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

(E) Confusion matrix showing the prediction errors on held out dataset. At a probability threshold of 50%, 0.84 precision (PPV), 0.92 True Positive Rate (also, “recall”/”sensitivity”) and 0.3 False Positive Rate are demonstrated. Specificity is 0.84 and 0.16 false discovery rate.

(F) AUC ROC of a model trained with only a subset of top 14 most predictive features reveals a stable performance with the selected final features.

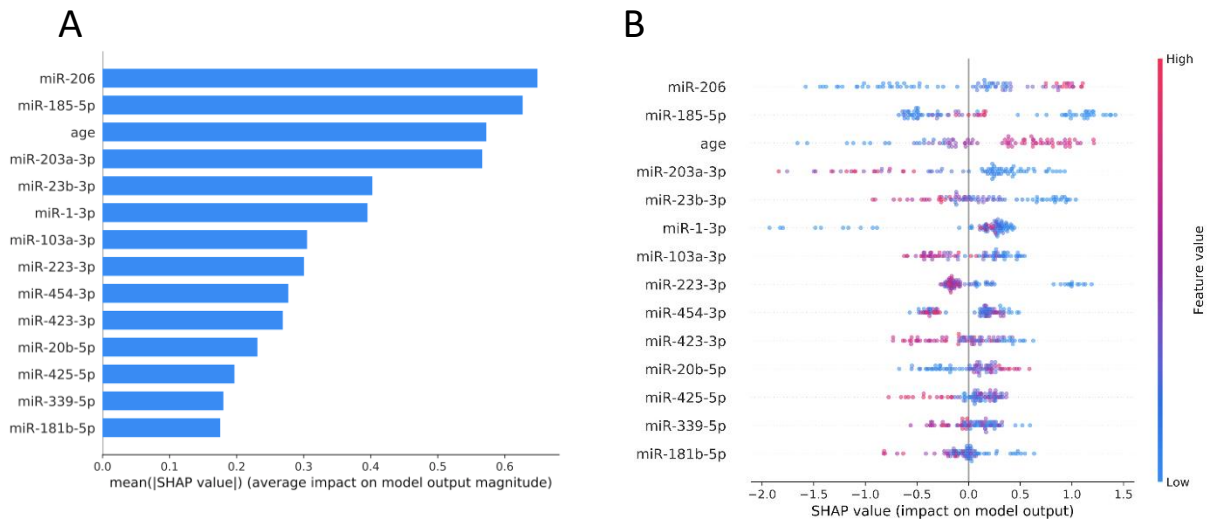


Figure 3: The contribution of individual miRNAs to the predictor of ALS diagnosis.
(A) Mean absolute SHapley Additive exPlanations (SHAP) values break down the impact of specific miRNAs on ALS disease predictor output in the held-out cohort. (B) Illustration of the relationship between the miRNA levels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

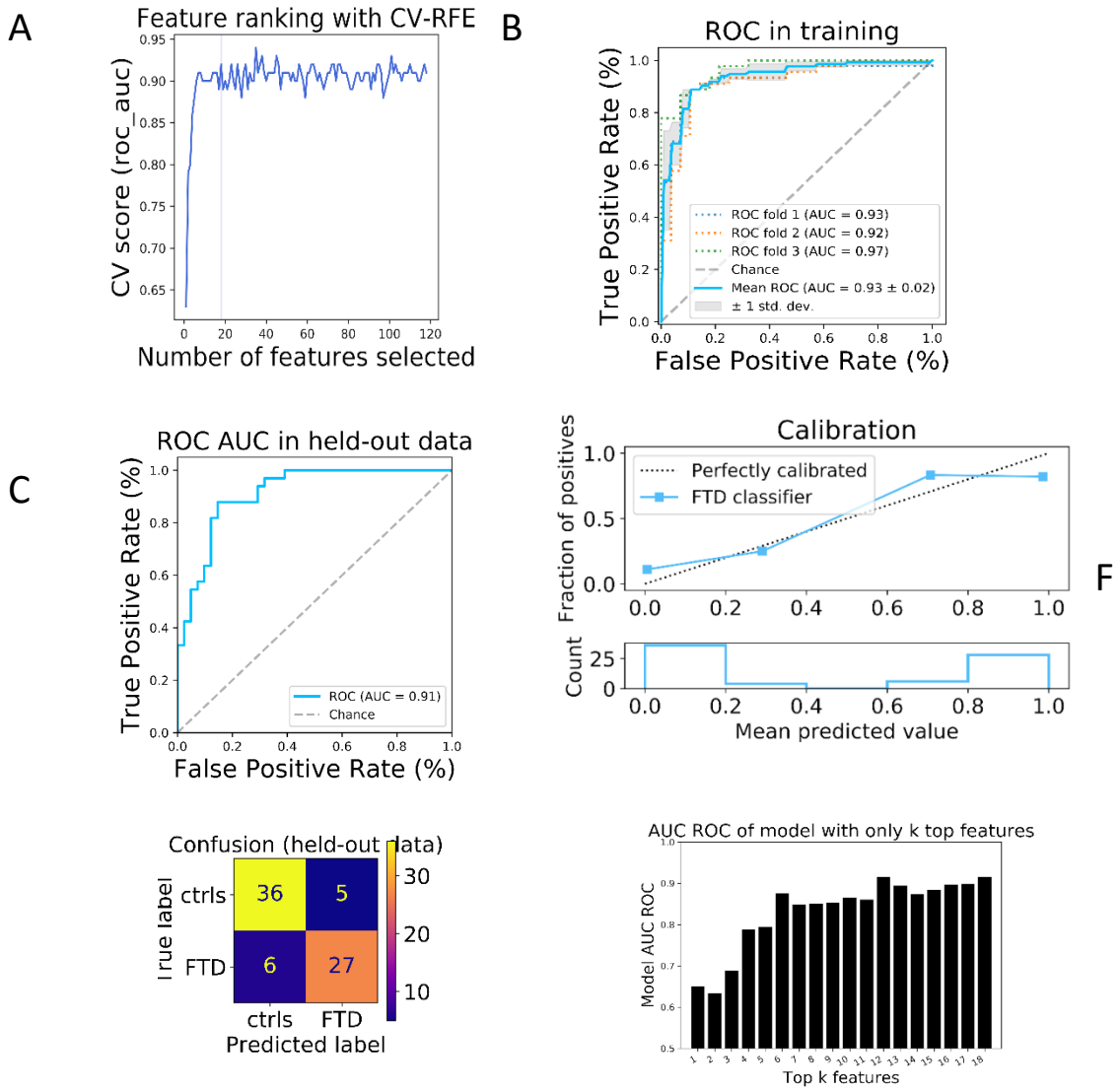


Figure 4: Diagnosis of FTD by a distinctive cell free miRNA signature.

(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 18 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 293 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 0.91 and average precision of 0.89.

(D) Reliability diagram, plotting truly observed vs. predicted average probability in five bins reveals a well 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 0.85 Negative predictive value (NPV), over 0.85 True Positive Rate (also, “recall”/”sensitivity”) and 0.15 False Positive Rate. Specificity is 0.875 and 0.18 false discovery rate.

(F) AUC ROC of a model trained with only a subset of top 18 most predictive features , reveals a stable performance with the selected final features.

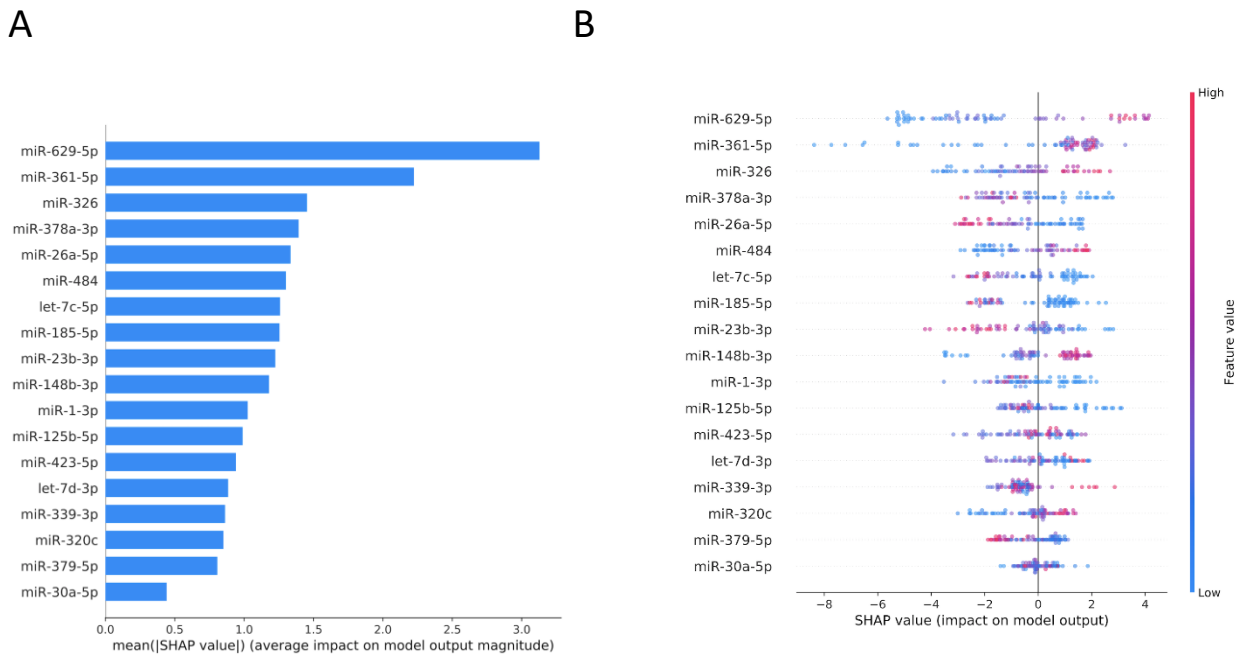
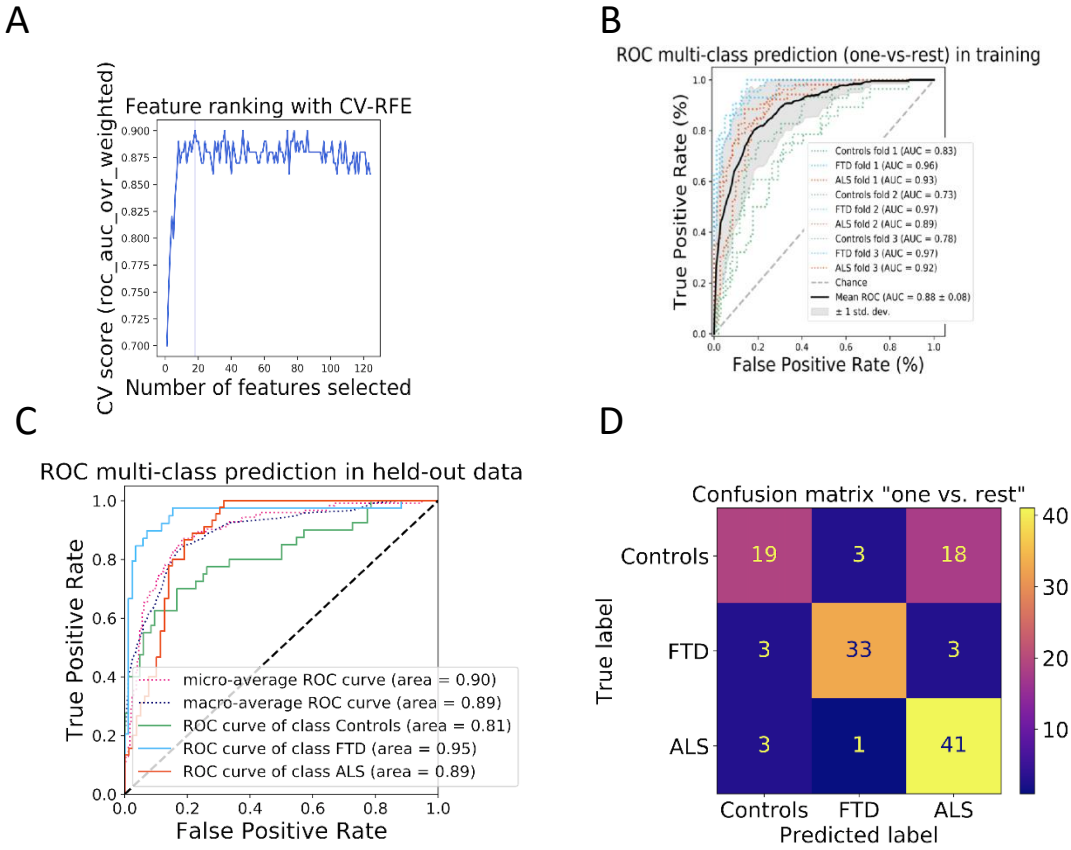


Figure 5: The contribution of individual miRNAs to the predictor of FTD diagnosis.
(A) Mean absolute SHapley Additive exPlanations (SHAP) values break down the impact of specific miRNAs on FTD disease predictor output in the held-out cohort. (B) Illustration of the relationship between the miRNA levels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



E

	Predicted label								
True label	Predicted / True	Non-controls	Controls	Predicted / True	Non-ALS	ALS	Predicted / True	Non-FTD	FTD
	Non-controls	78	6	Non-ALS	58	21	Non-FTD	81	4
	Controls	21	19	ALS	4	42	FTD	6	33

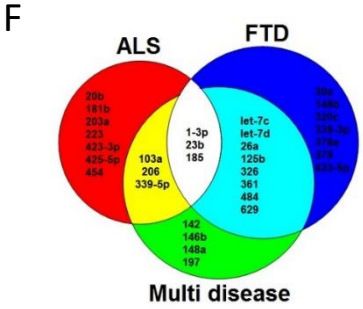


Figure 6: A multi-disease miRNA-based prediction model displays a superior diagnostic power over the binary models. A 3-fold cross-validated RFE with Extra Tree Classifier reveals that 18 features out of the original 134 approximate the same weighted mean one-vs-rest ROC AUC. (B) Fitting a Gradient Boosting multi-class classifier ("one-vs-all" strategy) on the training set. The multi-disease classifier achieved a weighted "one-vs-all" ROC AUC of 0.91, 0.86, 0.99 for each fold, respectively. Mean AUC of 88%±8% is obtained, indicating an overall good fit of the soft-classifier to all classes. Variance of the curve is shown in gray. (C) Model generalisability by ROC AUC per class in the held-out set (AUC of 0.81 for controls, 0.95 for FTD and 0.89 for ALS class), reassuring good accuracy for the soft-classifier over all three classes (micro-average AUC of 0.9 and macro-average AUC of 0.89), ensuring the model is not overfitting. (D) Confusion matrix (with a one-vs-rest strategy) in the held-out data and (E) class-specific confusion matrix (controls – green; ALS – blue; FTD – red). In the class-specific confusion matrix, the i-th confusion matrix considers class i as the positive class and all other classes j≠i as the negative class. Interestingly, 18/21 control subjects that were falsely classified were classified as "ALS". It can be observed that of the 21 false positives in 'ALS' class, 85.7% were actually healthy control. (F) a Venn diagram of miRNAs shared in the three classification models.

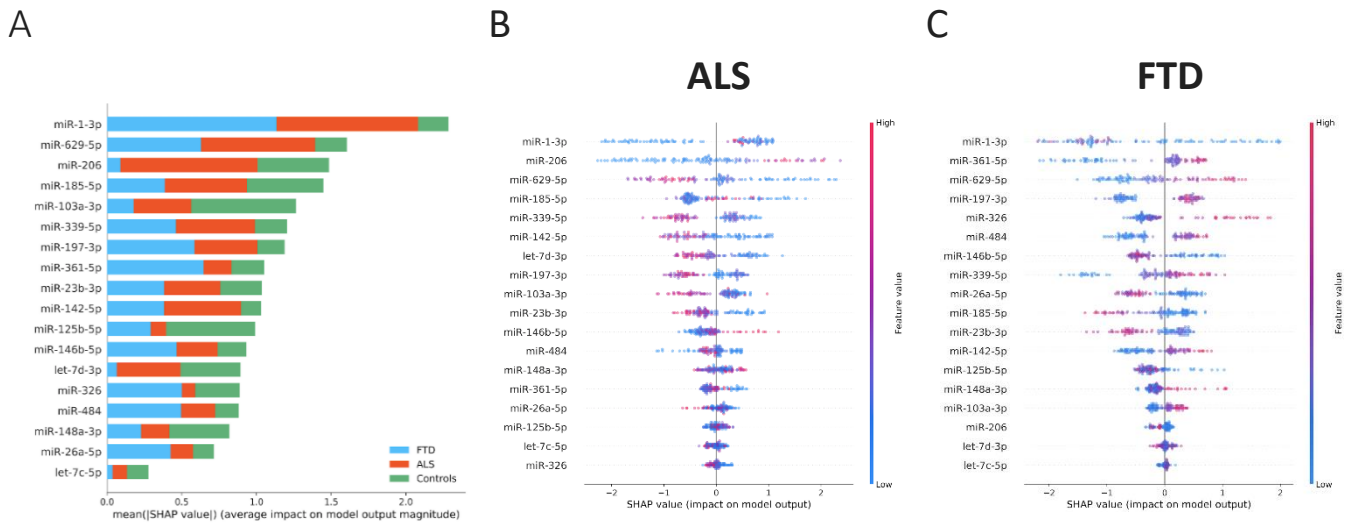


Figure 7: The contribution of individual features miRNAs shared and disease-specific miRNA predictors. ((B) SHAP analysis explains how features impact the multi-disease predictor output, namely, class probabilities. Average impact (mean absolute SHAP values) of miRNAs on model output for features, ordered according to their importance over all classes. For a specific class C-D), combining feature importance with feature effects to illustrate the relationship between the value of the feature and the impact on the prediction; 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-26a-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-26a-5p has the reversed impact with higher values reducing the risk of FTD.

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**	65.8	84	66.4	82	64.9±10.4, 95% CI [63.5, 66.4]
		M	63.4	122	63.9	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 [53.4, 58.0]
		M	46.6	29	57.0	18	
FTD cases (n=169)	65.58± 8.37, 95% CI [64.32, 66.84]	**F**	66.7	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	65.5	26	65.55± 7.2, 95% CI [63.49, 67.61]
		M	60.7	26	65.5	22	

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

Predictor	Single feature AUC
miR-206	0.74
miR-185-5p	0.678
miR-23b-3p	0.674
miR-1-3p	0.65
miR-454-3p	0.646
miR-103a-3p	0.641
miR-223-3p	0.637
miR-203a-3p	0.626
miR-339-5p	0.625
miR-181b-5p	0.562
miR-425-5p	0.521
miR-423-3p	0.515
miR-20b-5p	0.431

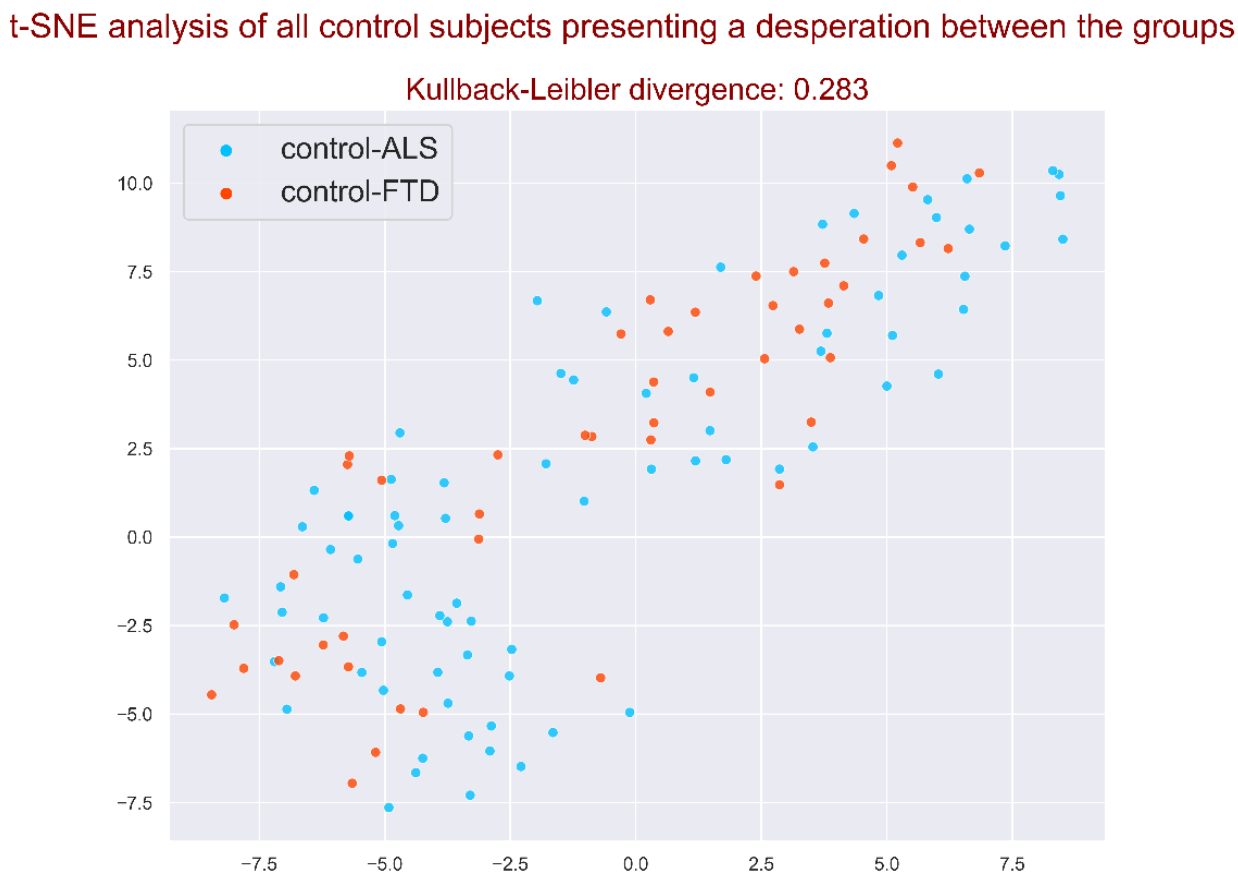
Supplementary Table 2. Predictive power of selected features, when used as a single predictor for ALS diagnosis on held-out data.

Predictor	Single feature AUC
miR-26a-5p	0.75
miR-185-5p	0.748
miR-629-5p	0.733
miR-320c	0.717
miR-125b-5p	0.694
miR-1-3p	0.67
miR-484	0.664
miR-379-5p	0.637
miR-23b-3p	0.633
let-7d-3p	0.623
miR-423-5p	0.606
miR-148b-3p	0.581
miR-361-5p	0.564
let-7c-5p	0.56
miR-326	0.536
miR-30a-5p	0.525
miR-378a-3p	0.524
miR-339-3p	0.505

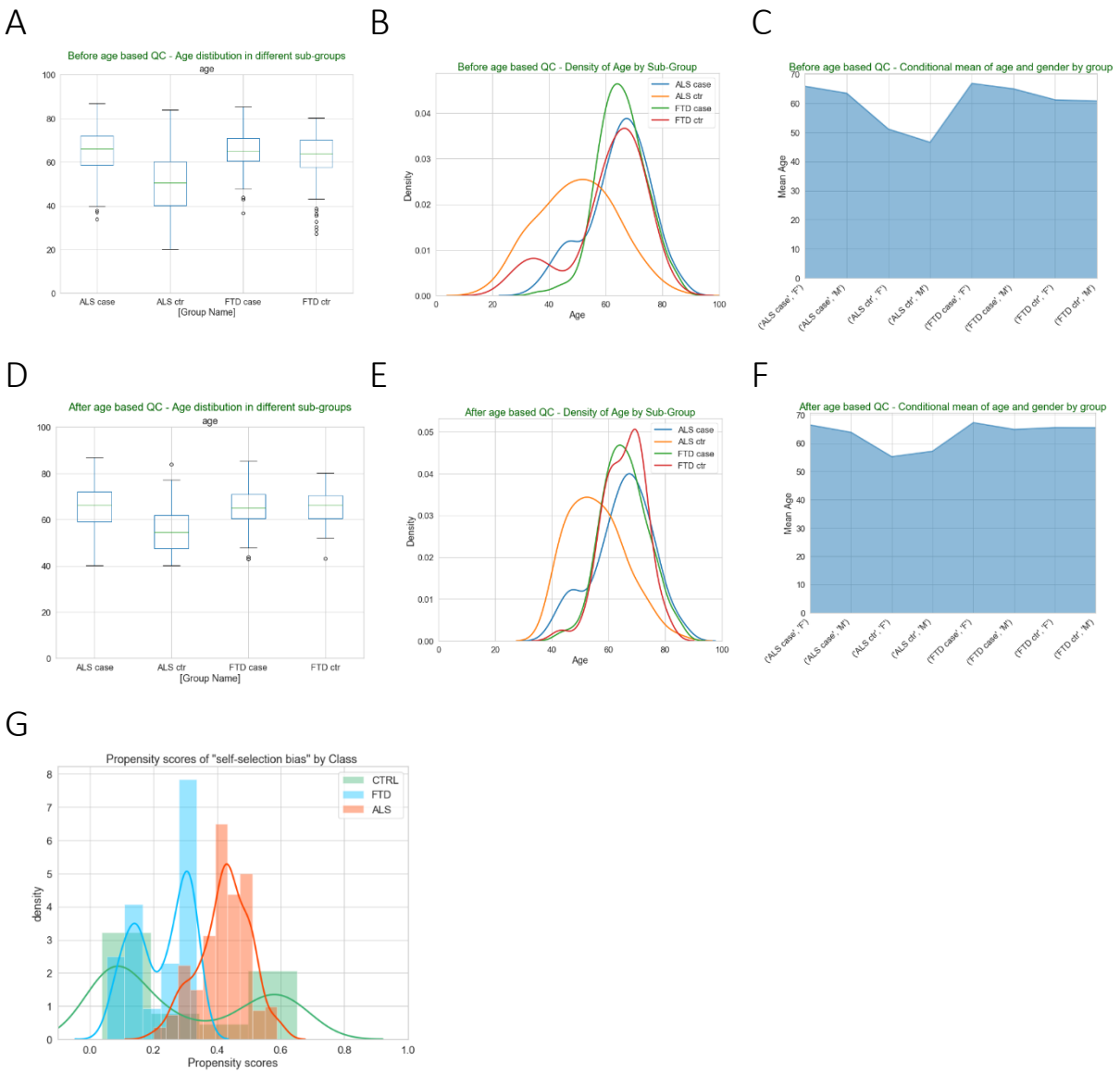
Supplementary Table 3. *Predictive power of selected features when used as a single predictor for FTD on held-out data. Among the most predictive features are miRNAs expressed in the brain, such as miR-26a-5p, miR-125b-5p and let-7c-5p.*

Predictor	Single feature AUC
miR-185-5p	0.748
miR-1-3p	0.718
miR-206	0.677
miR-125b-5p	0.667
miR-629-5p	0.643
miR-23b-3p	0.626
miR-339-5p	0.622
miR-26a-5p	0.618
let-7d-3p	0.616
let-7c-5p	0.608
miR-148a-3p	0.594
miR-484	0.589
miR-326	0.562
miR-146b-5p	0.553
miR-197-3p	0.544
miR-361-5p	0.534
miR-103a-3p	0.528
miR-142-5p	0.522

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. t-SNE analysis of all control subjects presenting a desperation between the control groups



Supplementary Figure 2: Meta-cohort adjusted for prognostic factors by inverse probability weighting using propensity score modeling. 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. (G) Splitting the subjects by quintiles of the propensity score and distribution of propensity scores by class shows an overlap in most propensity bins between comparison groups.