medRxiv preprint doi: https://doi.org/10.1101/2020.01.22.20018408; this version posted January 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Classification and prediction of frontotemporal dementia based on plasma microRNAs

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

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

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

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

5 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: 242.

Text word count (excluding references and figure legends): 3528.

No. of items (figures, tables): 4, 2

References: 39

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.01.22.20018408; this version posted January 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Abstract

Frontotemporal dementia (FTD) is a heterogeneous neurodegenerative disorder characterized by frontal and temporal lobe atrophy, typically manifesting with behavioural or language impairment. Because of its heterogeneity and lack of available diagnostic laboratory tests there can be a substantial delay in diagnosis. Cell-free, circulating, microRNAs are increasingly investigated as biomarkers for neurodegeneration, but their value in FTD is not yet established. In this study, we investigate microRNAs as biomarkers for FTD diagnosis. We performed next generation small RNA sequencing on cell-free plasma from 52 FTD cases and 21 controls. The analysis revealed the diagnostic importance of 20 circulating endogenous miRNAs in distinguishing FTD cases from controls. The study was repeated in an independent second cohort of 117 FTD cases and 35 controls. The combinatorial microRNA signature from the first cohort, precisely diagnosed FTD samples in a second cohort. To further increase the generalizability of the prediction, we implemented machine learning techniques in a merged dataset of the two cohorts, which resulted in a comparable or improved classification precision with a smaller panel of miRNA classifiers. In addition, there are intriguing molecular commonalities with cell free miRNA signature in ALS, a motor neuron disease that resides on a pathological continuum with FTD. However, the signature that describes the ALS-FTD spectrum is not shared with blood miRNA profiles of patients with multiple sclerosis. Thus, microRNAs are promising FTD biomarkers that might enable earlier detection of FTD and improve accurate identification of patients for clinical trials

medRxiv preprint doi: https://doi.org/10.1101/2020.01.22.20018408; this version posted January 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Introduction

Frontotemporal dementia (FTD) is a clinically and neuroanatomically heterogeneous neurodegenerative disorder characterized by frontal and temporal lobe atrophy. It typically manifests between the ages of 50 and 70 with behavioral or language problems, and below the age of 65 is the second most common form of dementia, after Alzheimer's disease (1).

Due to heterogeneity in clinical presentation FTD can be difficult to diagnose (2). Three main phenotypes 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 (2, 3).

FTD is also pathologically heterogeneous with inclusions seen containing hyperphosphorylated tau (4), TDP- 43 (5), or fused in sarcoma (FUS) (6, 7). Mutations in the genes encoding for these proteins, as well as in other genes such as progranulin (GRN), chromosome 9 open reading frame 72 (C9ORF72), valosin-containing protein (VCP), TANK-binding kinase 1 (TBK1) and charged multivesicular body protein 2B (CHMP2B) are also associated with FTD (8-11).

FTD overlaps clinically, pathologically and genetically with several other degenerative disorders. In particular, there is often overlap with amyotrophic lateral sclerosis (ALS): one in 5 ALS patients meets the clinical criteria for a concomitant diagnosis of FTD, and one in eight FTD patients is also diagnosed with ALS. TDP-43 inclusions are observed in the brains of both people

medRxiv preprint doi: https://doi.org/10.1101/2020.01.22.20018408; this version posted January 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

with FTD and ALS, and genetic evidence supports that these diseases reside along a continuum (5, 12-14).

Previous studies have aimed to develop cell-free biomarkers for FTD, including TDP-43 (15), tau (16), and neurofilament light chain (NfL) (17), but none of these have shown use for diagnosis. microRNAs (miRNAs), endogenous non-coding RNAs, can be quantified in biofluids (18), and have been shown previously to be dysregulated in amyotrophic lateral sclerosis (ALS) and in FTD (19). Furthermore, they may be biomarkers of disease progression in other brain diseases, including ALS (20). Previous studies have assessed the initial potential of microRNAs as diagnostic FTD biomarkers including miRNA analysis in plasma (21-23), CSF and serum (24), and CSF exosomes (25) but no definitive markers have so far been found. We therefore aimed to study a large cohort of patients with different clinical phenotypes and pathological forms of FTD, to see whether they are able to reliably distinguish cases from controls, and different forms of FTD from each other.

Here, we provide an unbiased signature of plasma miRNAs that has good diagnostic power in a large and heterogeneous cohort of patients with FTD, which is further predictive in an independent second cohort and may contribute to FTD subtyping. Therefore, circulating miRNAs hold a fascinating potential as diagnostic biomarkers and as means for patient stratification in clinical trials.

medRxiv preprint doi: https://doi.org/10.1101/2020.01.22.20018408; this version posted January 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Results

A plasma miRNA classifier for FTD

In order to characterize the potential of plasma miRNAs as biomarkers for FTD we assembled a cohort of 73 plasma samples (subject information in Table 1), purified RNA and performed next generation sequencing (NGS). As many as 2313 individual miRNA species were aligned to the human genome (GRCh37/hg19) across all samples. However, only 137 miRNA species exceeded a cut-off of ≥100 UMIs per sample averaged on all samples. Of the 137 detected miRNAs, 20 miRNA changed in a statistically significant manner in FTD plasma relative to control (p-value < 0.05, Wald test; Fig. 1A). Two miRNAs, whose levels decreased to the greatest extent in FTD compared to controls, namely, miR-379-5p and miR-654-3p (1.4 fold), remained significant after multiple hypothesis testing (Fig. 1B).

We next studied miRNA capacity as binary disease classifiers, by generating receiver-operating characteristic (ROC) curves. ROC area under the curve (AUC) suggested modest predictive capacity for miR-379-5p and for miR-654-3p (AUC for both: 0.71±0.07, p<0.01; Fig. 1C).

We further utilized the combinatorial signature of the 20 miRNAs that were differentially expressed between FTD patient plasma and control (Table S1). Using these, an ROC AUC of 0.79±0.05 (p<0.0001, Fig. 1C) was found, which was superior to the prediction capacity of any individual miRNA.

medRxiv preprint doi: https://doi.org/10.1101/2020.01.22.20018408; this version posted January 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Assessment of plasma miRNA classifier for FTD in a second cohort

We then performed a replication study in an independent cohort of 117 FTD cases and 35 age-matched controls (Table 1). In this study, the levels of 58 miRNAs decreased and 89 increased in a statistically significant manner in FTD, relative to control plasma (p-value <0.05, Wald test, Fig. 2A and Table S1). Noticeable miRNAs were miR-125b-2-3p (× 26 up, $p = 1.4x10^{-25}$), miR-34b-5p (× 23 up, adj. $p = 9.8x10^{-23}$) and miR-379-5p (× 2.2 down, $p = 1.9x10^{-14}$). 144 of the 147 miRNAs further survived adjustment for multiple comparisons by Benjamini–Hochberg procedure (adjusted p-value < 0.05).

The expression of the 20 miRNAs that were most differentially expressed in the first cohort correlated with their respective expression in the second cohort (Pearson R of log 2 fold-change = 0.75, p=0.0001, Fig. 2B). Furthermore, the combined predictive power of the 20 miRNAs, that were decided on as a classifier based on data of the first cohort, was slightly superior in the replication cohort, with an AUC of 0.82±0.04 (p<0.0001, Fig. 2C).

In addition to external validation, by testing a second cohort, we sought to guarantee the generalizability by building a machine learning gradient boosting classifier on a unified cohort. and applying K-fold cross-validation, which is an internal validation technique to evaluate performance and prevent overfitting (26, 27). Towards this we divided the 225 datasets (from 56 control and 169 FTD samples) randomly into three equal parts, or ‘folds’, of 75 datasets, each. A machine learning model was trained using each time 2 of the 3 data folds (150 samples) for building a prediction model and applying the prediction rule to estimate the prediction precision on the remaining 75 samples in the remaining third fold. This step was repeated k = 3 times

medRxiv preprint doi: https://doi.org/10.1101/2020.01.22.20018408; this version posted January 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

iteratively so all folds were used twice in training and once for the testing process. 136 miRNAs that were measured above noise levels in all cohorts were included, yielding the following AUCs: 0.90 for fold 1; 0.87 for fold 2; and 0.93 for fold 3, with an average AUC of 0.90 (Fig. 2D).

We next aimed to reduce the complexity of the measurements required for disease prediction by identifying the top 20 miRNA predictors per fold, i.e. the 20 miRNAs with the highest weighted importance in predicting disease status (Fig. 3A-C). We reduced the number of miRNAs gradually, starting from a 43 miRNA panel composed of the top 20 predictors in at least one fold (i.e., in one, two or three folds), which resulted in AUCs of 0.87, 0.87, 0.94 and an average AUC of 0.89 (Fig. 3D). We then utilized 13 miRNAs that were among the top 20 in at least two folds which resulted in AUCs of 0.85, 0.89 and 0.93, and an average of 0.89 (Fig. 3E). Finally, we used only four miRNAs - miR-26a-5p, miR-326, miR-203a-3p and miR-629-5p – that were among the top 20 predictors in all three folds. Their combinatorial AUCs after cross-validation were 0.81, 0.83 and 0.89 and 0.85 on average (Fig. 3F). All panels of miRNAs used for the cross-validation are listed in Table S1.

These measurements were comparable to the AUC obtained with 136 miRNAs (Fig. 2D), revealing that the diagnostic power was not compromised by a substantial reduction of the miRNA numbers.

medRxiv preprint doi: https://doi.org/10.1101/2020.01.22.20018408; this version posted January 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Overlap between FTD miRNA signature and ALS miRNA signature

FTD and ALS are two diseases on a neuropathology continuum. We aimed to determine whether the miRNA signatures found in FTD and in ALS reveal any molecular similarity. For this purpose, we sequenced and analyzed the differences between 115 ALS cases and 103 controls (see Table 2). We also sequenced 17 samples from patients with multiple sclerosis (MS), because this disease is mechanistically different from FTD and involves autoimmune-related demyelination, so molecular similarity to FTD is not expected to be seen. 161 miRNA species were differentially expressed in either one of the diseases (FTD, ALS or MS) vs controls. Differentially expressed miRNAs in either FTD or ALS were correlated in fold-change values between the diseases (Pearson R for log-transformed values = 0.35, $p<0.0001$, Fig. 4A), but no such correlation was found between FTD and MS (R= - 0.15, p=0.15, Fig. 4B). Intriguingly, muscle-specific miR-206 robustly increases in ALS, in agreement with previous reports (28-31) with no change at all in FTD.

We next tested the degree of overlap between miRNAs differentially expressed in FTD vs. ALS. Seven out of 20 miRNAs changed exclusively in FTD, and the remaining 13 miRNAs changed in a significant manner in both FTD and ALS (Fig. 4C; Table S1). Remarkably, the directionality of change for these miRNAs (increase/decrease) was consistent across diseases for all of the miRNAs but one, miR-29a-3p which decreased in FTD and increased in ALS (Fig. 4A). Moreover, the fold-change values in this subset of 13 miRNAs that have changed in both ALS and FTD, were highly correlated between the diseases (Pearson R = 0.90, $p<0.0001$). In contrast, only five out of the 20 miRNAs that changed in FTD, also changed in MS (Fig. 4D; Table S1).

medRxiv preprint doi: https://doi.org/10.1101/2020.01.22.20018408; this version posted January 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Taken together, the miRNA signature in FTD plasma shows a similarity to the ALS plasma signature, but not to the MS signature, in accordance with pathological and clinical similarities between FTD and ALS.

Finally, we employed the FTD predictor, based on 20 miRNAs that are changing in FTD on ALS and healthy control cohorts. The signature of 20 miRNAs was able to correctly call ALS from controls more than at random (ROC AUC = 0.63, $p<0.001$, Table S2), while the seven miRNAs that are exclusively changed in FTD were not able to distinguish between ALS and control in a statistically significant manner (ROC AUC = 0.57, $p=0.06$). Thus, miRNAs that are differentially expressed in FTD have a moderate capacity to predict ALS.

miRNAs signature of FTD subtypes and FTD patients with different pathologies

We next tested whether specific miRNAs changed in the main FTD subtypes, bvFTD, SD and PNFA. After statistical adjustment for multiple comparisons, four miRNAs decreased in a significant manner in PNFA, and two miRNAs decreased and one miRNA increased significantly in bvFTD, whereas the small SD sample numbers (n=8) did not allow to depict microRNAs that are changed in a significant manner after adjustment for multiple comparisons (Fig. S1A-C).

We calculated a decent combinatorial predictive power for the 20 miRNAs in distinguishing bvFTD / SD / PNFA from healthy controls: thus, for bvFTD vs. healthy controls in the original cohort we obtained an AUC of 0.85±0.06, $p<0.0001$; in the replication cohort AUC of

medRxiv preprint doi: https://doi.org/10.1101/2020.01.22.20018408; this version posted January 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

0.80±0.05, p<0.0001, Fig. S1D; for SD vs. controls, original cohort AUC was 0.86±0.08, p=0.003; replication cohort AUC was 0.79±0.06, p=0.0003 (Fig. S1E); for PNFA vs. controls, original cohort AUC was 0.81±0.08, p=0.002; replication cohort AUC was 0.81±0.05, p<0.0001 (Fig. S1F). We concluded that the combinatorial 20 miRNAs signature distinguishes FTD and its subtypes from controls with comparable AUCs, for all three subtypes.

The overlap of symptoms between subtypes of FTD poses a diagnostic challenge (32). We therefore tested whether FTD subtypes could be distinguished based on miRNA signature. We analyzed miRNA differential expression of PNFA cases vs. non-PNFA, which pooled together bvFTD and SD cases, due to a similar molecular signature of SD and bvFTD. Fourteen miRNAs changed in a significant manner in PNFA vs non-PNFA: miR-625-3p, miR-625-5p, miR-126-5p, miR-146a-5p, miR-146b-5p, miR-340-5p, miR-181a-5p (all increased in PNFA compared to non-PNFA) and miR-342-3p, let-7d-3p, miR-122-5p, miR-192-5p, miR-16-5p, miR-203a-3p (decreased; Fig. S1G). The combinatorial signature of these fourteen miRNAs yielded an AUC of 0.81±0.08 (Fig. S1H; p=0.0007), indicating that PNFA can be differentiated from other types of FTD with a high accuracy.

We also tested whether specific miRNAs changed between FTD cases with different likely underlying pathologies, i.e. tau and TDP-43. 19 FTD cases with predicted Tau pathology based on genetics (4 in cohort I + 15 in cohort II) were compared to 63 cases with predicted TDP-43 pathology (23 in cohort I + 40 in cohort II). Fourteen miRNAs changed in a statistically significant manner, but none remained significant after correction for multiple hypotheses (Fig. S2A). The combinatorial signature of these 14 miRNAs had a weak classification power, though

medRxiv preprint doi: https://doi.org/10.1101/2020.01.22.20018408; this version posted January 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

it was statistically significant (AUC of ROC = 0.7±0.06, p=0.009, Fig. S2B). Taken together, the miRNA profile in our dataset has limited diagnostic power for pathological subtypes of FTD, as opposed to FTD vs control and different clinical subtypes of FTD.

medRxiv preprint doi: https://doi.org/10.1101/2020.01.22.20018408; this version posted January 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Discussion

Our study utilizes a large cohort of FTD blood samples. It is the first work that employs next generation sequencing technology for FTD biomarkers. We defined a signature, composed of 20 miRNAs, that is able to classify FTD. This signature that was discovered in an initial cohort was informative when applied to a second cohort. These observations suggest that miRNAs can be potentially utilized in clinical sampling as diagnostic FTD markers, which is needed because of non-specific early symptoms and overlap with other degenerative and non-degenerative conditions. Ours is the largest cohort used for miRNA profile, and its use of unbiased exhaustive next generation sequencing can potentially explain the discrepancies from past studies with smaller cohorts and biased miRNA choices (21-25).

A classifier panel of 20 miRNAs had ~80% chance to correctly call FTD in the first cohort. Reassuringly, it was comparably informative in calling FTD correctly also on a second cohort. In addition to external (second cohort) validation, we applied machine learning to the whole dataset of 225 samples. Through iterative learning, we defined a signature created by 136 miRNAs that was able to call FTD correctly in 90% of cases. We then reduced the signature complexity to the usage of only 43 miRNAs with the highest classification power that kept a true FTD calling capacity of 90%. Toward clinical diagnostic usage it is warranted to test the predictor that was developed in machine learning on an independent cohort, preferentially of different ethnicity.

Interestingly, the miRNA signatures of FTD is akin of ALS perhaps reflecting on a shared patho-mechanism for these two neurodegenerative disorders on the ALS-FTD continuum. This similarity cannot be extended to multiple sclerosis, a disease that is driven by a different,

medRxiv preprint doi: https://doi.org/10.1101/2020.01.22.20018408; this version posted January 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

autoimmune, mechanism. Nonetheless, the two diseases are still two different entities and accordingly only 10% of the miRNAs that has changed in either disease were shared.

In summary, we have characterized a large FTD plasma cohort for miRNA expression by next generation sequencing and found specific patterns of changes that can contribute to diagnosis of FTD. These patterns seem to involve the ALS-FTD continuum, alluding to differences and commonalities in the underlying mechanisms that drive molecular changes in ALS and FTD.

medRxiv preprint doi: https://doi.org/10.1101/2020.01.22.20018408; this version posted January 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Materials and 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).

Study design

Based on power analysis, we found that about 20 control and 50 FTD samples are required to obtain an ROC of 0.7 with a power of 80% and a p-value of 0.05. We determined the sample size based on these calculations. Because sample processing was done in different batches, samples were randomly allocated to the batches and within each batch, the number of control and FTD/ALS/MS samples was balanced in order to reduce batch-associated bias.

Participants and sampling

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 I: 52 FTD patients, 21 healthy controls. Study cohort II: 117 FTD patients, 35 healthy controls. FTD patients were further assigned into two groups with predicted pathology of TDP-43 or tau, based on genetics and clinical phenotype. Patients positive for C9ORF72 repeats and progranulin (PRGN) mutations and/or presented with semantic dementia, were predicted to have TDP-43 pathology, while patients with MAPT mutations were predicted to have tau pathology. Demographic data are detailed in table 1.

ALS and MS samples and their respective healthy controls (N = 115, 17 and 103, respectively) were obtained from the ALS biomarker study. ALS patients were diagnosed according to

medRxiv preprint doi: https://doi.org/10.1101/2020.01.22.20018408; this version posted January 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

standard criteria by experienced ALS neurologists (33). Healthy controls were typically spouses or relatives of patients. Demographic data are detailed in table 2.

Plasma samples were stored in -80^{0}C until RNA extraction and subsequent small RNA next generation sequencing.

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 (NGS), 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. Following 3' and 5' adapter ligation, small RNA was reverse transcribed, using unique molecular identifier (UMI), primers of random 12-nucleotide sequences. This way, precise linear quantification miRNA is achieved, overcoming potential PCR-induced biases (18). cDNA libraries were amplified by PCR for 22 cycles, with a 3' primer that includes a 6-nucleotide unique index. Following size selection and cleaning of libraries with magnetic beads, quality control was performed by measuring library concentration with Qubit fluorometer using dsDNA high sensitivity (HS) assay kit (Thermo Fisher Scientific, Waltham, MA) and confirming library size with Tapestation D1000 (Agilent). Libraries with different indices were multiplexed and sequenced on a single NextSeq 500/550 v2 flow cell (Illumina), with 75bp single read and 6bp index read. Fastq files were demultiplexed using the User-friendly Transcriptome Analysis Pipeline (UTAP) developed at the Weizmann Institute (34). Sequences were mapped to the human genome using Qiagen GeneGlobe analysis web tool.

medRxiv preprint doi: https://doi.org/10.1101/2020.01.22.20018408; this version posted January 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Statistical analysis and machine learning

Plasma samples with ≥40,000 total miRNA UMIs were included. miRNA with average abundance of ≥100 UMIs per sample, across all samples, were considered above noise levels. miRNA NGS data was analyzed via DESeq2 package in R Project for Statistical Computing (35, 36), under the assumption that miRNA counts followed negative binomial distribution and data were corrected for library preparation batch in order to reduce its potential bias. Ratio of normalized FTD counts to the normalized control counts presented after logarithmic transformation on base 2. *P* values were calculated by Wald test (36, 37) and adjusted for multiple testing according to Benjamini and Hochberg (38). For binary classification by miRNAs, receiver operating characteristic (ROC) curves for individual miRNAs or combinations of miRNAs were plotted based on voom transformation of gene expression data in R (39). Graphs were generated with GraphPad Prism 5.

Machine learning approach was performed to build a Gradient Boosting classifier with 136 miRNAs. Cohorts were merged and case-control number imbalance was mitigated by applying ADASYN algorithm (https://imbalanced-learn.readthedocs.io/en/stable/api.html), which simulates synthetic new healthy sample data from the existing data. Then, K-Fold cross validation was performed on the pooled data set with K=3. An ROC was generated for each of the three folds and individual and mean AUCs were calculated.

medRxiv preprint doi: https://doi.org/10.1101/2020.01.22.20018408; this version posted January 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Acknowledgements

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 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

medRxiv preprint doi: https://doi.org/10.1101/2020.01.22.20018408; this version posted January 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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

Competing interests

The authors state that they have no competing interests.

medRxiv preprint doi: https://doi.org/10.1101/2020.01.22.20018408; this version posted January 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

References

1. Harvey RJ, Skelton-Robinson M, and Rossor MN. The prevalence and causes of dementia in people under the age of 65 years. *J Neurol Neurosurg Psychiatry.* 2003;74(9):1206-9.
2. Rabinovici GD, and Miller BL. Frontotemporal lobar degeneration: epidemiology, pathophysiology, diagnosis and management. *CNS Drugs.* 2010;24(5):375-98.
3. Snowden JS, Neary D, and Mann DM. Frontotemporal dementia. *Br J Psychiatry.* 2002;180(140-3.
4. Lee G, and Leugers CJ. Tau and tauopathies. *Prog Mol Biol Transl Sci.* 2012;107(263-93.
5. 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.
6. Neumann M, Rademakers R, Roeber S, Baker M, Kretzschmar HA, and Mackenzie IR. A new subtype of frontotemporal lobar degeneration with FUS pathology. *Brain.* 2009;132(Pt 11):2922-31.
7. Neumann M, Roeber S, Kretzschmar HA, Rademakers R, Baker M, and Mackenzie IR. Abundant FUS-immunoreactive pathology in neuronal intermediate filament inclusion disease. *Acta Neuropathol.* 2009;118(5):605-16.
8. Fontana F, Siva K, and Denti MA. A network of RNA and protein interactions in Fronto Temporal Dementia. *Front Mol Neurosci.* 2015;8(9.
9. Sieben A, Van Langenhove T, Engelborghs S, Martin JJ, Boon P, Cras P, De Deyn PP, Santens P, Van Broeckhoven C, and Cruts M. The genetics and neuropathology of frontotemporal lobar degeneration. *Acta Neuropathol.* 2012;124(3):353-72.
10. van Swieten J, and Spillantini MG. Hereditary frontotemporal dementia caused by Tau gene mutations. *Brain Pathol.* 2007;17(1):63-73.
11. Pottier C, Bieniek KF, Finch N, van de Vorst M, Baker M, Perkersen R, Brown P, Ravenscroft T, van Blitterswijk M, Nicholson AM, et al. Whole-genome sequencing reveals important role for TBK1 and OPTN mutations in frontotemporal lobar degeneration without motor neuron disease. *Acta Neuropathol.* 2015;130(1):77-92.
12. 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.
13. 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.
14. 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.
15. 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.
16. 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.
17. Skillback T, Mattsson N, Blennow K, and Zetterberg H. Cerebrospinal fluid neurofilament light concentration in motor neuron disease and frontotemporal dementia predicts survival. *Amyotroph Lateral Scler Frontotemporal Degener.* 2017;18(5-6):397-403.
18. 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.
19. Eitan C, and Hornstein E. Vulnerability of microRNA biogenesis in FTD-ALS. *Brain Res.* 2016.
20. Magen I, Coenen-Stass A, Yacovzada NS, Grosskreutz J, Lu C-H, Greensmith L, Malaspina A, Fratta P, and Hornstein E. Circulating miR-181a-5p is a prognostic biomarker for amyotrophic lateral sclerosis. *bioRxiv.* 2019:833079.

medRxiv preprint doi: https://doi.org/10.1101/2020.01.22.20018408; this version posted January 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

21. 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.

22. 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.

23. 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. *J Alzheimers Dis.* 2018;65(2):455-64.

24. 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.

25. 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. *J Neurol Neurosurg Psychiatry.* 2018;89(8):851-8.

26. Al-Hameed S, Benaissa M, Christensen H, Mirheidari B, Blackburn D, and Reuber M. A new diagnostic approach for the identification of patients with neurodegenerative cognitive complaints. *PLoS One.* 2019;14(5):e0217388.

27. Baek S, Tsai CA, and Chen JJ. Development of biomarker classifiers from high-dimensional data. *Brief Bioinform.* 2009;10(5):537-46.

28. 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.

29. 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.

30. 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.

31. 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.

32. Hodges JR, and Piguet O. Progress and Challenges in Frontotemporal Dementia Research: A 20-Year Review. *J Alzheimers Dis.* 2018;62(3):1467-80.

33. 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.

34. 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.

35. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria. *R Foundation for Statistical Computing* 2015.

36. 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.

37. Anders S, and Huber W. Differential expression analysis for sequence count data. *Genome Biol.* 2010;11(10):R106.

38. Benjamini Y, Drai D, Elmer G, Kafkafi N, and Golani I. Controlling the false discovery rate in behavior genetics research. *Behav Brain Res.* 2001;125(1-2):279-84.

39. Law CW, Chen Y, Shi W, and Smyth GK. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* 2014;15(2):R29.

medRxiv preprint doi: https://doi.org/10.1101/2020.01.22.20018408; this version posted January 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

medRxiv preprint doi: https://doi.org/10.1101/2020.01.22.20018408; this version posted January 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Magen et al Table 1

	Cohort I		Cohort II	
	Control	FTD	Control	FTD
Number of subjects (% males)	21 (62%)	52 (60%)	35 (51%)	117 (68%)
Age at enrolment	52±4 yr.	65±1 yr.	65±1 yr.	66±1 yr.
Age of onset (1st reported symptoms)	N/A	61±1 yr.	N/A	60±1 yr.
Disease duration at enrolment	N/A	4.5±0.3 yr	N/A	6±0.3 yr
Clinical subtype (bvFTD/PNFA/SD/FTD-ALS/others)	N/A	25/15/8/2/2	N/A	57/25/20/3/12
Mutation carriers (C9ORF72/MAPT/ GRN/TBK1)	N/A	6/4/8/1	N/A	12/10/5/1
Likely pathology (TDP-43/Tau)	N/A	23/4	N/A	40/15

Table 1. Summary of demographic and clinical characteristics of FTD Cohorts I and II and control samples. bvFTD: behavioural FTD; PNFA: progressive nonfluent aphasia; SD: semantic dementia

medRxiv preprint doi: https://doi.org/10.1101/2020.01.22.20018408; this version posted January 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Magen et al Table 2

	Controls	ALS	MS
Number of subjects (% males)	103 (27.2%)	115 (64.3%)	17 (29.4%)
Age at enrolment	52±2 yr.	64±1 yr.	57±2 yr.

Table 2. Summary of demographic and clinical characteristics of ALS, MS and control samples. ALSFRS-R: ALS functional rating scale. Demographic data recognizes that male ratio and age of first phlebotomy was significantly different between ALS cohort I and controls (proportion test: $p<0.0001$; t-test: $p<0.001$, respectively).

medRxiv preprint doi: https://doi.org/10.1101/2020.01.22.20018408; this version posted January 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Magen et al Figure 1

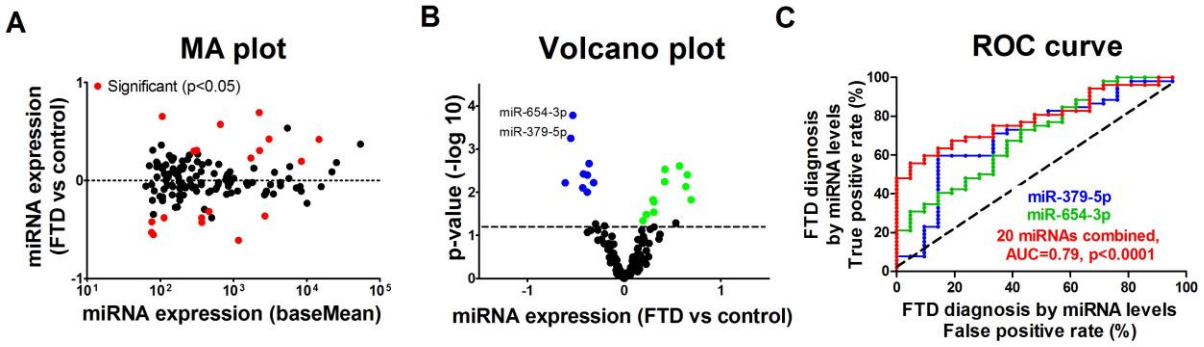

Figure 1. Predictive value of differential miRNA expression in FTD plasma. (A) MA plot of differential miRNA expression in FTD (n=52) and heathy control (n=21 plasma samples). Log-2 transformed fold-change, against the mean miRNA abundance. Red -significantly changed miRNAs (p-value ≤0.05). (B) A volcano plot of differentially expressed miRNAs between FTD (n=52) and heathy control (n=21 plasma samples). Each dot represents a single miRNA, plotted according to log 2 fold-change (FC) in FTD vs control (X-axis), and the negative log 10 transformation of p-value (Y-axis). Black horizontal line demarcates p<0.05 and dots denote miRNAs with statistically significant differential expression in FTD plasma; green - increased in FTD; blue - decreased in FTD. (C) Receiver operating characteristic (ROC) curves demonstrate the capacity of miR-379-5p (blue, AUC=0.71, p=0.005), miR-654-3p (green, AUC=0.71, p=0.006) and of a combinatorial signature of 20 miRNAs, whose differential expression is significant (red, AUC=0.79, p<0.0001; miRNAs listed in Table S1), to distinguish between FTD and healthy controls. True positive rate (sensitivity) as a function of the false positive rate (1-specificity) for different cut-off values. P-values are calculated given null hypothesis of area under the curve (AUC) =0.5.

medRxiv preprint doi: https://doi.org/10.1101/2020.01.22.20018408; this version posted January 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Magen et al Figure 2

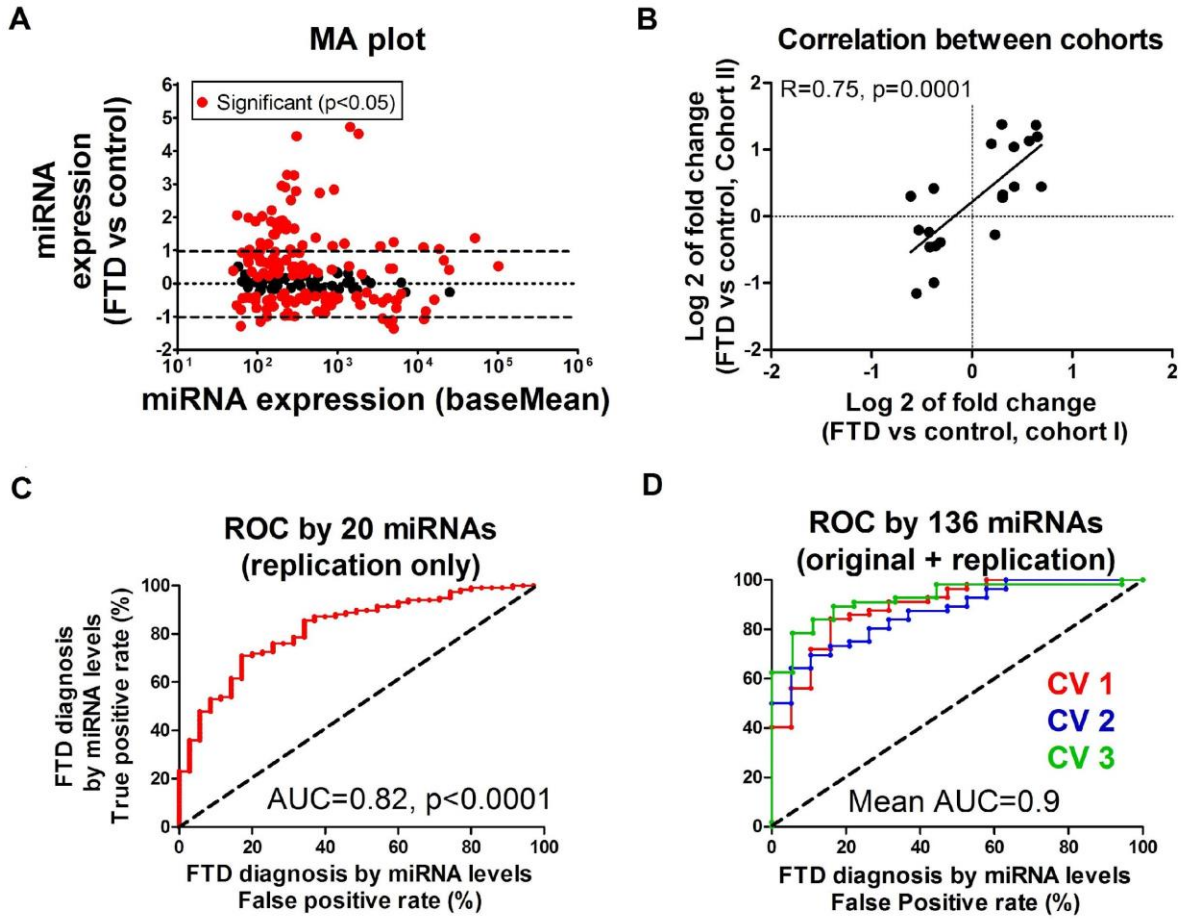

Figure 2. Replication of miRNA signature in a second cohort. (A) MA plot of differential miRNA expression in a second cohort of FTD (n=117) and a second heathy control cohort (n=35 plasma samples). Log-2 transformed fold-change, against the mean miRNA abundance. Red - significantly changed miRNAs (p-value ≤0.05). **(B)**. Scatter plot of correlation between fold change of 20 miRNAs, which classify FTD, between the first and second cohorts. **(C)** ROC curve of a combinatorial signature of 20 miRNAs, whose differential expression is significant in the first cohort, distinguishes between FTD and healthy controls of the second cohort. True positive rate (sensitivity) as a function of the false positive rate (1-specificity) for different cut-off values. P-values are calculated given null hypothesis of area under the curve (AUC) =0.5. **(D)** ROC curves of Gradient Boosting Classifier based on K-fold cross validation with K=3 for a merged data set of 136 miRNA expression in both the discovery and the replication cohort (169 FTD cases and 56 healthy controls). Red**,** ROC for fold 1; blue, ROC for fold 2; green, ROC for fold 3.

medRxiv preprint doi: https://doi.org/10.1101/2020.01.22.20018408; this version posted January 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Magen et al Figure 3

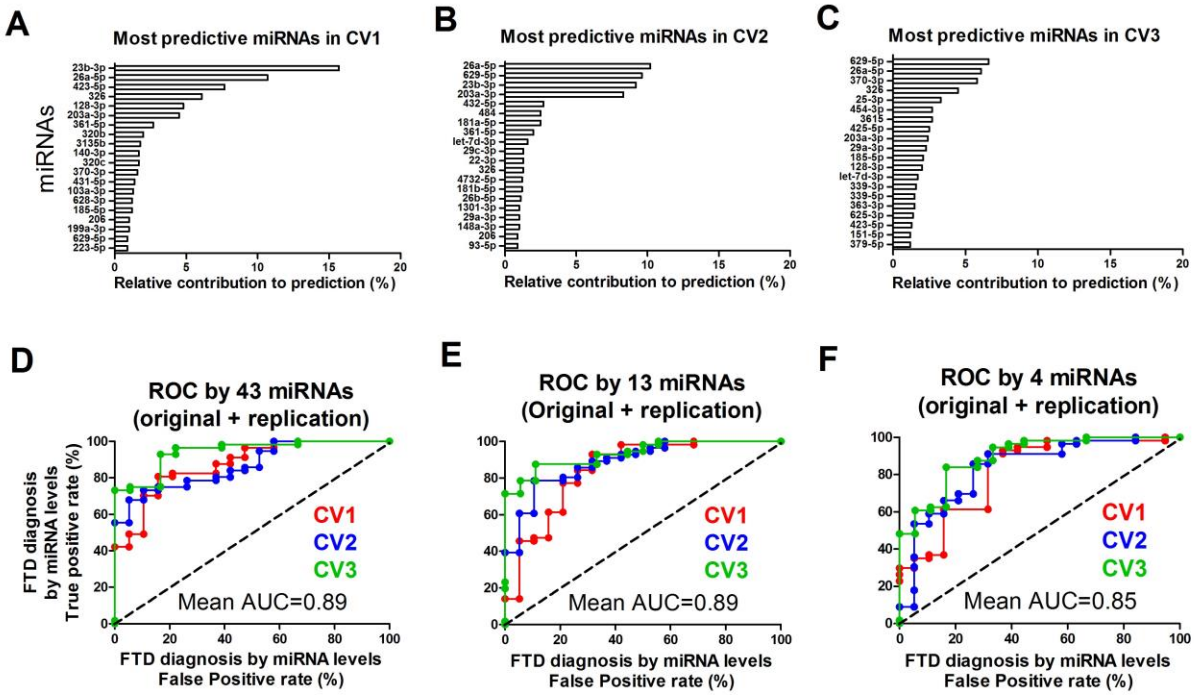

Figure 3. Most important miRNA predictors in gradient boosting classifier. Top 20 miRNA classifiers in (**A**) cross-validation fold 1 (**B**) fold 2 and **(C)** fold 3**. (D-F)** ROC curves based on K-fold cross validation with K=3 for a merged data set including both the discovery and the replication cohort (169 FTD cases and 56 healthy controls), when **(D)** ROC curves of Gradient Boosting Classifier with only 43 miRNAs depicted as "top 20" in at least one fold (all of the miRNAs shown in panels A-C) are selected for classification. **(E)** 13 miRNAs depicted as "top 20" in at least two folds are selected and **(F)** only the four miRNAs depicted as "top 20" in all three folds are selected. Red**,** ROC for fold 1; blue, ROC for fold 2; green, ROC for fold 3.

medRxiv preprint doi: https://doi.org/10.1101/2020.01.22.20018408; this version posted January 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Magen et al Figure 4

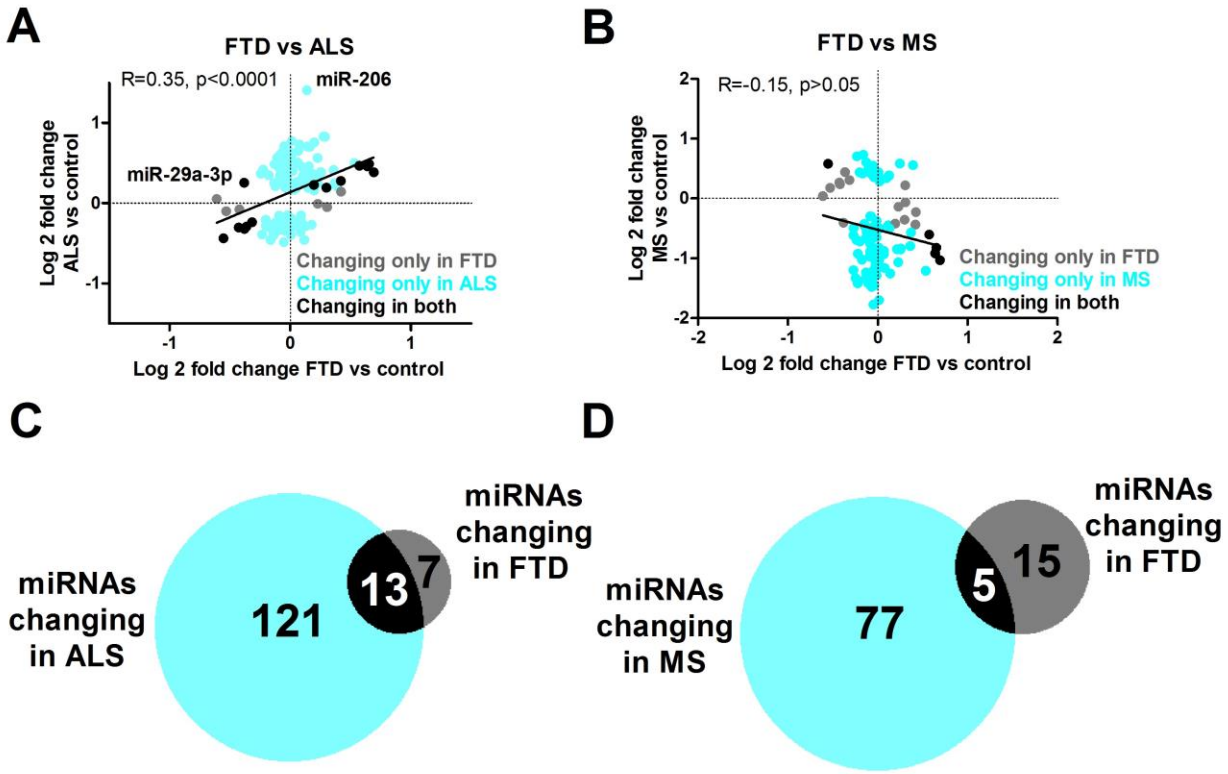

Figure 4. Differential miRNA expression in ALS vs. FTD plasma. (A) Scatter plot of correlation between ALS/control ratio and FTD/control ratio. (B) Scatter plot of correlation between MS/control ratio and FTD/control ratio. (C) Venn diagram of comparison between ALS and FTD with 13 shared miRNAs in the black portion. (D) Venn diagram of comparison between MS and FTD with 5 shared miRNAs in the black portion.

medRxiv preprint doi: https://doi.org/10.1101/2020.01.22.20018408; this version posted January 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Supplementary figures

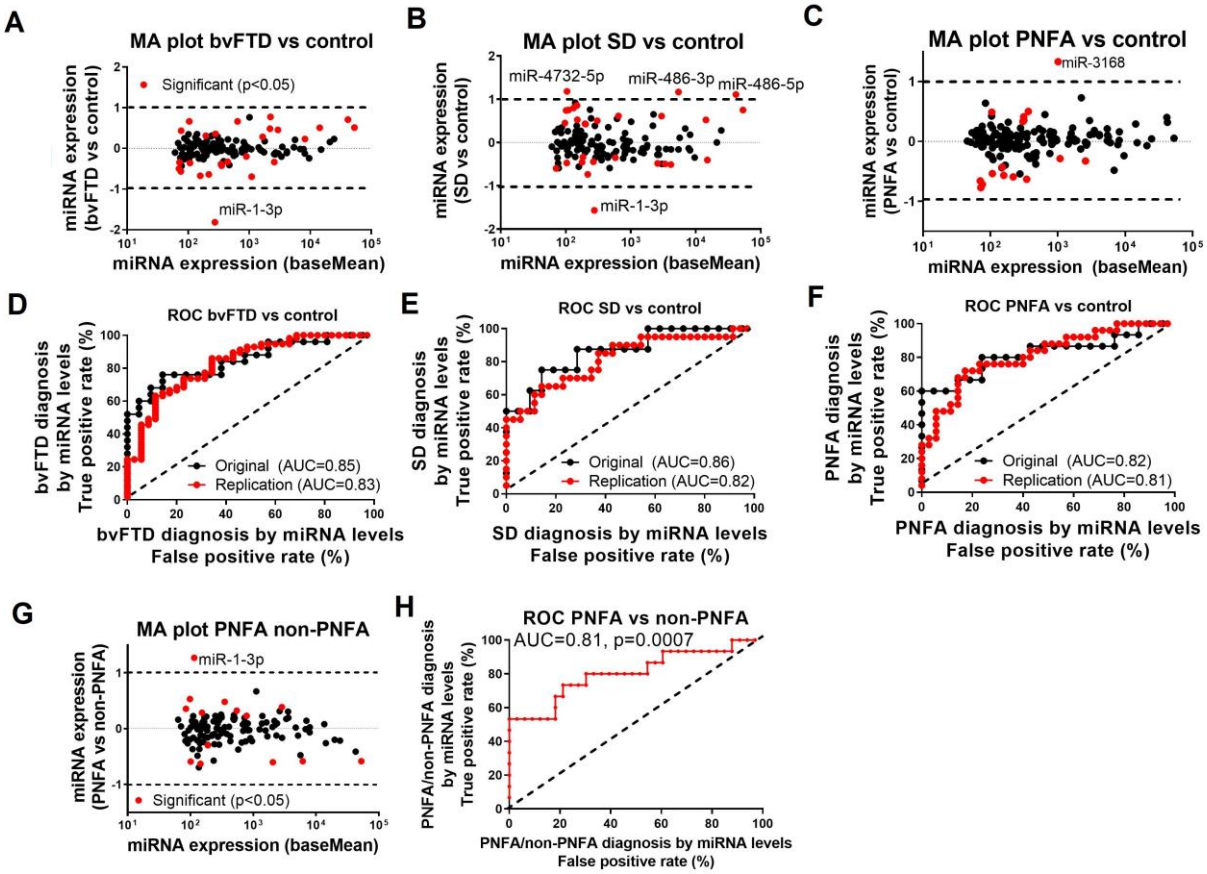

Figure S1. miRNA profiles of FTD subtypes. (A) MA plots of differential miRNA expression in bvFTD patients (n=25), (B) SD patient (n=8) and (C) PNFA patients (n=15) compared to healthy controls (n=21). Log-2 transformed fold-change, against mean miRNA abundance. Red - significantly changed miRNAs (p-value ≤0.05). ROC curve of a combinatorial signature of 20 miRNAs, whose differential expression is significant in the first cohort, distinguishes **(D)** bvFTD **(E)** SD and **(F)** PNFA from healthy controls, in the first (black) and second (red) studies. True positive rate (sensitivity) as a function of the false positive rate (1-specificity) for different cut-off values. P-values are calculated given null hypothesis of area under the curve (AUC) =0.5. **(G)** MA plot for differential miRNA expression between PNFA (n=15) and non-PNFA FTD cases (bvFTD + SD (n=33)). Red- significantly changed miRNAs (p<0.05). **(H)** ROC curve based on combinatorial signature of 14 significant miRNAs for distinguishing PNFA and non-PNFA FTD cases.

medRxiv preprint doi: https://doi.org/10.1101/2020.01.22.20018408; this version posted January 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

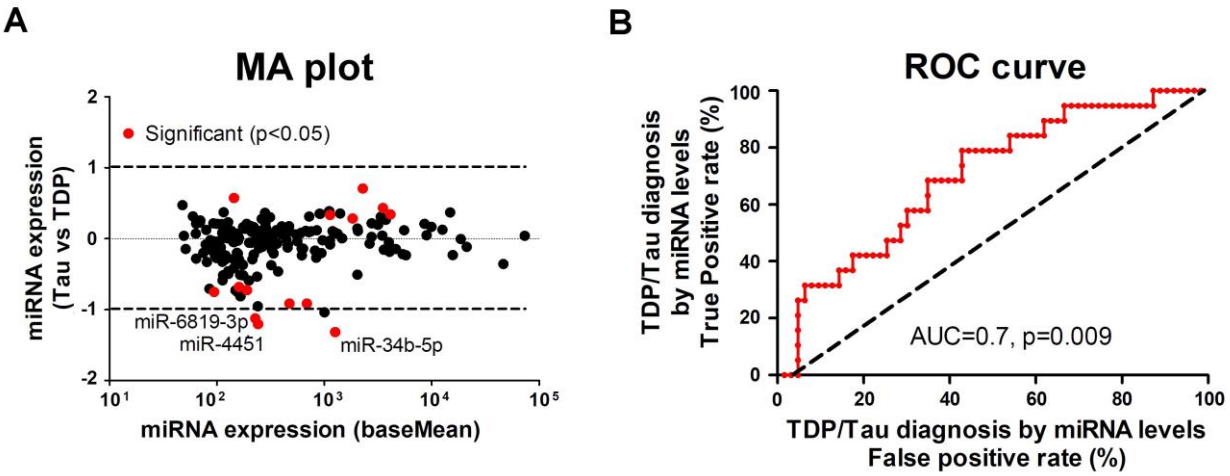

Figure S2. miRNA profiles associated with FTD neuropathologies. (A) MA plots of differential miRNA expression in FTD patients with predicted Tau pathology (n=19) vs patients with predicted TDP-43 pathology (n=63) from both cohorts used in the study. Log-2 transformed fold-change, against mean miRNA abundance. Red -significantly changed miRNAs (p-value ≤0.05). **(B)** ROC curve based on combinatorial signature of 14 significant miRNAs for distinguishing FTD cases with Tau pathology from those with TDP-43 pathology.