microRNA-based predictor for diagnosis of frontotemporal dementia

Running title: miRNA biomarkers for 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

Correspondence to: Eran Hornstein, Arnold R. Meyer Building, Room 314, Weizmann Institute of Science, Rehovot 76100, Israel, email: eran.hornstein@weizmann.ac.il.

Abstract word count: 150

Total word count (excluding title page, abstract, references and figure legend): 4870

No. of items (Figures, Tables, Supplementary): 5, 2, 6

References: 101

Key words: frontotemporal dementia; biomarker; microRNA; predictor; feature elimination.

Abstract

BACKGROUND: Frontotemporal dementia (FTD) is an early onset dementia that is diagnosed in ~20% of the progressive dementia cases. Heterogeneity in FTD 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.

METHODS: We initially studied a training cohort of 219 subjects (135 FTD and 84 non-neurodegenerative controls) and then validated the results in a cohort of 74 subjects (33 FTD and 41 controls).

RESULTS: Based on cell-free plasma miRNA profiling by next generation sequencing and machine learning approaches, we develop a nonlinear prediction model that accurately distinguishes FTD from non-neurodegenerative controls in ~90% of cases.

DISCUSSION: 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.

Background

Frontotemporal dementia (FTD) is a neurodegenerative disease that is characterized by frontal and temporal lobe atrophy, and typically manifests with behavioral or language deficits [1-4]. The key known genetic drivers of FTD include hexanucleotide repeat expansion in the first intron of the gene chromosome 9 open reading frame 72 (*C9ORF72*) [5,6], and mutations in the genes encoding for microtubule associated protein Tau **(*MAPT*)** [7], Valosin-containing protein (*VCP),* TANK-binding kinase 1 *(TBK1),* progranulin (*GRN*), charged multivesicular body protein 2B (*CHMP2B*) and the RNA-binding protein TAR DNA-binding protein 43 (*TDP-43*) [8-12] . Relatedly, inclusions of Tau and TDP-43 are hallmarks of FTD neuropathology[13],[14].

FTD can be difficult to diagnose, due to heterogeneity in clinical presentation [1]. Three main phenotypes of FTD are described: behavioral variant frontotemporal dementia (bvFTD), characterized by changes in social behavior 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 [1,15]. Related motor variants within the FTD spectrum are corticobasal syndrome (CBS) and progressive supranuclear palsy (PSP) [16]. FTD further resides on a genetic and a clinico-pathological continuum with amyotrophic lateral sclerosis (ALS) [17].

Brain imaging and several biofluid proteins have been proposed as biomarkers for FTD [18-48]. Among the proteins, noticeable are neurofilament light chain (NfL), TDP-43, and phospho-tau, amyloid beta and glial fibrillary acidic protein (GFAP). A recent study concluded that high NfL blood levels are indicative of the intensity of neurodegeneration or the extent of the degenerated axons in FTD [26].

microRNAs (miRNAs) are a class of small, non-coding RNAs, that can be quantified in biofluids in a massively parallel fashion, yielding fine-grained profiles [49]. miRNAs were suggested as biomarkers in neurodegeneration and psychiatry [50], and we too have shown their potential as cell-free biomarkers in neurodegeneration, focusing on the motor neuron diseases spinal muscular atrophy (SMA) and ALS [51,52]. Thus, low levels of miR-133 and miR-206 in the cerebrospinal fluid (CSF) of patients with SMA, predicted clinically meaningful response to nusinersen therapy [51], whereas low plasma levels of miR-181 predict longer survival and slower progression in patients with ALS [52].

Several studies suggested plasma or CSF miRNAs as diagnostic biomarkers for FTD[53-62]. However, initial studies of miRNAs in diagnosis of FTD were confounded by cohort size, sample

heterogeneity biases or pre-selection of candidate miRNAs. Furthermore, these studies did not address the potential non-linear relationships between miRNAs in developing a predictor.

In the current study, we profiled blood plasma miRNAs, and developed miRNA-based classifier for diagnosing FTD in a training cohort of 219 subjects, that was further validated in another cohort of 74 subjects. We implemented an ensemble machine learning approach, to address biomarker nonlinearity and were able to expose unrevealed disease-associated signals. We then confirmed that these signals are similar between clinical subtypes of FTD. The diagnostic power of the study roots from unbiased miRNA signature discovered by advanced machine learning, on a large and heterogeneous cohort, and validation in an independent held-out cohort, according to the TRIPOD guidelines [63]. Therefore, circulating miRNAs hold a fascinating potential as diagnostic biomarkers that may shorten diagnostic delay in FTD.

Methods

Participants and Sampling

Demographic data of study participants are detailed in Table 1.

FTD subjects and their respective controls were enrolled in the longitudinal FTD cohort studies at UCL. Study cohort included 169 FTD patients and 56 controls. Additional controls (N=102) were obtained from the ALS biomarker study (Ethics approval 09/H0703/27) in Queen Mary Hospital, totaling 158 non-neurodegenerative controls. Controls were typically spouses or relatives of patients and were not reported to have any clinical signs of ALS or FTD. Informed consent was obtained from all participants. Study inclusion period was from 2009 to 2018.

All FTD patients were seen at the National Hospital for Neurology and Neurosurgery, a national referral center for young and genetic dementias in the UK; the clinic has high diagnostic accuracy in cases that come to post mortem; for the majority of cases seen in the clinic CSF biomarkers of amyloid and Tau are used to distinguish FTD and AD whenever there is any question over accuracy of diagnosis.

Blood was collected by venipuncture in EDTA tubes, and plasma was recovered from the whole blood sample by centrifugation for 10 minutes at 3500 RPM at 4°C within 1 hour of sampling, and stored at −80°C until RNA extraction and subsequent small RNA next generation sequencing. Frozen plasma samples of FTD and controls from the UCL Biobanks were shipped to the Weizmann Institute of Science for molecular analysis.

Study Design

Based on power analysis calculations, we found that 150 controls and 150 cases are sufficient 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 [49]. 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) [64]. Human miRNAs, as defined by miRBase [65], were mapped using the GeneGlobe pipeline (https://geneglobe.qiagen.com/us/analyze). We defined "true positive" miRNAs and reduced the likelihood of considering "false positive" miRNAs, following previous works on miRNA biomarkers in neurodegeneration[58] and other conditions[66-68]. To this end, we included only miRNAs with an average UMI counts > 100 across all samples and with at least a single UMI across all samples, similar to our previous works[51,52]. Data were further corrected for the library preparation batch in order to reduce its potential bias, and normalized with DESeq2 package [69] under the assumption that miRNA counts followed negative binomial distribution

Constructing Cohorts and Restricting Age and Sex Biases

We observed a younger mean age in controls (53.8±14.5, 95% CI [51.5- 56.1]) than in FTD (65.6±8.4, 95% CI [64.4-67.0], Table S1). We reduced age-variance by excluding 34 participants younger than 40, which reduced differences in mean age across the remaining meta-cohort of 293 subjects by 45%. Thus, 168 out of 169 FTD patients and 125 out of 158 non-neurodegenerative control samples were included in the analysis (Table S1). In order to verify that a merged dataset of controls, collected in two different clinical centers, does not introduce biases, we employed the t-distributed stochastic neighbor embedding (t-SNE) algorithm and measured Kullback–Leibler divergence, the difference between probability distributions. In addition, a higher prevalence of males was observed among FTD patients (65%) than among the controls (35%). Therefore, sex and age variables were added to the prediction model as covariates, in addition to the selected 13 miRNA predictors.

Gradient Boosted Trees for the Development of Disease Binary Classifiers

The FTD-disease binary classifier was 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 [63] (https://www.tripod-

statement.org/). Gradient Boosting trees [70,71], 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. 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.
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 hyper-parameters: 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.

miRNA 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 3-fold cross validation. ExtraTreesClassifier algorithm is used in the cross-validated RFE procedure, with hyper-parameters: criterion="entropy", max_features=0.9, and n_estimators=20. 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 in order to rank features, and thus reduced the dimension of miRNA measurements needed for prediction by ~90% (13 miRNA features in a model in total). Additionally, the final model included age and sex as predictors (resulting in a of total 15 predictors).

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) [72], the current state of the art in Machine Learning explainability tools. 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.

Linear regression model

A linear classifier (Logistic Regression) was developed with a feature set of 95 miRNA predictors, that were differentially expressed (adjusted p-value ≤ 0.05) between FTD cases and controls. Each miRNA was binarized by its mean value. Training and validation were identical as in the gradient boosting classifier: 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 for Logistic Regression: L2 penalty, tol=1e-4, fit_intercept=True, solver='lbfgs', max_iter=500.

Results

We sought to determine the overall diagnostic capability of miRNA measurements in FTD. To this end, we based our study on analysis of plasma miRNA expression and the development of computational diagnostic models. The cohort included a total of 293 participants, enrolled between 2009 and 2018. Summary of participants' basic characteristics is shown in Table 1. Since the non-neurodegenerative controls were collected in two centers, we first verified that they could have been sampled from a single population and can therefore be considered a single cohort. We estimated the Kullback–Leibler divergence, a measure of the difference between two probability distributions. This analysis indicated only a small difference between the two sets in question (KL = 0.293, Figure S1), that is further visualized by t-distributed stochastic neighbor embedding (t-SNE) analysis of the cohorts.

Differential expression of miRNAs

Out of the >2000 miRNA species that were aligned to the human genome, only 132 fulfilled QC criteria of an average UMI count ≥ 100 across all samples and non-zero counts in all samples (see Methods). Next, we quantified the differential miRNAs that may distinguish FTD from controls. Ninety-five miRNAs were differentially expressed between plasma of patients with FTD and controls, with an adjusted p-value<0.05 (Figure 1A). Additional analysis of only subsets of C9ORF72-FTD cases (Figure 1B), FTD females (Figure 1C) or FTD males (Figure 1D) vs. relevant controls revealed that the miRNA signature was comparable between the full FTD cohort and subcategories (Table S2: Source Data Fig. 1). This was also true for subsets of FTD patients with predicted TDP or Tau pathology (Figure S2).

Development of machine learning classifier for the diagnosis of FTD

We established a diagnostic prediction model for FTD on a randomly selected training set of 135 FTD cases and 84 controls, comprising 75% of the total cohort (168 cases, 125 controls). For model validation, the remaining 25% of the data were held out as a replication cohort (33 FTD, 41 control samples).

The 132 miRNAs were tested as potential predictors, using an ensemble machine learning approach for ranking miRNAs predictive value in the diagnosis of FTD vs. individuals that did not suffer from neurodegeneration and were considered healthy. For selecting the most predictive features during model development, we used Recursive Feature Elimination (RFE), an efficient multivariate approach that iteratively removes miRNAs (features) and identifies those that

contribute the most to prediction accuracy. Cross validated RFE on the training set with Extra Tree classifiers, obtained a set of 13 miRNAs with highest feature importance (Figure 2A). Age and sex were added as predictors in the model, in addition to the selected 13 miRNAs features.

The prediction model presented a mean receiver operating characteristic area under the curve (ROC AUC) of 0.94 under internal cross validation on training dataset (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 0.88 (Figure 2C). Additional analysis reveals that the model is well calibrated and satisfactorily performs on held-out data (Figure 2D). At a defined probability threshold (0.6), the classifier exhibits a 0.81 precision (PPV) over 0.92 negative predictive value (NPV) with a 0.91 recall (sensitivity), a 0.83 specificity and a 0.08 false omission rate over 0.19 false discovery rate values (Figure 2E). Furthermore, a family of models trained with only a subset of the 15 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 have further compared the non-linear machine-learning strategy (with 13 miRNA features selected by multivariate RFE, Figure 2) to a logistic regression model with univariate feature selction (Figure S3). A subset of 95 miRNAs out of the 132 passing QC, were differentially expressed between FTD and controls (passing a threshold of $p<0.05$ after correction for multiple hypotheses) and therefore were used in the logostic regression as features. Similarly, age and sex were also included as predictors. The linear model with 95 differentially-expressed miRNAs, age and sex was inferior to the gradient boosting classifier model, particularly by sensitivity (0.73 vs 0.91) and negative predictive value (0.8 vs 0.92) (Figure S3). Moreover, the non-linear approach showed better robustness in the different training folds and also outperforms the linear model in utilizing a significantly smaller number of miRNA features (15 vs. 97), to obtain its accuracy.

We then sought to better understand the machine-learning-based non-linear 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). Extreme SHAP values inform that the model predicts a more likely FTD (positive values) or healthy control (negative values). The key predictors revealed by SHAP, which contribute the most to calling FTD vs controls, are multiple sclerosis-associated miR-629 [73,74], brain enriched miR-125b [75] and the astrocyte-derived exosomal miR-361 [76]. We further tested each individual miRNA as a single predictor of FTD

diagnosis. A model with only miR-423-5p or miR-125b-5p presented the highest AUC values of ~0.69 (Table 2). Boxplots depicting the underlying distribution of each of the 13 miRNA predictors, along with scatter plots of the expression values of individual subjects in these miRNAs, confirm the non-linear relationship between miRNA expression and the phenotype, for some miRNAs (e.g., miR-107 and miR-26a), while other miRNAs can be divided to low/high values predominated by either controls or FTD cases (e.g., miR-629-5p, Figure 4). These observations emphasize the power of a model, which is able to take into account non-linear relationship between features (miRNAs).

Significant pairwise differences were noted in miR-107 levels in both bvFTD and SD vs other FTD subtypes (primary progressive aphasia, corticobasal syndrome, progressive supranuclear palsy and FTD with motor neuron disease. ANOVA: p=0.009, bvFTD vs other FTD subtypes: p=0.01, SD vs other FTD subtypes: p=0.02, Figure S4C). In addition, significant pairwise differences in miR-26a levels were found in both bvFTD (p=0.04) and PNFA (p=0.02) vs other FTD subtypes (ANOVA: p=0.024, bvFTD vs other FTD subtypes: p=0.04, PNFA vs other FTD subtypes: p=0.02, Figure S4F).

None of the other miRNAs showed differences between clinical subtypes (Figure S4), between different mutation carriers, or between mutations carriers and patients with no known FTD mutations (Figure S5). In conclusion, a non-linear signature based on only 13 miRNAs is able to call the diagnosis of FTD with high accuracy and generalizability, for all FTD subtypes and independent of the underlying genetic background.

Finally, we tested the performance of plasma miRNA classifiers that were reported in previous works [56,58], relative to the classifier we have reported here. We replicated the logistic regression model (with L2 regularization[77,78]) from the study of Kmetzsch et al. [58] on our data. Taking miR-34a-5p, miR-345-5p, miR-200c-3p and miR-10a-3p as features, displayed inferiority to our model. Moreover, these miRNAs were not differentially expressed between cases and controls in our study. Next, we replicated the "microRNA pair" approach from the study of Sheinerman et al. [56]: miR-335/let-7e, miR-99b/let-7e and miR-9-3p/miR-181a and assessed the ROC curves, showing they are inferior to our non-linear classifier in distinguishing between FTD and controls (Figure 5). Thus, when tested in comparison to reported miRNA classifiers from the literature, our panels of miRNA and non-linear classifier perform better, with AUC values of 0.9.

Discussion

The need to facilitate the diagnosis of FTD in the face of clinical heterogeneity raises the hope for new effective biomarkers. Circulating miRNAs hold a fascinating potential as diagnostic biomarkers that were not fully explored, including for brain disorders [50]. We recently demonstrated the power of miRNAs in prognosis and prediction of clinical response to therapy in motor neuron diseases [51,52]. In the present work, we sought to discover diagnostic biomarkers for FTD. By using an unbiased, next generation sequencing approach and advanced computation in a discovery cohort of 219 subjects and 74 additional subjects for a replication study, we overcame limitations of past works in developing biomarkers for FTD. A recursive non-linear approach was utilized to find the smallest set of miRNA features that obtain the highest accuracy of the prediction model, leading to a signature of 13 miRNAs only. Thus, our miRNA-based binary classifier is more accurate and robust than published miRNA-based predictors [56,58].

The use of gradient boosting trees, an ensemble learning approach, allows discovering nonlinear relationships between miRNAs and disease status that gained affirmation by cross validation in the training dataset. Furthermore, our model is externally validated on held-out data, which was not used during feature selection and model development, according to the TRIPOD guidelines [63]. SHAP analysis [72] further unfolds the relative contribution of individual features to the predictive model. Of note, the performance of the ensemble learning was superior to that of a logistic regression model, and enabled the use of only 15 features, instead of 97, for prediction.

According to the human miRNA tissue atlas [75], most of the selected miRNAs are CNS-enriched (let-7c-5p, miR-26a-5p, miR-107, miR-125b-5p), suggesting that they might be directly involved in the disease state of the CNS. 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 [79-85]. In mice, miR-326 inhibited Tau phosphorylation[80], a hallmark of FTD as well as AD, supranuclear palsy (PSP), conticobasal syndrome (CBD) and chronic traumatic encephalopathy (CTE)[86]. Brain-enriched miR-107 was also implicated in AD [87-92]. Changes in blood levels of miR-326, miR-26a, and miR-629 are associated with multiple sclerosis (MS), a condition characterized by demyelination [93-100]. Moreover, serum miR-629 was negatively correlated with MS patient brain volume and lesion severity, respectively [101]. Thus, some of the miRNA predictors proposed here are associated directly or indirectly with other degenerative brain diseases.

Conclusion

Strengths and Limitations

We have found specific molecular miRNA patterns that can contribute to the diagnosis of FTD. Therefore, the work encourages testing if circulating miRNAs biomarkers can establish a cost-effective screening approach to increase speed or precision in the diagnosis of suspected FTD. Early-stage diagnosis may be useful towards design of 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. Finally, while ongoing biomarker studies in FTD highlight protein markers such as NfL and Tau, combined protein-RNA markers may present increased accuracy, as we previously showed[52].

We would like to put forward a few notable limitations of our study: First, patients in our cohorts were most likely recruited in different phases of the disease, which results in significant phenotypic heterogeneity. At the same time, since we had no record of disease severity, we could not stratify patients by their disease phase and depict stage-dependent changes in miRNAs. Additionally, we did not find any miRNAs, which can differentiate between FTD-Tau and FTD-TDP, which might result from small, underpowered subset cohorts or could highlight a wider role for these novel biomarkers within dementia syndromes. We hope that in the future, larger cohorts can be used to reveal such differences. Finally, a long route is expected from this initial study and until miRNA can be used in personalized diagnosis. These steps shall include prospective studies, quantitative calibration of absolute miRNA concentrations and simple bed-side methods for quantification of miRNAs.

Unanswered questions

It remains to be determined why our findings are discrepant with conclusions of past studies [53-60,62]. For studies in other bodily fluids the answer might be trivially-related to the different biofluid composition[53,57,60]. Furthermore, we emphasize the progress we presented in our study in terms of power (larger cohort) and the unbiased analysis by next generation microRNA sequencing, that contrasts past biased choices of miRNA candidates. A second unanswered question is why defined nonlinear patterns are strongly predictive of disease states. Future works should address the differential diagnosis between FTD and other dementias, such as Alzheimer's disease. Lastly, protein based markers, such as neurofilaments, lack specificity to particular neurodegenerative diseases. However, it is plausible that miRNAs might demonstrate a disease-specific pattern in the circulation. Future studies with large AD and FTD cohorts might address this hypothesis.

Recommendations

In our study we took care to keep a held-out cohort for external validation. To substantiate our work and towards clinical diagnostic usage, it is warranted to validate the predictor by testing it on an independent cohort of different ethnicity and create means for quantification of absolute miRNA concentrations. In addition, miRNA levels should be compared in follow up works, and combined with other experimental circulating markers of neurodegeneration such as neurofilaments. In addition, miRNAs could be explored also as prognostic markers in FTD and in predicting disease severity.

Availability of supporting data

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Appendix

IDs of 34 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'.

Acknowledgements

We thank Vittoria Lombardi (UCL) for technical assistance. We acknowledge patients with FTD, 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 and of the Andi and Larry Wolfe Center for Neuroimmunology and Neuromodulation.

Funding

Research at the Hornstein laboratory is supported by the CReATe consortium and ALSA (program: "Prognostic Value of miRNAs in Biofluids From ALS Patients"), the RADALA Foundation; AFM Telethon (20576); the Weizmann Center for Research on Neurodegeneration

at Weizmann Institute of Science; the Minerva Foundation, with funding from the Federal German Ministry for Education and Research; the ISF Legacy Heritage Fund (828/17); the Israel Science Foundation (135/16, 3497/21, 424/22, 425/22); United States - Israel Binational Science Foundation (#2021181); A research grant from the Anita James Rosen Foundation; Target ALS (118945); the Thierry Latran Foundation for ALS Research; the European Research Council under the European Union's Seventh Framework Program ([FP7/2007–2013]/ERC grant agreement number 617351); ERA-Net for Research Programs on Rare Diseases (eRARE FP7) via the Israel Ministry of Health; Dr Sydney Brenner and friends; Edward and Janie Moravitz; A. Alfred Taubman through IsrALS; Yeda-Sela; Yeda-CEO; the Israel Ministry of Trade and Industry; the Y. Leon Benoziyo Institute for Molecular Medicine; the Nella and Leon Benoziyo Center for Neurological Disease; the Kekst Family Institute for Medical Genetics; the David and Fela Shapell Family Center for Genetic Disorders Research; the Crown Human Genome Center; the Nathan, Shirley, Philip, and Charlene Vener New Scientist Fund; the Julius and Ray Charlestein Foundation; the Fraida Foundation; the Wolfson Family Charitable Trust; the Adelis Foundation; Merck (UK); M. Halphen; the estates of F. Sherr, L. Asseof, and L. Fulop; the Goldhirsh-Yellin Foundation; the Redhill Foundation–Sam and Jean Rothberg Charitable Trust; Dr. Dvora and Haim Teitelbaum Endowment Fund; A research grant from the Anita James Rosen Foundation. The ALS-Therapy Alliance, Motor Neuron Disease Association (UK); this work was supported by the Motor Neuron Disease Association (MNDA) 839-791.

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 a Medical Research Council Senior Clinical Fellowship, an MRC/MND LEW Fellowship, the NIHR UCLH BRC and the Lady Edith Wolfson Fellowship scheme (MR/M008606/1 and MR/S006508/1).

NSY was supported by the Israeli Council for Higher Education (CHE) via the Weizmann Data Science Research Center, by a research grant from the Estate of Tully and Michele Plesser and by Maccabim Foundation. I.M. was supported by Teva Pharmaceutical Industries as part of the Israeli National Network of Excellence in Neuroscience (fellowship no. 117941).

Authors' contribution

I.M., P.F., J.D.R. and E.H. conceived research.

I.M., N.S.Y. and E.H. analyzed the data.

A.M. and J.D.R. established cohort, gained ethical approval and collected human samples for research.

Y.B., C.H. and I.S. assisted research.

I.M., N.S.Y., P.F., A.M., J.D.R. and E.H. wrote the manuscript, with comments and final approval by all other authors.

J.D.W. provided resources for research and input in research development.

A.M. and J.D.R. are corresponding authors for cohorts and clinical data. P.F. and E.H. are corresponding authors for all other facets of the work.

All authors read and approved the final manuscript.

Declarations of interest

None.

References

1. McKhann GM, Albert MS, Grossman M, et al. Clinical and pathological diagnosis of frontotemporal dementia: report of the Work Group on Frontotemporal Dementia and Pick's Disease. *Arch Neurol*. Nov 2001;58(11):1803-9. doi:10.1001/archneur.58.11.1803
2. Seelaar H, Rohrer JD, Pijnenburg YA, Fox NC, van Swieten JC. Clinical, genetic and pathological heterogeneity of frontotemporal dementia: a review. *J Neurol Neurosurg Psychiatry*. May 2011;82(5):476-86. doi:10.1136/jnnp.2010.212225
3. Convery R, Mead S, Rohrer JD. Review: Clinical, genetic and neuroimaging features of frontotemporal dementia. *Neuropathol Appl Neurobiol*. Feb 2019;45(1):6-18. doi:10.1111/nan.12535
4. Deleon J, Miller BL. Frontotemporal dementia. *Handb Clin Neurol*. 2018;148:409-430. doi:10.1016/B978-0-444-64076-5.00027-2
5. DeJesus-Hernandez M, Mackenzie IR, Boeve BF, et al. Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron*. Oct 20 2011;72(2):245-56. doi:10.1016/j.neuron.2011.09.011
6. Renton AE, Majounie E, Waite A, et al. A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron*. Oct 20 2011;72(2):257-68. doi:10.1016/j.neuron.2011.09.010
7. Hutton M, Lendon CL, Rizzu P, et al. Association of missense and 5'-splice-site mutations in tau with the inherited dementia FTDP-17. *Nature*. Jun 18 1998;393(6686):702-5. doi:10.1038/31508
8. Skibinski G, Parkinson NJ, Brown JM, et al. Mutations in the endosomal ESCRTIII-complex subunit CHMP2B in frontotemporal dementia. *Nat Genet*. Aug 2005;37(8):806-8. doi:10.1038/ng1609
9. Baker M, Mackenzie IR, Pickering-Brown SM, et al. Mutations in progranulin cause tau-negative frontotemporal dementia linked to chromosome 17. *Nature*. Aug 24 2006;442(7105):916-9. doi:10.1038/nature05016
10. Borroni B, Bonvicini C, Alberici A, et al. Mutation within TARDBP leads to frontotemporal dementia without motor neuron disease. *Hum Mutat*. Nov 2009;30(11):E974-83. doi:10.1002/humu.21100
11. Pottier C, Bieniek KF, Finch N, et al. Whole-genome sequencing reveals important role for TBK1 and OPTN mutations in frontotemporal lobar degeneration without motor neuron disease. *Acta Neuropathol*. Jul 2015;130(1):77-92. doi:10.1007/s00401-015-1436-x
12. Watts GD, Wymer J, Kovach MJ, et al. Inclusion body myopathy associated with Paget disease of bone and frontotemporal dementia is caused by mutant valosin-containing protein. *Nat Genet*. Apr 2004;36(4):377-81. doi:10.1038/ng1332
13. Lippa CF, Zhukareva V, Kawarai T, et al. Frontotemporal dementia with novel tau pathology and a Glu342Val tau mutation. *Ann Neurol*. Dec 2000;48(6):850-8.
14. Neumann M, Sampathu DM, Kwong LK, et al. Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science*. Oct 6 2006;314(5796):130-3. doi:10.1126/science.1134108
15. Snowden JS, Neary D, Mann DM. Frontotemporal dementia. *Br J Psychiatry*. Feb 2002;180:140-3.
16. Olney NT, Spina S, Miller BL. Frontotemporal Dementia. *Neurol Clin*. May 2017;35(2):339-374. doi:10.1016/j.ncl.2017.01.008
17. Strong MJ, Abrahams S, Goldstein LH, et al. Amyotrophic lateral sclerosis - frontotemporal spectrum disorder (ALS-FTSD): Revised diagnostic criteria. *Amyotroph Lateral Scler Frontotemporal Degener*. May 2017;18(3-4):153-174. doi:10.1080/21678421.2016.1267768

18. Teunissen CE, Elias N, Koel-Simmelink MJ, et al. Novel diagnostic cerebrospinal fluid biomarkers for pathologic subtypes of frontotemporal dementia identified by proteomics. *Alzheimers Dement (Amst)*. 2016;2:86-94. doi:10.1016/j.dadm.2015.12.004
19. Sheikh-Bahaei N, Sajjadi SA, Pierce AL. Current role for biomarkers in clinical diagnosis of Alzheimer disease and frontotemporal dementia. *Current treatment options in neurology*. 2017;19(12):46.
20. Meeter LH, Kaat LD, Rohrer JD, Van Swieten JC. Imaging and fluid biomarkers in frontotemporal dementia. *Nature Reviews Neurology*. 2017;13(7):406.
21. Meeter LHH, Vijverberg EG, Del Campo M, et al. Clinical value of neurofilament and phospho-tau/tau ratio in the frontotemporal dementia spectrum. *Neurology*. Apr 3 2018;90(14):e1231-e1239. doi:10.1212/WNL.0000000000005261
22. Feneberg E, Gray E, Ansorge O, Talbot K, Turner MR. Towards a TDP-43-Based Biomarker for ALS and FTLD. *Mol Neurobiol*. Oct 2018;55(10):7789-7801. doi:10.1007/s12035-018-0947-6
23. Forgrave LM, Ma M, Best JR, 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. *Alzheimers Dement (Amst)*. Dec 2019;11:730-743. doi:10.1016/j.dadm.2019.08.009
24. Delaby C, Alcolea D, Carmona-Iragui M, et al. Differential levels of Neurofilament Light protein in cerebrospinal fluid in patients with a wide range of neurodegenerative disorders. *Sci Rep*. Jun 8 2020;10(1):9161. doi:10.1038/s41598-020-66090-x
25. van der Ende EL, Xiao M, Xu D, et al. Neuronal pentraxin 2: a synapse-derived CSF biomarker in genetic frontotemporal dementia. *J Neurol Neurosurg Psychiatry*. Jun 2020;91(6):612-621. doi:10.1136/jnnp-2019-322493
26. Ashton NJ, Janelidze S, Al Khleifat A, et al. A multicentre validation study of the diagnostic value of plasma neurofilament light. *Nat Commun*. Jun 7 2021;12(1):3400. doi:10.1038/s41467-021-23620-z
27. Katisko K, Cajanus A, Huber N, et al. GFAP as a biomarker in frontotemporal dementia and primary psychiatric disorders: diagnostic and prognostic performance. *J Neurol Neurosurg Psychiatry*. Dec 2021;92(12):1305-1312. doi:10.1136/jnnp-2021-326487
28. Verde F, Otto M, Silani V. Neurofilament Light Chain as Biomarker for Amyotrophic Lateral Sclerosis and Frontotemporal Dementia. *Front Neurosci*. 2021;15:679199. doi:10.3389/fnins.2021.679199
29. Wilson KM, Katona E, Glaria I, et al. Development of a sensitive trial-ready poly(GP) CSF biomarker assay for C9orf72-associated frontotemporal dementia and amyotrophic lateral sclerosis. *J Neurol Neurosurg Psychiatry*. Jul 2022;93(7):761-771. doi:10.1136/jnnp-2021-328710
30. Katisko K, Cajanus A, Jaaskelainen O, et al. Serum neurofilament light chain is a discriminative biomarker between frontotemporal lobar degeneration and primary psychiatric disorders. *J Neurol*. Jan 2020;267(1):162-167. doi:10.1007/s00415-019-09567-8
31. Hansson O, Santillo AF, Meeter LH, et al. CSF placental growth factor - a novel candidate biomarker of frontotemporal dementia. *Ann Clin Transl Neurol*. May 2019;6(5):863-872. doi:10.1002/acn3.763
32. Al Shweiki MR, Steinacker P, Oeckl P, et al. Neurofilament light chain as a blood biomarker to differentiate psychiatric disorders from behavioural variant frontotemporal dementia. *J Psychiatr Res*. Jun 2019;113:137-140. doi:10.1016/j.jpsychires.2019.03.019
33. Borroni B, Benussi A, Archetti S, et al. Csf p-tau181/tau ratio as biomarker for TDP pathology in frontotemporal dementia. *Amyotroph Lateral Scler Frontotemporal Degener*. Mar 2015;16(1-2):86-91. doi:10.3109/21678421.2014.971812
34. Bourbouli M, Rentzos M, Bougea A, et al. Cerebrospinal Fluid TAR DNA-Binding Protein 43 Combined with Tau Proteins as a Candidate Biomarker for Amyotrophic Lateral Sclerosis and

Frontotemporal Dementia Spectrum Disorders. *Dement Geriatr Cogn Disord*. 2017;44(3-4):144-152. doi:10.1159/000478979

35. Hu WT, Watts K, Grossman M, et al. Reduced CSF p-Tau181 to Tau ratio is a biomarker for FTLD-TDP. *Neurology*. Nov 26 2013;81(22):1945-52. doi:10.1212/01.wnl.0000436625.63650.27

36. Jiskoot LC, Panman JL, Meeter LH, et al. Longitudinal multimodal MRI as prognostic and diagnostic biomarker in presymptomatic familial frontotemporal dementia. *Brain*. Jan 1 2019;142(1):193-208. doi:10.1093/brain/awy288

37. McCarthy J, Collins DL, Ducharme S. Morphometric MRI as a diagnostic biomarker of frontotemporal dementia: A systematic review to determine clinical applicability. *Neuroimage Clin*. 2018;20:685-696. doi:10.1016/j.nicl.2018.08.028

38. Niikado M, Chrem-Mendez P, Itzcovich T, et al. Evaluation of Cerebrospinal Fluid Neurofilament Light Chain as a Routine Biomarker in a Memory Clinic. *J Gerontol A Biol Sci Med Sci*. Mar 14 2019;74(4):442-445. doi:10.1093/gerona/gly179

39. Benussi A, Karikari TK, Ashton N, et al. Diagnostic and prognostic value of serum NfL and p-Tau(181) in frontotemporal lobar degeneration. *J Neurol Neurosurg Psychiatry*. Sep 2020;91(9):960-967. doi:10.1136/jnnp-2020-323487

40. Das S, Dewit N, Jacobs D, et al. A Novel Neurofilament Light Chain ELISA Validated in Patients with Alzheimer's Disease, Frontotemporal Dementia, and Subjective Cognitive Decline, and the Evaluation of Candidate Proteins for Immunoassay Calibration. *Int J Mol Sci*. Jun 29 2022;23(13)doi:10.3390/ijms23137221

41. Bellini S, Saraceno C, Benussi L, et al. Plasma Small Extracellular Vesicles with Complement Alterations in GRN/C9orf72 and Sporadic Frontotemporal Lobar Degeneration. *Cells*. Jan 30 2022;11(3)doi:10.3390/cells11030488

42. Chouliaras L, Thomas A, Malpetti M, et al. Differential levels of plasma biomarkers of neurodegeneration in Lewy body dementia, Alzheimer's disease, frontotemporal dementia and progressive supranuclear palsy. *J Neurol Neurosurg Psychiatry*. Jun 2022;93(6):651-658. doi:10.1136/jnnp-2021-327788

43. Eratne D, Loi SM, Li QX, et al. Cerebrospinal fluid neurofilament light chain differentiates primary psychiatric disorders from rapidly progressive, Alzheimer's disease and frontotemporal disorders in clinical settings. *Alzheimers Dement*. Feb 1 2022;doi:10.1002/alz.12549

44. Krishnan G, Raitcheva D, Bartlett D, et al. Poly(GR) and poly(GA) in cerebrospinal fluid as potential biomarkers for C9ORF72-ALS/FTD. *Nat Commun*. May 19 2022;13(1):2799. doi:10.1038/s41467-022-30387-4

45. Oeckl P, Anderl-Straub S, Von Arnim CAF, et al. Serum GFAP differentiates Alzheimer's disease from frontotemporal dementia and predicts MCI-to-dementia conversion. *J Neurol Neurosurg Psychiatry*. Apr 27 2022;doi:10.1136/jnnp-2021-328547

46. Silva-Spinola A, Lima M, Leitao MJ, et al. Serum neurofilament light chain as a surrogate of cognitive decline in sporadic and familial frontotemporal dementia. *Eur J Neurol*. Jan 2022;29(1):36-46. doi:10.1111/ene.15058

47. Thijssen EH, Verberk IMW, Kindermans J, et al. Differential diagnostic performance of a panel of plasma biomarkers for different types of dementia. *Alzheimers Dement (Amst)*. 2022;14(1):e12285. doi:10.1002/dad2.12285

48. Meeter LH, Dopper EG, Jiskoot LC, et al. Neurofilament light chain: a biomarker for genetic frontotemporal dementia. *Ann Clin Transl Neurol*. Aug 2016;3(8):623-36. doi:10.1002/acn3.325

49. Coenen-Stass AML, Magen I, Brooks T, et al. Evaluation of methodologies for microRNA biomarker detection by next generation sequencing. *RNA Biol*. 2018;15(8):1133-1145. doi:10.1080/15476286.2018.1514236

50. van den Berg MMJ, Krauskopf J, Ramaekers JG, Kleinjans JCS, Prickaerts J, Briede JJ. Circulating microRNAs as potential biomarkers for psychiatric and neurodegenerative disorders. *Prog Neurobiol*. Feb 2020;185:101732. doi:10.1016/j.pneurobio.2019.101732

51. Magen I, Aharoni S, Yacovzada NS, et al. Muscle microRNAs in the cerebrospinal fluid predict clinical response to nusinersen therapy in type II and type III spinal muscular atrophy patients. *Eur J Neurol*. Aug 2022;29(8):2420-2430. doi:10.1111/ene.15382
52. Magen I, Yacovzada NS, Yanowski E, et al. Circulating miR-181 is a prognostic biomarker for amyotrophic lateral sclerosis. *Nat Neurosci*. Nov 2021;24(11):1534-1541. doi:10.1038/s41593-021-00936-z
53. Denk J, Oberhauser F, Kornhuber J, 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. doi:10.1371/journal.pone.0197329
54. Grasso M, Piscopo P, Talarico G, et al. Plasma microRNA profiling distinguishes patients with frontotemporal dementia from healthy subjects. *Neurobiol Aging*. Feb 5 2019;doi:10.1016/j.neurobiolaging.2019.01.024
55. Piscopo P, Grasso M, Puopolo M, et al. Circulating miR-127-3p as a Potential Biomarker for Differential Diagnosis in Frontotemporal Dementia. *J Alzheimers Dis*. 2018;65(2):455-464. doi:10.3233/JAD-180364
56. Sheinerman KS, Toledo JB, Tsivinsky VG, et al. Circulating brain-enriched microRNAs as novel biomarkers for detection and differentiation of neurodegenerative diseases. *Alzheimers Res Ther*. Nov 9 2017;9(1):89. doi:10.1186/s13195-017-0316-0
57. Schneider R, McKeever P, Kim T, et al. Downregulation of exosomal miR-204-5p and miR-632 as a biomarker for FTD: a GENFI study. *J Neurol Neurosurg Psychiatry*. Aug 2018;89(8):851-858. doi:10.1136/jnnp-2017-317492
58. Kmetzsch V, Anquetil V, Saracino D, et al. 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):485-493.
59. Sproviero D, Gagliardi S, Zucca S, et al. Different miRNA Profiles in Plasma Derived Small and Large Extracellular Vesicles from Patients with Neurodegenerative Diseases. *Int J Mol Sci*. Mar 8 2021;22(5)doi:10.3390/ijms22052737
60. Tan YJ, Wong BYX, Vaidyanathan R, et al. Altered Cerebrospinal Fluid Exosomal microRNA Levels in Young-Onset Alzheimer's Disease and Frontotemporal Dementia. *J Alzheimers Dis Rep*. 2021;5(1):805-813. doi:10.3233/ADR-210311
61. Muller M, Kuiperij HB, Versleijen AA, et al. Validation of microRNAs in Cerebrospinal Fluid as Biomarkers for Different Forms of Dementia in a Multicenter Study. *J Alzheimers Dis*. Apr 16 2016;52(4):1321-33. doi:10.3233/JAD-160038
62. Pounders J, Hill EJ, Hooper D, et al. MicroRNA expression within neuronal-derived small extracellular vesicles in frontotemporal degeneration. *Medicine (Baltimore)*. Oct 7 2022;101(40):e30854. doi:10.1097/MD.0000000000030854
63. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med*. Jan 6 2015;162(1):55-63. doi:10.7326/M14-0697
64. Kohen R, Barlev J, Hornung G, et al. UTAP: User-friendly Transcriptome Analysis Pipeline. *BMC bioinformatics*. 2019;24:154.
65. Kozomara A, Griffiths-Jones S. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res*. Jan 2014;42(Database issue):D68-73. doi:10.1093/nar/gkt1181
66. Kho AT, Sharma S, Davis JS, et al. Circulating MicroRNAs: Association with Lung Function in Asthma. *PLoS One*. 2016;11(6):e0157998. doi:10.1371/journal.pone.0157998
67. Wander PL, Enquobahrie DA, Bammler TK, et al. Associations of plasma miRNAs with waist circumference and insulin resistance among women with polycystic ovary syndrome - Pilot study. *Mol Cell Endocrinol*. Aug 20 2022;554:111723. doi:10.1016/j.mce.2022.111723

68. Wang A, Kwee LC, Grass E, et al. Whole blood sequencing reveals circulating microRNA associations with high-risk traits in non-ST-segment elevation acute coronary syndrome. *Atherosclerosis*. Jun 2017;261:19-25. doi:10.1016/j.atherosclerosis.2017.03.041
69. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):550. doi:10.1186/s13059-014-0550-8
70. Elith J, Leathwick JR, Hastie T. A working guide to boosted regression trees. *J Anim Ecol*. Jul 2008;77(4):802-13. doi:10.1111/j.1365-2656.2008.01390.x
71. Witten IH, Frank E, Hall MA, Pal CJ. Ensemble Learning. In: Witten IH, Frank E, Hall MA, Pal CJ, eds. *Data Mining (Fourth Edition)*. Morgan Kaufmann; 2017:479-501:chap 12.
72. Lundberg SM, Erion G, Chen H, et al. From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence*. 17 january 2020 2020;2:56-67.
73. Liu F, Li T, Hu P, Dai L. Upregulation of Serum miR-629 Predicts Poor Prognosis for Non-Small-Cell Lung Cancer. *Dis Markers*. 2021;2021:8819934. doi:10.1155/2021/8819934
74. Jingushi K, Ueda Y, Kitae K, et al. miR-629 Targets TRIM33 to Promote TGFbeta/Smad Signaling and Metastatic Phenotypes in ccRCC. *Mol Cancer Res*. Mar 2015;13(3):565-74. doi:10.1158/1541-7786.MCR-14-0300
75. Ludwig N, Leidinger P, Becker K, et al. Distribution of miRNA expression across human tissues. *Nucleic Acids Research*. 2016;44(8):3865-3877. doi:10.1093/nar/gkw116
76. Bu X, Li D, Wang F, Sun Q, 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;16:1863-1877. doi:10.2147/NDT.S260748
77. Hoerl AE, Kennard RW. Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*. 1970/02/01 1970;12(1):55-67. doi:10.1080/00401706.1970.10488634
78. Hoerl AE, Kennard RW. Ridge Regression: Applications to Nonorthogonal Problems. *Technometrics*. 1970/02/01 1970;12(1):69-82. doi:10.1080/00401706.1970.10488635
79. Chang WS, Wang YH, Zhu XT, Wu CJ. Genome-Wide Profiling of miRNA and mRNA Expression in Alzheimer's Disease. *Med Sci Monit*. Jun 4 2017;23:2721-2731. doi:10.12659/msm.905064
80. He B, Chen W, Zeng J, Tong W, 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*. Jan 2020;235(1):480-493. doi:10.1002/jcp.28988
81. Ludwig N, Fehlmann T, Kern F, et al. Machine Learning to Detect Alzheimer's Disease from Circulating Non-coding RNAs. *Genomics Proteomics Bioinformatics*. Aug 2019;17(4):430-440. doi:10.1016/j.gpb.2019.09.004
82. Allach El Khattabi L, Heide S, Caberg J-H, et al. 16p13.11 microduplication in 45 new patients: refined clinical significance and genotype–phenotype correlations. *Journal of Medical Genetics*. 2020;57(5):301-307. doi:10.1136/jmedgenet-2018-105389
83. Goncalves TF, Piergiorge RM, Dos Santos JM, Gusmao J, Pimentel MMG, Santos-Reboucas CB. Network Profiling of Brain-Expressed X-Chromosomal MicroRNA Genes Implicates Shared Key MicroRNAs in Intellectual Disability. *J Mol Neurosci*. Feb 2019;67(2):295-304. doi:10.1007/s12031-018-1235-7
84. Wingo TS, Yang J, Fan W, et al. Brain microRNAs associated with late-life depressive symptoms are also associated with cognitive trajectory and dementia. *NPJ Genom Med*. 2020;5:6. doi:10.1038/s41525-019-0113-8
85. Cogswell JP, Ward J, Taylor IA, et al. Identification of miRNA changes in Alzheimer's disease brain and CSF yields putative biomarkers and insights into disease pathways. *J Alzheimers Dis*. May 2008;14(1):27-41. doi:10.3233/jad-2008-14103
86. Rawat P, Sehar U, Bisht J, Selman A, Culberson J, Reddy PH. Phosphorylated Tau in Alzheimer's Disease and Other Tauopathies. *Int J Mol Sci*. Oct 25 2022;23(21)doi:10.3390/ijms232112841

668 87. Muller M, Kuiperij HB, Claassen JA, Kusters B, Verbeek MM. MicroRNAs in Alzheimer's
669 disease: differential expression in hippocampus and cell-free cerebrospinal fluid. *Neurobiol Aging*.
670 Jan 2014;35(1):152-8. doi:10.1016/j.neurobiolaging.2013.07.005
671 88. Nelson PT, Wang WX. MiR-107 is reduced in Alzheimer's disease brain neocortex:
672 validation study. *J Alzheimers Dis*. 2010;21(1):75-9. doi:10.3233/JAD-2010-091603
673 89. Prendecki M, Florczak-Wyspianska J, Kowalska M, et al. APOE genetic variants and
674 apoE, miR-107 and miR-650 levels in Alzheimer's disease. *Folia Neuropathol*. 2019;57(2):106-
675 116. doi:10.5114/fn.2019.84828
676 90. Wang J, Chen C, Zhang Y. An investigation of microRNA-103 and microRNA-107 as
677 potential blood-based biomarkers for disease risk and progression of Alzheimer's disease. *J Clin*
678 *Lab Anal*. Jan 2020;34(1):e23006. doi:10.1002/jcla.23006
679 91. Yilmaz SG, Erdal ME, Ozge AA, Sungur MA. Can Peripheral MicroRNA Expression Data
680 Serve as Epigenomic (Upstream) Biomarkers of Alzheimer's Disease? *OMICS*. Aug
681 2016;20(8):456-61. doi:10.1089/omi.2016.0099
682 92. Wang WX, Rajeev BW, Stromberg AJ, et al. The expression of microRNA miR-107
683 decreases early in Alzheimer's disease and may accelerate disease progression through
684 regulation of beta-site amyloid precursor protein-cleaving enzyme 1. *J Neurosci*. Jan 30
685 2008;28(5):1213-23. doi:10.1523/JNEUROSCI.5065-07.2008
686 93. Lindberg RL, Hoffmann F, Mehling M, Kuhle J, Kappos L. Altered expression of miR-17-
687 5p in CD4+ lymphocytes of relapsing-remitting multiple sclerosis patients. *Eur J Immunol*. Mar
688 2010;40(3):888-98. doi:10.1002/eji.200940032
689 94. Yang D, Wang WZ, Zhang XM, et al. MicroRNA expression aberration in Chinese patients
690 with relapsing remitting multiple sclerosis. *J Mol Neurosci*. Jan 2014;52(1):131-7.
691 doi:10.1007/s12031-013-0138-x
692 95. Honardoost MA, Kiani-Esfahani A, Ghaedi K, Etemadifar M, Salehi M. miR-326 and miR-
693 26a, two potential markers for diagnosis of relapse and remission phases in patient with relapsing-
694 remitting multiple sclerosis. *Gene*. Jul 10 2014;544(2):128-33. doi:10.1016/j.gene.2014.04.069
695 96. Niwald M, Migdalska-Sek M, Brzezianska-Lasota E, Miller E. Evaluation of Selected
696 MicroRNAs Expression in Remission Phase of Multiple Sclerosis and Their Potential Link to
697 Cognition, Depression, and Disability. *J Mol Neurosci*. Dec 2017;63(3-4):275-282.
698 doi:10.1007/s12031-017-0977-y
699 97. Regev K, Healy BC, Paul A, et al. Identification of MS-specific serum miRNAs in an
700 international multicenter study. *Neurol Neuroimmunol Neuroinflamm*. Sep 2018;5(5):e491.
701 doi:10.1212/NXI.0000000000000491
702 98. Sharaf-Eldin WE, Kishk NA, Gad YZ, et al. Extracellular miR-145, miR-223 and miR-326
703 expression signature allow for differential diagnosis of immune-mediated neuroinflammatory
704 diseases. *J Neurol Sci*. Dec 15 2017;383:188-198. doi:10.1016/j.jns.2017.11.014
705 99. Baulina N, Kulakova O, Kiselev I, et al. Immune-related miRNA expression patterns in
706 peripheral blood mononuclear cells differ in multiple sclerosis relapse and remission. *J*
707 *Neuroimmunol*. Apr 15 2018;317:67-76. doi:10.1016/j.jneuroim.2018.01.005
708 100. Dolati S, Aghebati-Maleki L, Ahmadi M, et al. Nanocurcumin restores aberrant miRNA
709 expression profile in multiple sclerosis, randomized, double-blind, placebo-controlled trial. *J Cell*
710 *Physiol*. Jul 2018;233(7):5222-5230. doi:10.1002/jcp.26301
711 101. Regev K, Healy BC, Khalid F, et al. Association Between Serum MicroRNAs and Magnetic
712 Resonance Imaging Measures of Multiple Sclerosis Severity. *JAMA Neurol*. Mar 1
713 2017;74(3):275-285. doi:10.1001/jamaneurol.2016.5197

Main figures and tables

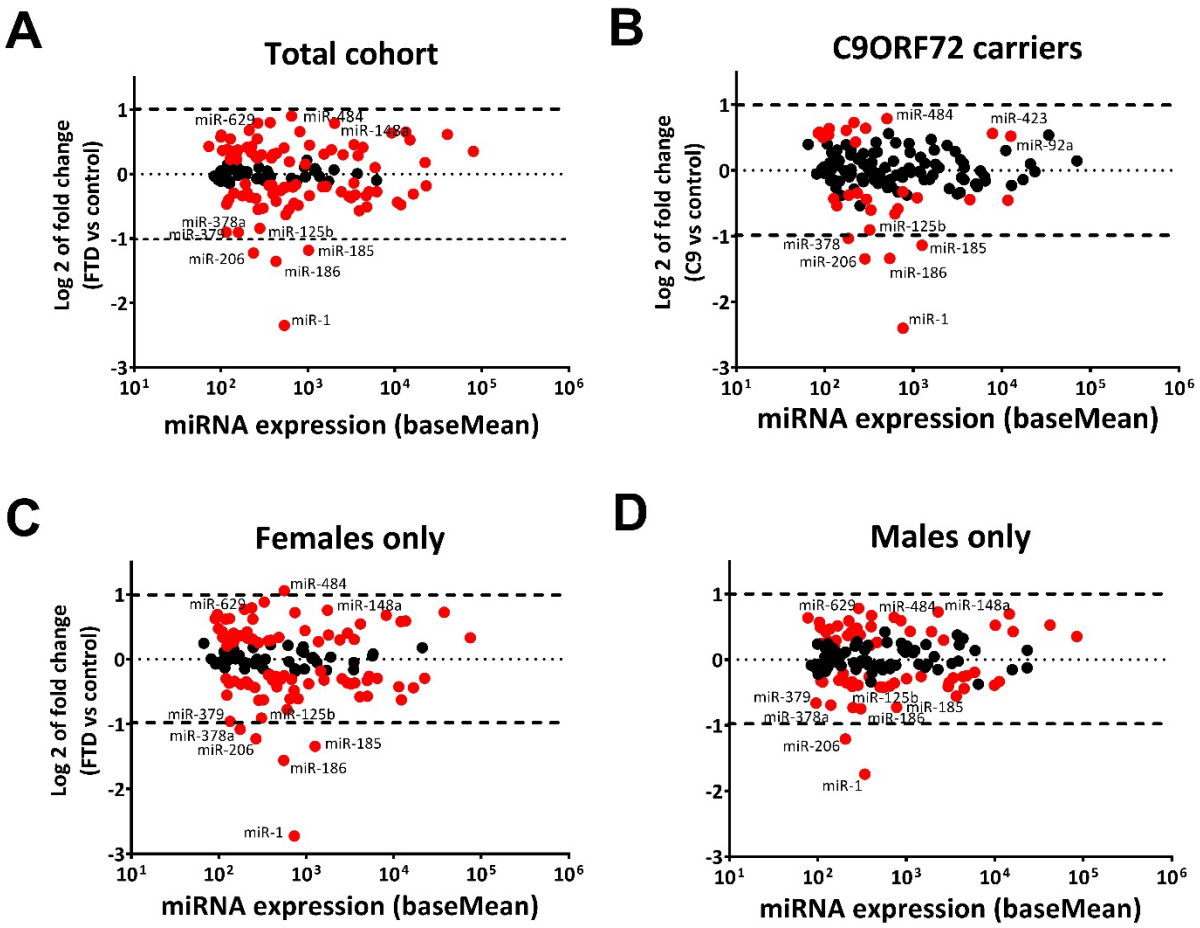


Figure 1. miRNA signature associated with FTD. MA plot of differential miRNA expression in plasma of patients with FTD vs non-neurodegenerative healthy controls in the **(A)** total cohort of FTD patients and controls (N=168 and N=125, respectively). **(B)** C9ORF72 mutations carriers only (N=18) vs healthy controls (N=125). **(C)** Female patients with FTD and healthy female controls (N=58 and N=85, respectively) **(D)** male patients with FTD and healthy male controls (N=110 and N=40, respectively). Log 2 transformed fold change (y-axis), against mean miRNA abundance (x-axis). Red: significantly changed miRNAs (adjusted p<0.05, Wald test). Black: miRNAs showing insignificant change.

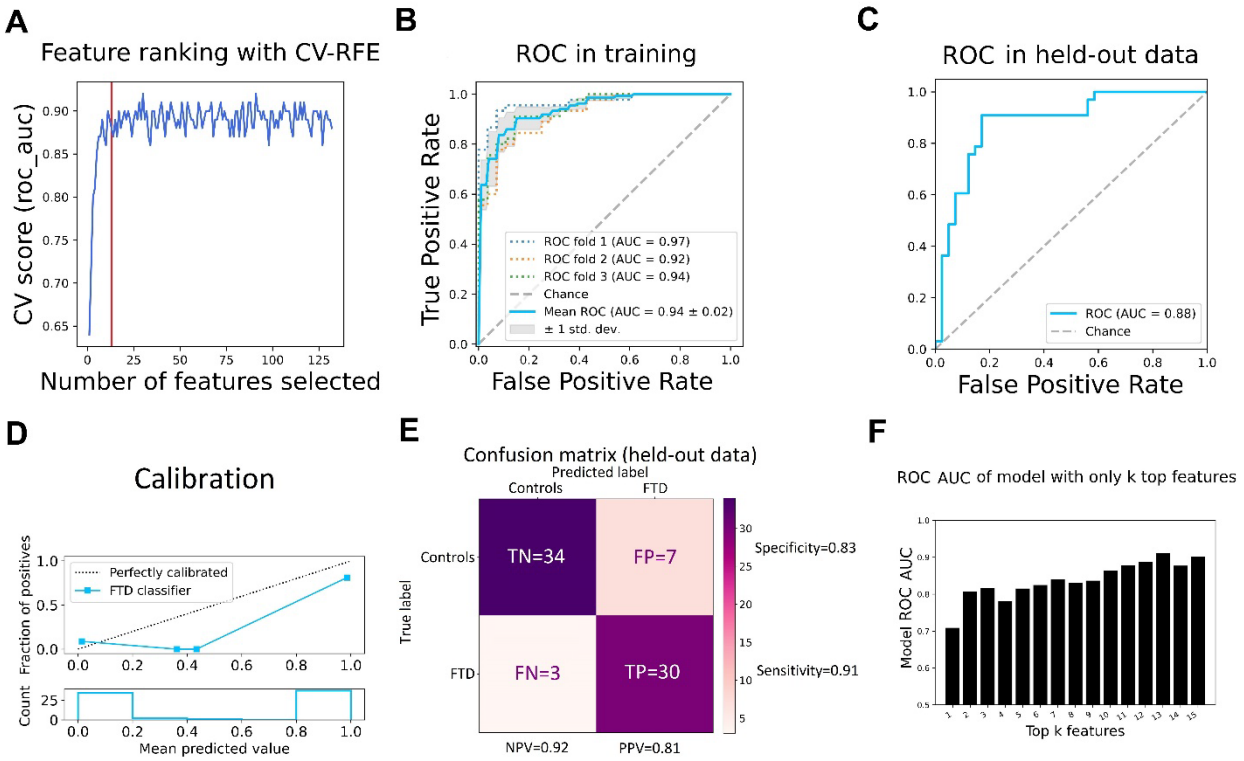


Figure 2: Diagnosis of FTD by a distinctive cell free miRNA signature. **(A)** Mean accuracy under 3-fold cross validation (AUC ROC, y-axis) as a function of the number of plasma miRNAs used in the FTD prediction model. Increasing numbers of miRNA features (x-axis) successively selected in the recursive feature elimination process reveals a saturation at 13 miRNA features (red line). Final model included 15 features (including age and sex as covariates). **(B)** ROC curves in the training set: true positive rate (y-axis) vs. false positive rate (x-axis). Mean values and variance of data from 219 samples with 3-fold cross validation. Mean AUC 0.94 ± 0.02. 95% CI is 0.92-0.96. **(C)** Performance and generalizability on replication held-out data revealed by an ROC curve with AUC of 0.88. **(D)** Reliability diagram, plotting truly observed fraction of cases (upper plot, y-axis) vs. predicted probability by the prediction model (upper plot, x-axis), in five probability bins, reveals a sufficiently calibrated model on the held-out set. Lower plot shows the number of individuals at each of the five predicted probability bins. **(E)** Confusion matrix showing the prediction errors on held out dataset. At a probability threshold of 0.6, we observed 0.92 Negative predictive value (NPV), 0.81 Positive Predictive Value (PPV/precision), 0.91 recall (sensitivity) and 0.83 specificity. In addition, a 0.17 False Positive Rate is obtained over 0.08 False Omission Rate and 0.19 false discovery rate. **(F)** AUC ROC of a model trained with only a subset of top 15 most predictive features, reveals a stable performance with the selected final features.

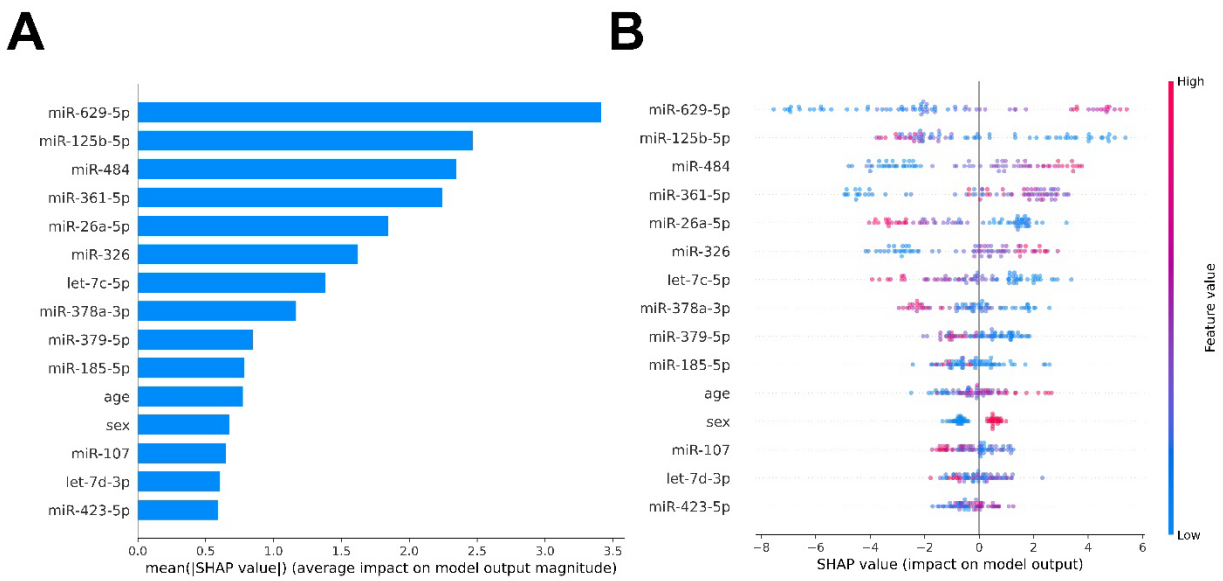


Figure 3: The contribution of individual miRNAs to the predictor of FTD diagnosis. (A) Mean absolute SHapley Additive exPlanations (SHAP) values break down the impact of specific miRNAs on FTD disease non-linear predictor output in the held-out cohort. **(B)** Illustration of the relationship between the miRNA levels (low – blue to high – red), SHAP values and the impact on the prediction in the held-out cohort. Positive or negative SHAP value lead the model to a more likely FTD or healthy control predictions, respectively.

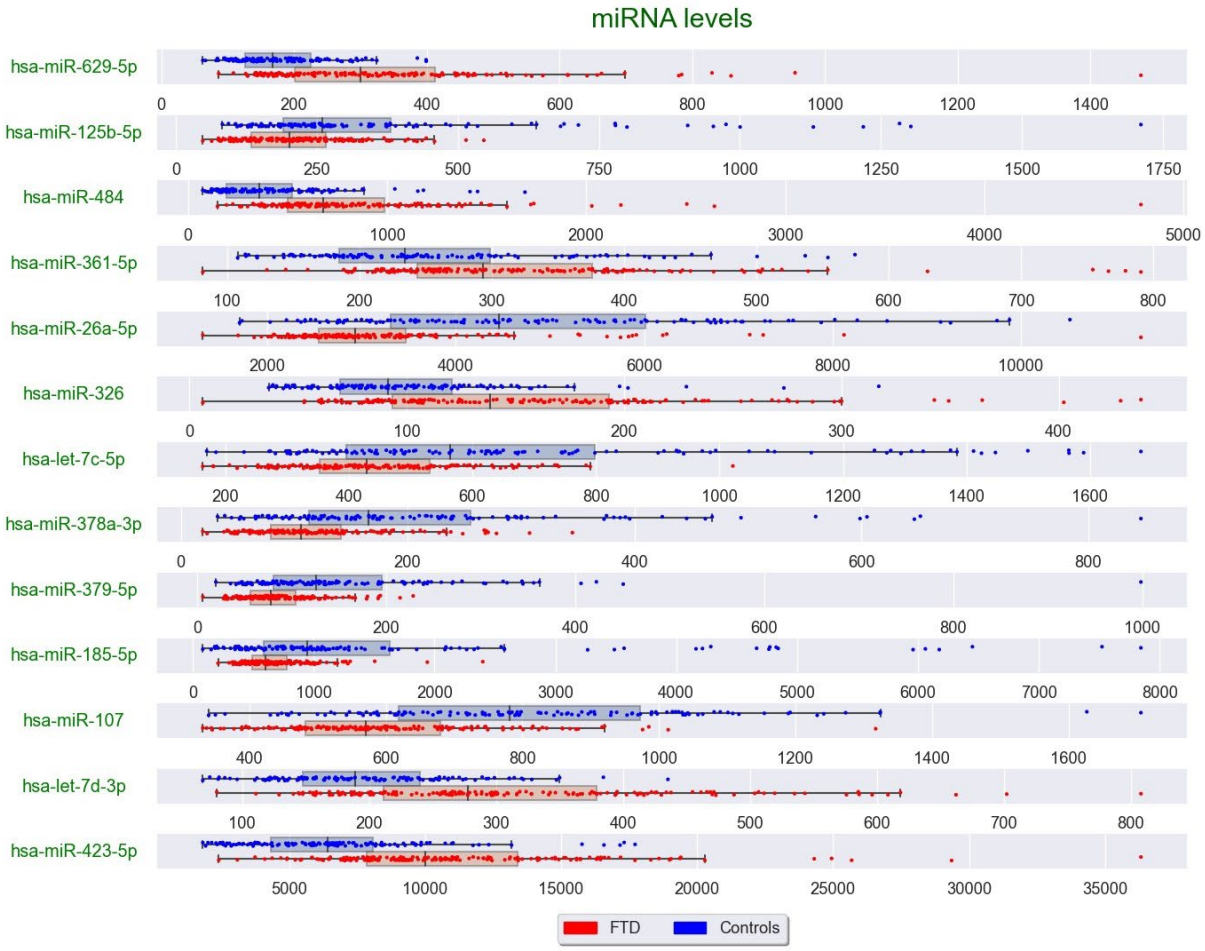


Figure 4. Distribution of the top 13 miRNA predictors in patients with FTD and controls. Box plots with complement scatter plots of the 13 miRNA features predicting FTD, depicting expression values of all cohort along with distribution of each miRNA. Blue dots represent the levels of a given miRNA measured in healthy control participants. Red dots represent the levels of a given miRNA measured in patients with FTD. Box – two central data quartiles, with a line at the median (Q2). Whiskers extend to show the rest of the distribution, except for points that are determined to be outliers using a method that is a function of the inter-quartile range.

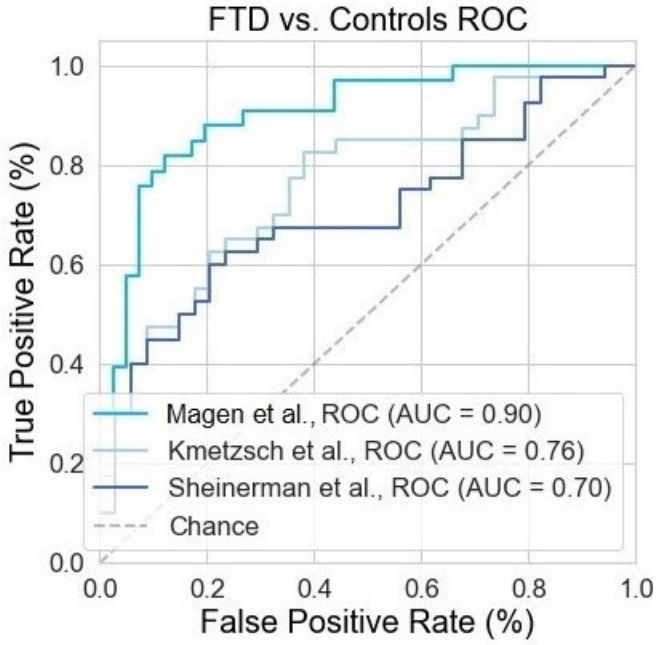


Figure 5. AUC values for different miRNA classifiers for discriminating FTD vs controls in the held-out data. The classifiers were either the 13 selected in the FTD model in our data, a combination of miR-34a-5p, miR-345-5p, miR-200c-3p and miR-10a-3p [from], or a combination of miR-335/let-7e, miR-99b/let-7e and miR-9-3p/miR-181a [from];

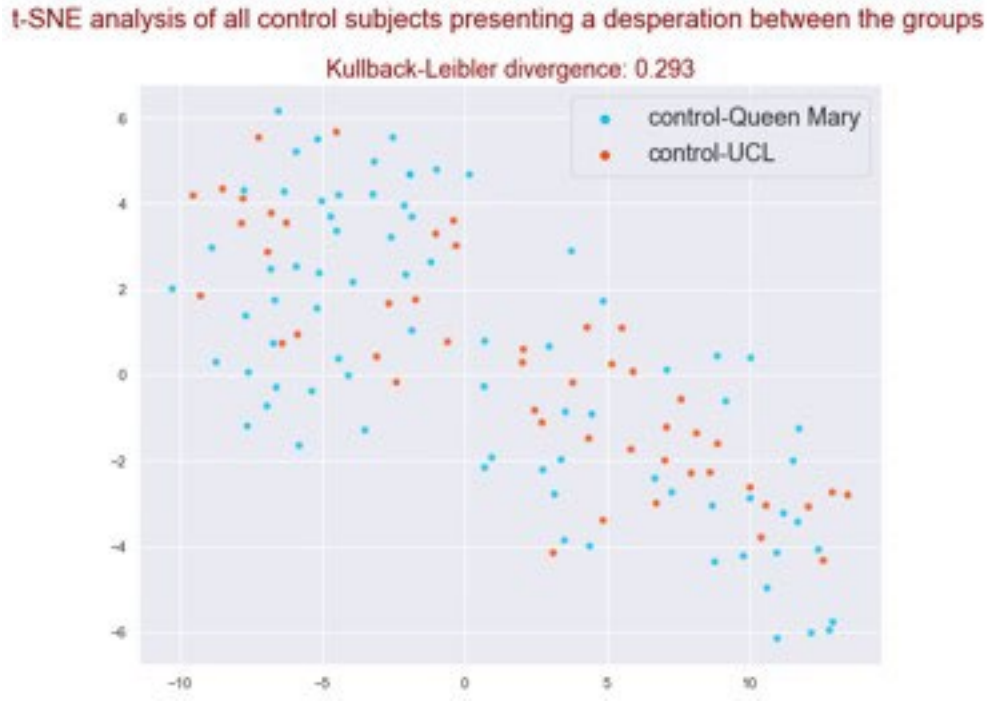
	Control	FTD
Number of subjects (% males)	125 (32%)	168 (65.5%)
UCL controls	48 (45.8%)	
Queen Mary Hospital controls	77 (23.4%)	
Age at enrolment	59.5±10.2 yr.	65.8±8.1 yr.
UCL controls	65.6±7.3 yr.	
Queen Mary Hospital controls	55.7±10.0 yr	
Age of onset (1st reported symptoms)		60.3±8.3 yr.
Disease duration at enrolment		5.5±3.4 yr.
FTD clinical subtype (bvFTD/PNFA/SD/FTD-ALS/others)		81/40/28/5/14
FTD Mutation carriers (C9ORF72/MAPT/GRN/TBK1)		18/13/13/2
Likely FTD pathology (TDP-43/Tau)		63/18

Table 1. Summary of demographic and clinical characteristics of participants suffering from FTD and control samples. bvFTD: behavioural FTD; PNFA: progressive nonfluent aphasia; SD: semantic dementia. Mean±SD.

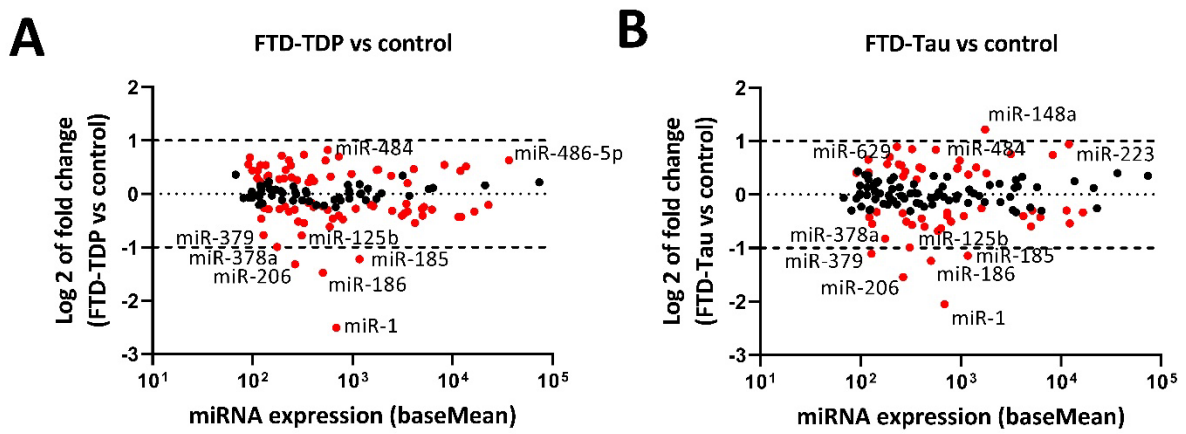
Predictor	Single Feature AUC
miR-423-5p	0.69
miR-125b-5p	0.67
miR-26a-5p	0.66
sex	0.66
miR-326	0.65
miR-185-5p	0.64
age	0.64
miR-629-5p	0.63
miR-484	0.62
let-7d-3p	0.62
miR-107	0.62
let-7c-5p	0.61
miR-361-5p	0.59
miR-379-5p	0.56
miR-378a-5p	0.54

Table 2. Predictive power of selected miRNA 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.

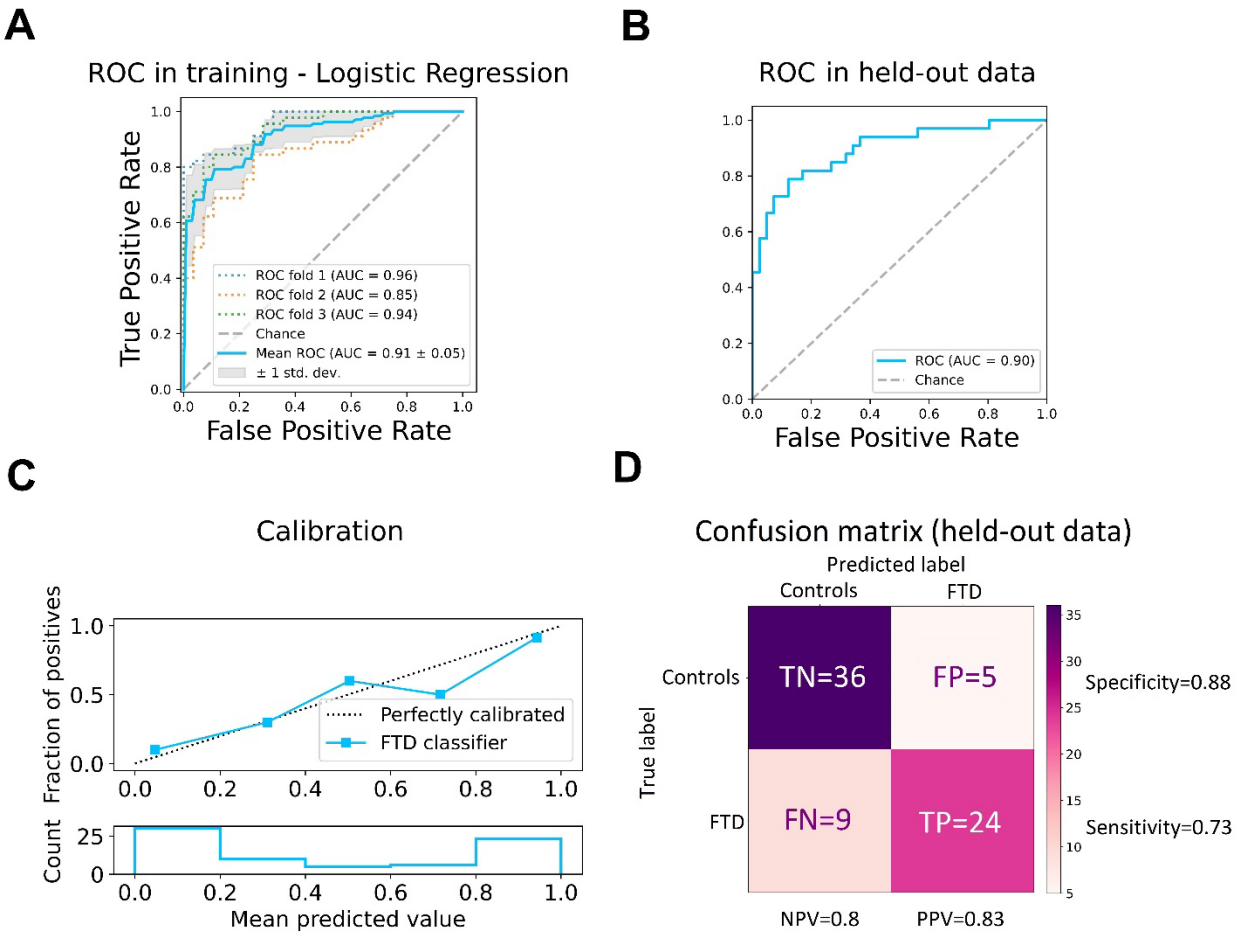
Supplementary figure and tables



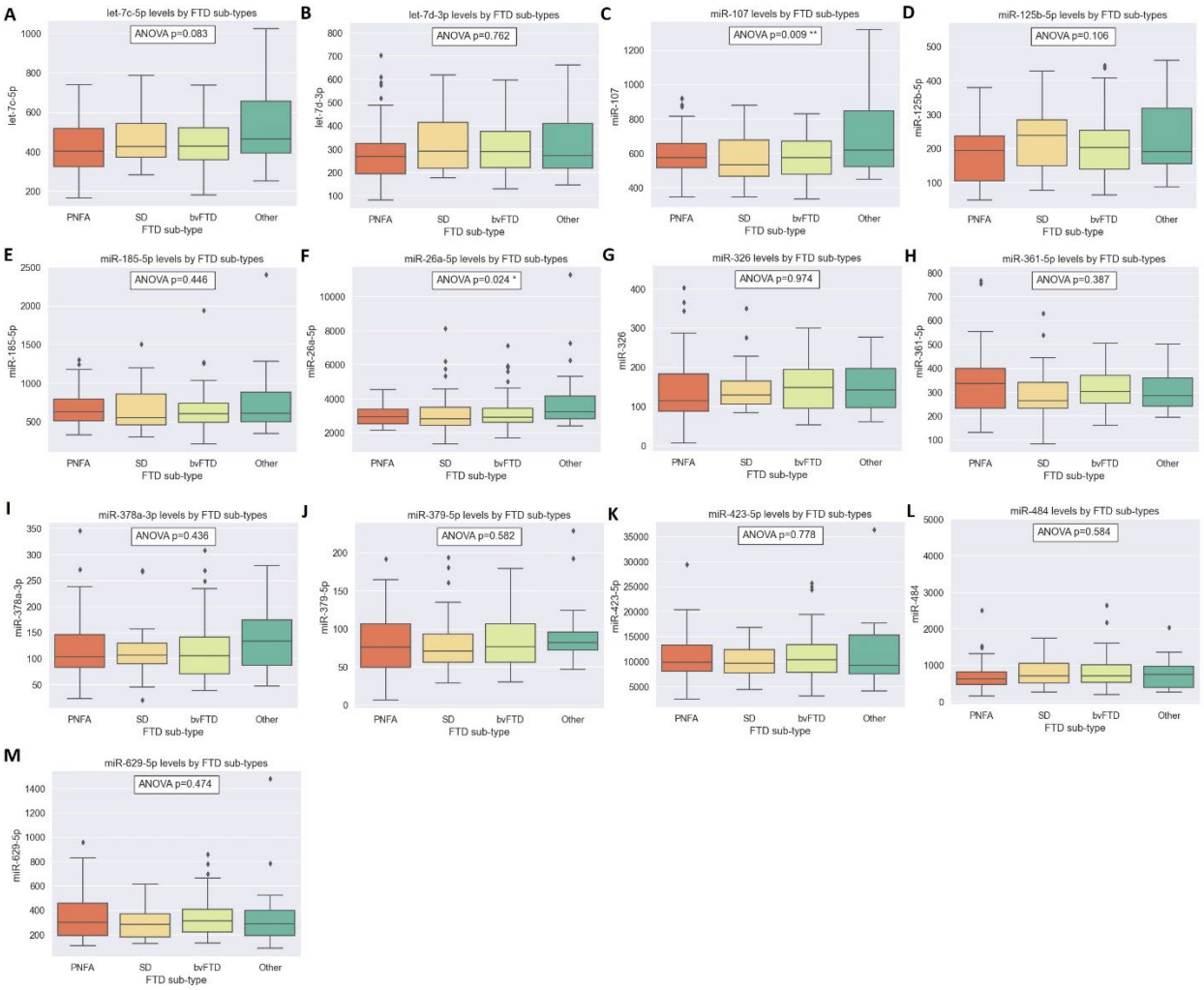
Supplementary Figure 1. t-SNE analysis of control subjects from Queen Mary Hospital (blue) and from UCL (red). Similarity of the two control groups suggests they could have been taken from a single distribution which is justified further by an associated Kullback–Leibler divergence value of 0.293, a measure of the difference between two probability distributions.



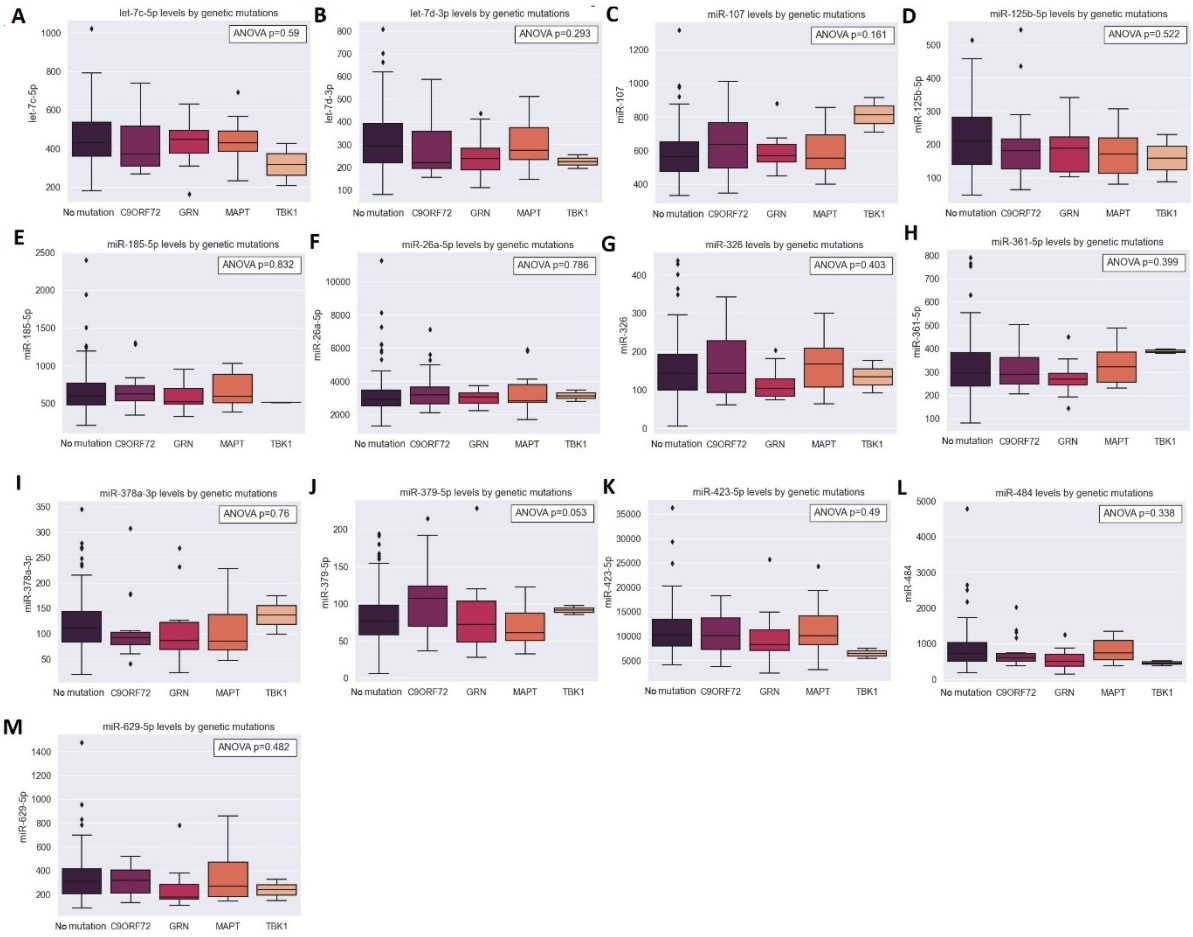
Supplementary Figure 2. miRNA signature associated with FTD patients with predicted pathology. MA plot of differential miRNA expression in plasma of patients with predcited TDP pathology (**A; N=63**) or Tau pathology (**B; N=19**) vs non-neurodegenerative healthy controls (N=125). Log 2 transformed fold change (y-axis), against mean miRNA abundance (x-axis). Red: significantly changed miRNAs (adjusted p<0.05, Wald test). Black: miRNAs showing insignificant change.



Supplementary Figure 3. Diagnosis of FTD by age, sex and 95 differentially-expressed miRNA in a logistic regression model. ***(A)*** ROC curves in the training set: true positive rate (y-axis) vs. false positive rate (x-axis). Mean values and variance of data from 219 samples with 3-fold cross validation. Mean AUC 0.91 ± 0.05. 95% CI is 0.86-0.96. **(B)** Performance and generalizability on held-out data revealed by a ROC curve with AUC of 0.9. **(C)** Reliability diagram, plotting truly observed fraction of cases (upper plot, y-axis) vs. predicted probability by the prediction model (upper plot, x-axis), in five probability bins, reveals a sufficiently calibrated model on the held-out set. Lower plot shows the number of individuals at each of the five predicted probability bins. **(D)** Confusion matrix showing the prediction errors on held out dataset. At a probability threshold of 0.6, we observed 0.8 Negative predictive value (NPV), 0.83 Positive Predictive Value (PPV/precision), 0.73 recall (sensitivity) and 0.88 specificity. In addition, 0.12 False Positive Rate is obtained, over 0.2 False Omission Rate and 0.17 false discovery rate.



Supplementary Figure 4. Levels of 13 miRNA predictors in different FTD subtypes. miRNA levels for PNFA (progressive nonfluent aphasia), SD (semantic dementia), behavioral FTD (bvFTD) and the following subtypes: primary progressive aphasia, corticobasal syndrome, progressive supranuclear palsy and FTD with motor neuron disease, commonly referred to as "Other". Box plot with two central data quartiles, with a line at the median (Q2). Whiskers extend to show the rest of the distribution, except for points that are determined to be outliers using a method that is a function of the inter-quartile range. Data were analyzed by one-way ANOVA. Post-hoc Bonferroni was conducted only for miR-107 **(C)** and miR-26a-5p **(F)** which showed significant differences in ANOVA. Significant pairwise differences in miR-107 levels were found between other FTD subtypes and both bvFTD (p=0.01) and SD (p=0.02). Significant pairwise differences in miR-26a were found between other FTD subtypes and both bvFTD (p=0.04) and PNFA (p=0.02).



Supplementary Figure 5. Levels of 13 miRNA predictors in mutations carriers. miRNA levels for FTD patients without known FTD mutations, or with mutations in C9ORF72, Progranulin (GRN), Tau (MAPT) and TBK1. Box plot with two central data quartiles, with a line at the median (Q2). Whiskers extend to show the rest of the distribution, except for points that are determined to be outliers using a method that is a function of the inter-quartile range. Data were analyzed by one-way ANOVA.

	Before Exclusion				***After Exclusion***		
Group Name	***Age [95% CI]***	***Sex***	***Mean Age***	***Group Size***	***Age [95% CI]***	***Mean Age***	***Group Size***
controls total (n=158)	53.8±14.5, 95% CI [51.5, 56.1]	**F**	54.0	103	59.5±10.2, 95% CI [57.7, 61.3]	58.4	85
		M	53.4	55		61.7	40
UCL controls (n=56)	61.0± 13.3, 95% CI [57.4, 64.5]	**F**	61.1	30	65.55± 7.3, 95% CI [63.4, 67.7]	65.6	26
		M	60.8	26		65.5	22
Queen Mary Hospital controls (n=102)	49.9±13.7, 95% CI [47.2, 52.6]	**F**	51.1	73	55.7±10.0, 95% CI [53.4, 58.0]	55.3	59
		M	46.8	29		57.0	18
FTD cases (n=169)	65.6± 8.3, 95% CI [64.4, 67.0]	**F**	66.8	59	65.8±8.1, 95% CI [64.6, 67.0]	67.4	58
		M	65.0	110		65.0	110

806 **Supplementary Table 1.** Age and sex characteristics of the cohort and exclusion by age of 34 individuals
807 that were below 40 years of age during blood collection.