Parkinsonism and Related Disorders 65 (2019) 203–209

Contents lists available at ScienceDirect

Parkinsonism and Related Disorders

journal homepage: www.elsevier.com/locate/parkreldis

TRNA-derived fragments as sex-dependent circulating candidate biomarkers for Parkinson's disease

Rogan Magee, Eric Londin, Isidore Rigoutsos*

Computational Medicine Center, Jefferson Alumni Hall #M81, Sidney Kimmel Medical College, Thomas Jefferson University, 1020 Locust Street, Philadelphia, PA, 19107, USA

ARTICLE INFO

Keywords:
Parkinson's disease
Transfer RNA
tRNA
tRNA-derived fragments
tRF
i-tRF
5′-tRF
3′-tRF
tRNA halves
tRHs
Biomarkers
Parkinson's disease biomarkers

ABSTRACT

Introduction: Parkinson's Disease (PD) is diagnosed clinically. Reliable non-invasive PD biomarkers are actively sought. Transfer RNAs produce short non-coding RNAs, the tRNA-derived fragments (tRF). tRF have been shown to play diverse roles, including in amyotrophic lateral sclerosis, and the response to ischemic stroke. Rich tRF populations are being reported in biofluids. We explored the possibility that tRF can serve as non-invasive biomarkers for PD.

Methods: We collected existing RNA-seq samples and re-analyzed a total of 254 legacy datasets from 3 previous studies, from male and female PD patients and controls that belong to three categories: prefrontal cortex samples from 29 patients and 33 controls; cerebrospinal fluid (CSF) samples from 63 patients and 64 controls; and, serum samples from 34 patients and 31 controls. First, we identified tRF exhaustively and deterministically in every dataset. Second, we determined tRF that are differentially abundant (DA) between PD and control samples, using uncorrected *t*-tests. Lastly, we assessed all the DA tRF from the previous step with Partial Least Squares – Discriminant Analysis (PLS-DA) to stringently sub-select tRF that can distinguish PD patients from controls.

Results: We show that PLS-DA identified tRF from prefrontal cortex, CSF, and serum that can distinguish PD patients from controls. A handful of identified tRF were previously investigated in neurological contexts. Signatures built from relatively few tRF suffice to distinguish PD from control in each category of samples with high sensitivity (89-100%) and specificity (79-98%).

Conclusion: tRF-based signatures are promising candidates that warrant further evaluation as non-invasive PD biomarkers.

1. Introduction

tRNA halves (tRHs), the first reported category of tRNA-derived fragments (tRF) [1,2], are produced in response to cellular stress. The endonuclease Angiogenin (*ANG*) has been linked to production of 5′-tRHs and 3′-tRHs from mature tRNAs in several contexts [1,3,4]. In recent years, several more structural categories of tRF were identified. They include: 5′-tRF, whose 5′ termini are the 5′ ends of the mature tRNA; and 3′-tRF, whose 3′ termini are the 3′ end of the mature tRNA [5]. Our work identified a fifth structural category of tRF, known as "internal tRF" (i-tRF). i-tRF can begin and end anywhere along the mature tRNA sequence [6].

In neurologic contexts, tRHs were first described in amyotrophic lateral sclerosis (ALS), where they provide a protective role for motor neurons. Either loss of protective tRHs or interaction between protective tRHs and pathogenic RNA repeats may contribute to the pathogenesis of ALS [7]. tRF were also shown to be active in a vascular response to ischemia in the rat brain [8]. Finally, mutations in CLP1 cause production of tRF that sensitize neurons to p53-mediated oxidative damage, resulting in a rare familial neurodegenerative disorder [9].

With regard to PD, *ANG* is linked to increased risk for sporadic PD [10]. Sporadic *ANG* mutations observed in ALS and PD affect its ribonuclease activity. Experiments in mice demonstrate that alpha-synuclein decreases the activity of ANG in a dose dependent manner [11,12]. Additionally, these experiments show that angiogenin is neuroprotective in a chemical model of PD, in a manner independent of Akt signaling. Thus, whether through mutation or functional inhibition, the normal production of tRHs by ANG may be lost in PD.

Currently, PD is diagnosed by movement disorder specialists [13,14]. Several studies have explored biomarkers, from serum alpha-synuclein to CSF cytokines. One of the most promising candidates is

* Corresponding author.
E-mail address: isidore.rigoutsos@jefferson.edu (I. Rigoutsos).

https://doi.org/10.1016/j.parkreldis.2019.05.035
Received 4 December 2018; Received in revised form 21 May 2019; Accepted 23 May 2019

serum uric acid [13,15,16]. Low uric acid levels correlate with risk for PD. Serum uric acid may proxy an individual's capacity to compensate for oxidative stress [13]. The molecular cause of PD is debated [14,17,18]. Gene knockout and environmental exposure models support the idea that PD involves improper management of cellular stresses - due to SNc-specific high energy demands - by mitochondrial mechanisms, protein turnover, and vesicular trafficking [19,20].

Because of the links between tRHs, stress responses and PD, we hypothesized that tRF exhibit differential abundances in PD, and that some of the tRF behind those differences may manifest themselves in biofluids. Below, we leverage MINTmap [21], Threshold-seq [22], and MINTbase [23] to determine whether tRF in prefrontal cortex, CSF, or serum can serve as PD biomarkers. A PD-specific serum signature raises the possibility of devising a non-invasive biomarker.

2. Methods

2.1. Datasets

We re-analyzed short RNA-sequencing data that were generated by previous research efforts. We downloaded the data from two repositories at the National Institutes of Health (NIH): the Gene Expression Omnibus (GEO) and the database of Genotypes and Phenotypes (dbGaP). Specifically, we downloaded: datasets corresponding to prefrontal cortex samples from 29 PD patients (NIH GEO #GSE72962 [24]) and 33 controls (NIH GEO #GSE64977 [25]); datasets corresponding to CSF samples from 67 PD patients and 69 controls (NIH dbGap #phs000727 [26]); and, datasets corresponding to serum samples from 61 PD samples and 71 controls (NIH dbGap #phs000727 [26]). Data were acquired from the GEO repository without any restrictions. Data were acquired from the dbGap repository by Eric Londin, with an accompanying statement of purpose submitted and approved as of November 19, 2014.

2.2. Patient diagnosis and characteristics

Patients in the dbGap study were enrolled in the Banner Sun Health Research Institute (BSHRI) Brain and Body Donation Program. Samples in the NIH GEO studies were obtained from the BSHRI National Brain and Tissue Resource for PD and Related Disorders, Harvard Brain and Tissue Resource Center (Belmont, MA), and Human Brain and Spinal Fluid Resource (Los Angeles, CA).

All NIH GEO patients were negative for known GBA and LRRK2 mutations. PD diagnosis in these studies was made on the basis of histopathologic criteria at the time of enrollment of the tissue samples. In the dbGap study, diagnoses were subsequently confirmed prior to the study [26]. Dementia was diagnosed by enrolling physicians using chart review.

2.3. Preprocessing of sequenced reads and tRF profiling

First, we processed samples using the *cutadapt* (v1.5) tool [27]. Briefly, *cutadapt* trims low quality bases from the 5′ and 3′ ends of reads. Then, *cutadapt* removes 5′ and 3′ adapters. Additionally, we filtered out reads in which adapters could not be located. We then used *MINTmap* to profile and quantify tRF [21,28,29]. *MINTmap* uses only exact matches and distinguishes between *exclusive* tRF (whose sequences can only be found among the known tRNA loci of the human genome) and *ambiguous* tRF (whose sequences can be found among the known tRNA loci but also elsewhere in the human genome). We used *Threshold-seq* [22] to calculate an adaptive threshold for each sample. Threshold-seq calculates a different threshold for each sample, by determining the level above which raw sequence read events are unlikely to be noise and are more likely to represent biologically relevant RNAs. The threshold is adaptive in that it is calculated using the unique distribution of read counts in each sample. Finally, for all tRF in each sample that exceeded the *Threshold-seq* threshold, we normalized their abundance to sample read depth using a “reads per million” (RPM) calculation.

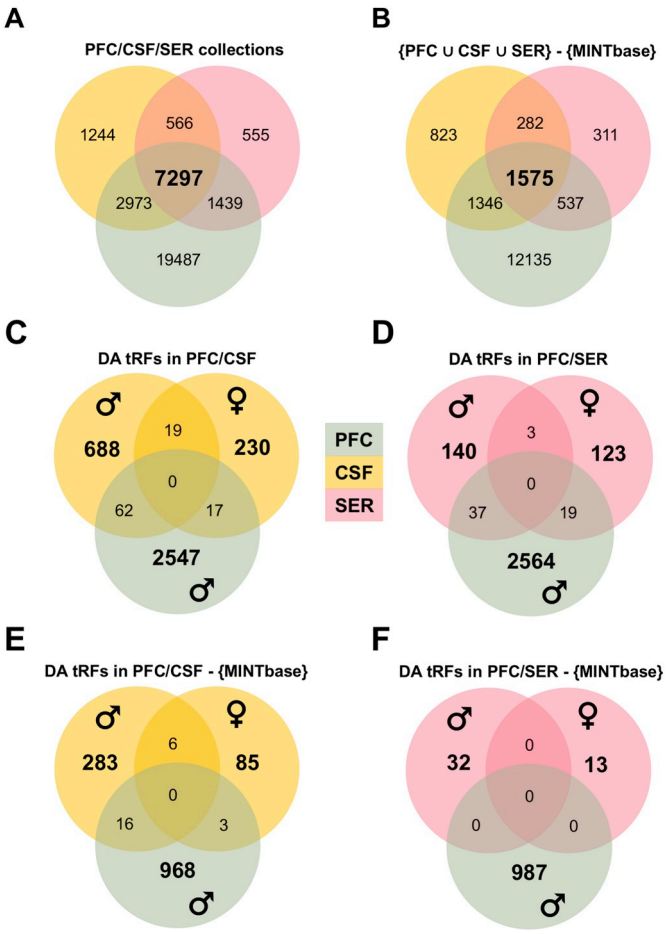


Fig. 1. tRF identified in the prefrontal cortex, CSF and serum collections. We used the MINTmap and Threshold-seq algorithms to identify the set of tRF recovered from all prefrontal cortex (PFC in figure) samples, all CSF samples, and all serum (SER in figure) samples. (A) The Venn diagram describes the overlap of tRF discovered in these three collections. (B) We recompute the overlap shown in (A) after filtering out tRF that are contained in the most recent Release of MINTbase and which have been reported in other tissue contexts. We then used *t*-tests to identify tRF that are DA between control and PD samples in prefrontal cortex, CSF, and serum samples. DA tRF are those with a difference in means per a *t*-test with p-value ≤ 0.05. The shown Venn diagrams capture the overlap between tRF identified in each context (C-F). Each panel describes the overlap of DA tRF when splitting samples by patient sex (male or female) and sample tissue type (prefrontal cortex, CSF, or serum), as indicated. Green circles represent prefrontal cortex samples. Gold circles represent CSF samples. Red circles represent serum samples. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

2.4. Sample quality control

Extracting RNA from biofluids can be challenging. Earlier work [26] noted the instability of miRNA profiles derived from CSF and serum samples. In our analyses, we observed a positive correlation between the number of reads after using *cutadapt* on the CSF and serum collections and the number of reads identified as tRF by MINTmap (Supp. Fig. 1A–B). The correlation breaks down in those samples where fewer reads survive our preprocessing. Looking at the average numbers of reads that land in the tRNA space allows us to determine the minimum number of reads needed to identify all unique tRF. This appears to be ≥ 100,000 reads mapping to tRF (Supp. Fig. 1C–D). Importantly,

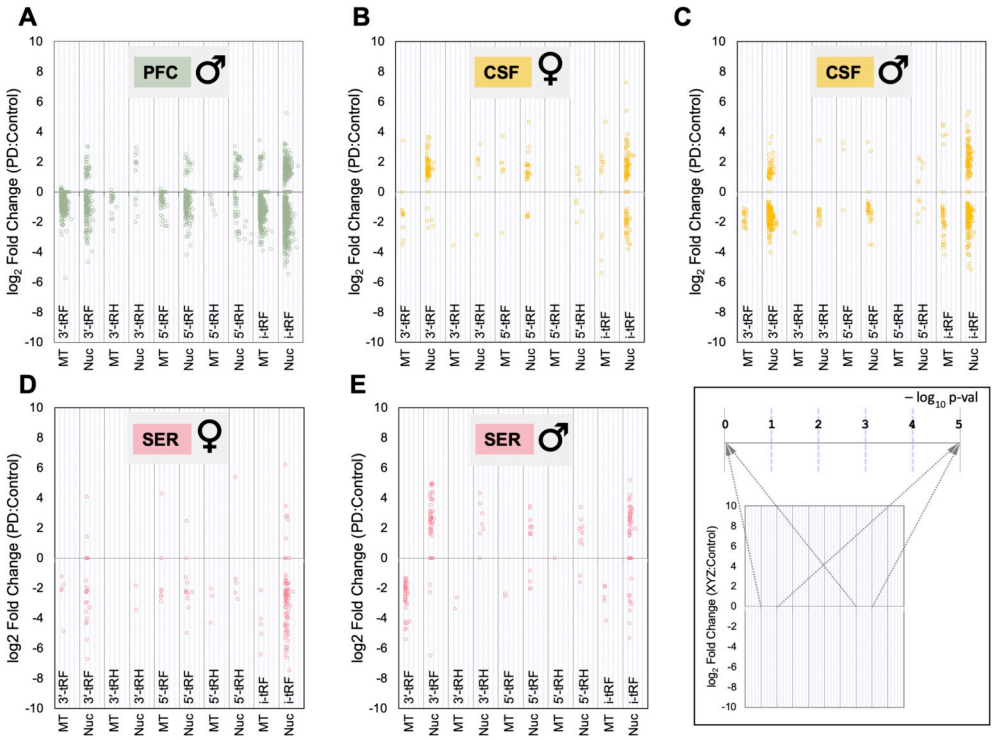


Fig. 2. DA tRF in the prefrontal cortex, CSF, and serum collections. We identified DA tRF as described in Fig. 1. Here, we plot the results, showing fold changes and p-values, in combination with a breakdown by tRF type. Specifically, we plot the $\log_2$ fold change (PD:CNTRL) for each DA tRF in each sample set. Circles are displaced from the vertical line according to their p-value, with lower p-values being further shifted to the right (up to a minimum observed p-value of $\sim 10^{-5}$) (see explanation panel for an illustration). We plot the results for prefrontal cortex male comparisons (A), CSF female comparisons (B), CSF male comparisons (C), serum female comparisons (D), and serum male comparisons (E). Green circles refer to prefrontal cortex samples. Gold circles refer to CSF samples. Red circles refer to serum samples. As can be seen, many tRF exhibit considerable fold changes that are also statistically-significant. Each of the shown probability intervals spans five $\log_{10}$ units – see detailed plot in the bottom right of the Figure. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

enforcing this minimum required number of reads allows us to confidently identify samples in which tRF are either present at very low levels, or not expressed at all. Clinical characteristics for samples from both studies that passed quality control are listed in Supp. Table 1. All five structural categories mentioned in the Introduction are represented among the identified tRF (Supp. Material and Supp. Figure 2).

2.5. Sequence batch correction

Prefrontal cortex samples were sequenced in three annotated batches (Supp. Table 1). To account for and correct possible batch effects, we processed RPM expression data using *ComBat* [30] from Bioconductor (release 3.6). *ComBat*-processed reads were used in all of our downstream analyses. Sequencing batch information was not available for CSF and serum samples.

2.6. Statistical analyses

All analyses were run in R version 3.3.0. We used two-tailed *t*-tests to identify differentially-abundant tRF (p-value ≤ 0.05). In this first step, we did not correct for multiple testing, in order to retain the greatest number of tRF possible for use in downstream classifier model construction. To construct classifier models, we used the *DiscriMiner* package for Partial Least Squares-Discriminant Analysis (PLS-DA). PLS-DA is a method of sample classification that first transforms the observable and predicted variables by using partial least squares.

We used Monte Carlo simulations to evaluate PLS-DA model performance. For each collection (prefrontal cortex, CSF, serum), we randomly split the control and disease samples of the collection into a *training* (60%) and a *test* (40%) set. We built a classifier using the *training* set and assessed it using the respective *test* set. We repeated this process of random splitting for a total of N^2 times where N is the total number of control and disease samples in the collection. We then tallied true positives, true negatives, false positives, and false negatives to compute sensitivity and specificity. We did this by counting the number of times each sample in each collection was classified as control or PD, and comparing this to the known classification of the sample.

3. Results

We worked with legacy datasets corresponding to samples from three collections (prefrontal cortex - PFC in figures, CSF, and serum - SER in figures) that we processed and re-analyzed as described in Methods. As noted in the original publication [24], NIH GEO PD patients were significantly older at death (77.6 years ± 8.96) than control patients (68.1 years ± 14.9) (*t*-test, p-value ≤ 0.05). The dbGap samples were age-matched for both CSF (control: 81.6 years ± 10.1, PD: 80.0 years ± 5.13) and serum (control: 78.5 years ± 11.1, PD: 80.1 years ± 5.53) samples. Enforcing quality control criteria left us with 62 prefrontal cortex, 127 CSF, and 65 serum samples from PD patients and controls (CNTRL in figures). These samples contain enough tRF reads for us to confidently identify tRF that are truly absent from these samples, and not absent due to the quality of sequencing. Note that all prefrontal cortex samples were obtained from male patients.

Information on the duration of disease was available for 21 of 29 prefrontal cortex samples, and all CSF and serum samples. We note that duration of disease was not significantly different across the three sample types. Duration of disease was as follows: prefrontal cortex: 10.5 years ± 6.5; CSF: 12.7 years ± 8.0; and, serum: 10.6 ± 8.4. Further, we note that motor onset of disease was marked in 21 of 29 prefrontal cortex samples (66.5 years ± 9.8), occurring roughly 10 years on average prior to death. Age at motor onset did not significantly differ between prefrontal cortex patients who developed dementia (69.8 years ± 12.2) and those who did not (64.1 years ± 7.2). Duration of disease did not significantly differ either between prefrontal cortex patients who developed dementia (9.2 years ± 6.7) and those who did not (11.5 years ± 6.4). Additional clinical characteristics for these samples are listed in Supp. Table 1.

3.1. tRF abundance profiles differ among prefrontal cortex, CSF and serum

Using *MINTmap* and *Threshold-seq*, we identified 33,561 unique tRF across the three categories of samples (Fig. 1A). These are denoted by "{PFC ∪ CSF ∪ SER}" in Fig. 1. 31,196 tRF were contributed by the prefrontal cortex samples, 12,608 by the CSF samples, and 9,857 by the

serum samples (Supp. Table 2). 58% of all these tRF are unique to the prefrontal cortex samples. To gauge the tissue-specificity of these tRF, we recalculated the Venn diagram of Fig. 1A after excluding all tRF that are contained in MINTbase [23,29], currently the largest repository of tRF from human samples (Fig. 1B). Release 2.0 of MINTbase contains 26,744 distinct tRF (here denoted by "{MINTbase}") from 11,721 samples. MINTbase's samples come from The Cancer Genome Atlas (TCGA), the 1000 Genomes Project, and multiple other studies, both non-cancer and cancer [29]. Several dozen human tissue contexts are represented in MINTbase including 535 lower grade glioma tissue samples that give rise to 7,405 tRF.

3.2. Multiple tRF are differentially abundant between control and PD in all three collections

We next identified differentially abundant (DA) tRF between healthy and PD states in each collection (*t*-test, p-value ≤ 0.05). Because we have reported that sex modulates tRF expression [6], we analyzed male and female samples separately. As mentioned above, during the first step we did not correct for multiple testing, in order to retain as many tRF as possible prior to filtering them with the more stringent PLS-DA method (see below).

The numbers of identified DA tRF and their overlaps are summarized in Fig. 1C–D. The sequences of these tRF and other information are described in Supp. Table 3. Again, we leveraged MINTbase [23,29] – see previous section. Fig. 1E–F shows how the overlaps change when we subtract all previously-seen tRF that are contained in MINTbase. In the CSF samples, 19 DA tRF are common to both males and females; of these, six are not in MINTbase. By contrast, in the serum samples, only three DA tRF are common to both males and females, and all three of them are in MINTbase.

DA tRF from the prefrontal cortex samples (all male donors) overlap with male comparisons in biofluids: 62 DA tRF, of which 16 are not in MINTbase, are common to both CSF and prefrontal cortex. Fig. 2 presents a panorama of the $\log_2$ fold changes between PD and control samples for all DA tRF. In each plot, individual points represent DA tRF, with the point's Y-value capturing the observed fold change ($\log_2$) and the X-value capturing the $-\log_{10}$ of the tRF's p-value. Each interval shows the findings separately for each tRF category and spans five $\log_{10}$ units.

3.3. Multiple tRF are differentially abundant in the context of dementia associated with PD

We next compared patients who had PD with dementia (PDD) and PD without dementia (PND). Fig. 3 describes DA tRF shared among PDD, PND and control in all three collections.

Of note, in the prefrontal cortex samples we find 980 DA tRF that are common to all three comparisons: control vs. PD, control vs. PND, and control vs. PDD. This indicates that these tRF are DA independently of the stage of disease (Fig. 3A). On the other hand, when we compare control vs. PD and control vs. PND, we find that 934 DA tRF are common to both. Lastly, when we compare control vs. PD and control vs. PDD, we find 512 tRF are DA and common to both (Fig. 3A). This suggests that the presence or absence of dementia is captured by a characteristically different profile of DA tRF.

Our analyses of CSF (Fig. 3B–C) and serum (Fig. 3D–E) samples produced analogous results. We emphasize that DA tRF between PND and PDD in CSF and in serum exhibit a very strong sex dependence (Fig. 3B–E). We also find tRF that are DA between PND and PDD yet are not captured by any of the other comparisons (Fig. 3F). Interestingly, these DA tRF are virtually unique to prefrontal cortex, CSF, and serum, respectively.

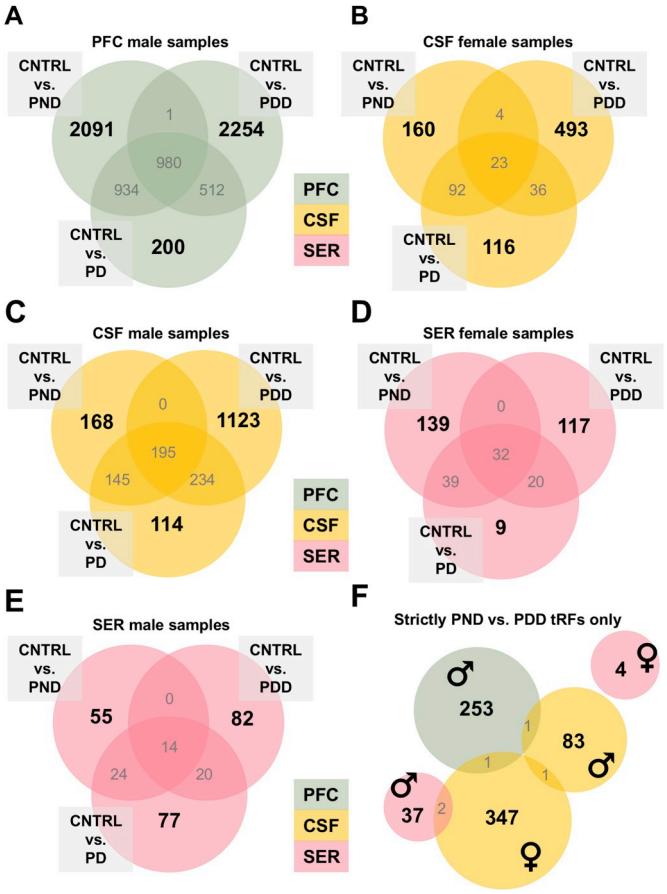
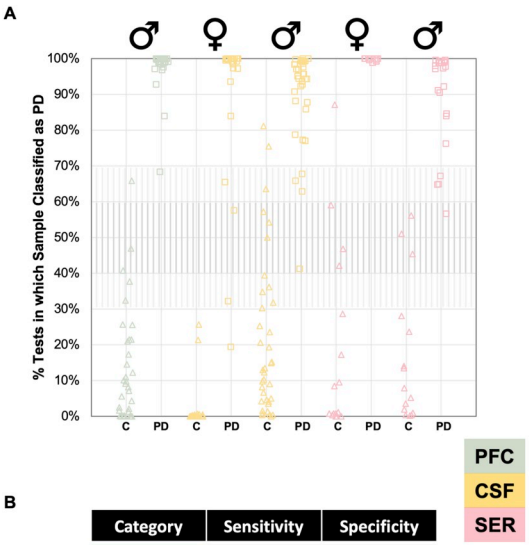


Fig. 3. tRF in prefrontal cortex, CSF, and serum samples that are differentially abundant between control, PND, and PDD. We used *t*-tests to identify the DA tRF. We carried out these analyses separately for prefrontal cortex, CSF, and serum. The Venn diagrams describe the overlap between DA tRF identified in each context (A-E) with those originally identified in Fig. 1C–F. Each panel describes the overlap of DA tRF when splitting samples by patient sex, and by collection (prefrontal cortex, CSF, or serum). Specifically, sample groups are plotted as follows: prefrontal cortex male (A), CSF female (B), CSF male (C), serum female (D), and serum male (E). Panel F shows the complement of DA tRF in PND vs. PDD comparisons. Specifically, we retained and plotted the intersections of DA tRF not part of the set of DA tRF from control (CNTRL in figure) vs. PD, control vs. PND, or control vs. PDD comparisons. Green circles represent prefrontal cortex samples. Gold circles represent CSF samples. Red circles represent serum samples. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

3.4. tRF could serve as candidate PD biomarkers

We next determined whether DA tRF could serve as biomarkers of PD (see Methods). We used a PLS-DA model approach using as input the DA tRF derived from the *t*-test of the first step (see above). Several biomarkers have been explored for PD diagnosis, including but not limited to: serum uric acid, DJ-1 isoforms, ApoA1, EGF, and interleukin-8 [14]. All of these promising biomarker candidates have also been investigated either in tissues outside the brain or in roles outside of PD, indicating that non-specific markers may confer diagnostic power. Thus, we retained tRF that have been discussed in other contexts, in order to maximize the potential for PLS-DA models to identify promising disease classifiers.

We evaluated the PLS-DA model by randomly assigning disease and control labels to the samples followed by training and testing over the course of N^2 iterations to determine the model's response to noise. When we did that performance decreased dramatically, indicating that the PLS-DA model had identified true signal (data not shown).



B

Category	Sensitivity	Specificity
Male PFC	98%	86%
Female CSF	90%	98%
Male CSF	91%	79%
Female SER	100%	81%
Male SER	89%	83%

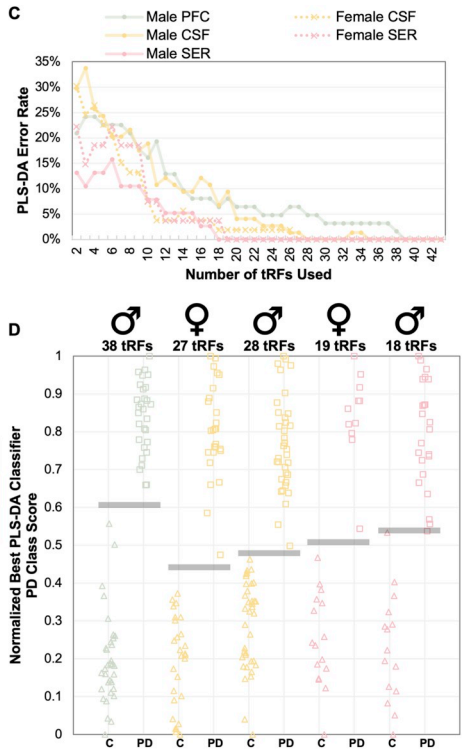


Fig. 4. Small signatures of tRF suffice to correctly distinguish control from PD samples. In panel A, we demonstrate the efficacy of classification of each sample throughout the N^2 PLS-DA simulations, by plotting the percentage of simulations in which each sample was classified as PD, when it was included in a simulation. We show this for all five comparisons and split control and PD samples, to show the relative performance of classification. We then count the total number of times each sample was correctly classified (true positive or true negative) and the number of times each sample was incorrectly classified (false positive or false negative). In panel B, we use the true positive, true negative, false positive, and false negative sums from the data in panel A, to calculate the sensitivity and specificity of the five models. To create panel C, we first summed the number of times the VIP score of each tRF used in panel A was ≥ 1.5 in a simulation, and ranked each sum in descending order, for all five comparisons. We then built PLS-DA models using 100% of samples and using the top X tRF by sum of VIP votes, from $X = 2$ to 44 (44 is the maximum number of tRF available from the female serum samples). Panel C shows the error rate of all models, for all five comparisons. Panel D shows the discriminant score calculated by the best performing PLS-DA model from panel C that uses the smallest number of tRF as a signature.

To determine how well each model classified unknown samples, we counted the number of times a given sample in a test set was classified as PD or control through the N^2 simulations in each of the five comparisons (Fig. 4A). Next, we compared these totals to the true identity (PD or control) of the sample and tallied correct and incorrect assignments that we used to calculate sensitivity and specificity (Fig. 4B). We find that across N^2 simulations, we classify male prefrontal cortex samples with 98% sensitivity and 86% specificity, female CSF samples with 90% sensitivity and 98% specificity, male CSF samples with 91% sensitivity and 79% specificity, female serum samples with 100% sensitivity and 81% specificity, and male serum samples with 89% sensitivity and 83% specificity.

3.5. tRF could serve as candidate PND and PDD biomarkers

We next asked whether tRF could also differentiate PND and PDD from control. Specifically, we examined how sensitivity and specificity change when sub-dividing PD into PDD and PND. For the prefrontal cortex samples, we find that classification is slightly less specific in control vs. PND than in control vs. PDD (Supp. Figure 3B).

When using CSF samples, the sensitivity in distinguishing control cases from female PDD patients increases to 98% and to 94% for PND patients. However, specificity is slightly lower (95%) in control vs. female PDD patients. For male patients, the sensitivity and sensitivity trends for PDD and PND are the same; however, in absolute terms, the values of sensitivity and specificity that can be achieved are lower than the values for female patients.

For female and male serum samples, incorporating knowledge of the presence or absence of dementia left the sensitivity and specificity figures essentially unchanged (Supp. Figure 3).

3.6. Prioritizing the tRF that can distinguish PD, PND, and PDD from control samples

We next prioritized tRF based on their ability to discriminate PD from control. For PLS-DA models using all tRF from {PFC ∪ CSF ∪ SER}, we asked how many times each tRF showed a "Variable Importance in Projection" (VIP) score ≥ 1.5 in all N^2 PLS-DA models (Supp. Table 4). A tRF VIP score is a measure of tRF contribution to the model's ability to classify samples: the higher the VIP score of a tRF, the higher its relevance as a biomarker. Those tRF that are assigned consistently high VIP scores during the simulation are ideal components with which to build a biomarker signature.

We ranked tRF by the number of times they exceeded VIP threshold in each of the N^2 models across the five comparisons (Supp. Table 5). Drawing on this ranked list, we ran PLS-DA simulations with 100% of the samples in each category, using the top X tRF by number of VIP votes received, from $X = 3$ to 44. The maximum value of 44 is due to the availability of only 44 VIP tRF in female serum samples. For each value of X, we fixed the maximum number of components in each PLS-DA model at X, and used cross-validation to allow the PLS-DA algorithm to build the best possible model from the available data. Finally, we plotted the error rate of each model as a function of the number of tRF used (Fig. 4C).

As a final step in the selection of the different biomarker panels for PD, we chose the smallest number of tRF contributing to the minimum error rate achieved in Fig. 4C. We plotted the normalized discriminant score for each sample (100% of samples) in this best possible model using the smallest number of tRF (Fig. 4D). Specifically, we find that X = 38 tRF are necessary to model samples from male prefrontal cortex, whereas the numbers become 27, 28, 19 and 18 for female CSF, male CSF, female serum, and male serum samples respectively (Fig. 4D; listed in Supp. Table 5).

4. Discussion

In this exploratory study, we investigate the relevance of tRF in the context of PD. Our findings suggest that tRF have the potential to serve as non-invasive biomarkers for PD. As is the case with biomarkers, additional and independent validation of tRF-based biomarkers for PD

will be necessary. Currently, our data suggest that such a biomarker would serve a diagnostic goal. Many more datasets will be required before it can be determined how well tRF perform as prognostic biomarkers or whether they can serve as therapeutic targets. The apparent sex dependence further adds to the number of datasets that are needed. For the time being, we posit based on the available data that tRF biomarkers for PD could be used to provide independent confirmation after clinical investigation has raised suspicion of disease. Our reasoning is based on the presence of an observable disease signal in a large group of samples characterized by a wide range of disease durations (Supp. Table 1).

Our analysis uncovered a large number of tRF in prefrontal cortex, CSF, and serum from both healthy control patients and PD patients. Several tRF are shared among the three collections. Some of the tRF in these three collections have been identified in other tissue/cellular contexts and are currently catalogued in MINTbase [23,29]. Even so, a large portion of the tRF we identified in the samples we analyzed are specifically observed in only one of prefrontal cortex, CSF, or serum.

The majority of the tRF that are differentially abundant (DA) between control and PD are unique to exactly one of prefrontal cortex, CSF, or serum categories. We identified a first group of DA tRF using a simple *t*-test without p-value correction for multiple testing in order to generate the largest possible pool of biomarker candidates. We then relied on the stringency of a robust downstream method, PLS-DA to filter out DA tRF as needed. Specifically, all DA tRF (by *t*-test) from each of five separate PD vs. control comparisons became the input for the PLS-DA analysis, the latter having been carried out with the appropriate cross-validation. The resulting models were tested using standard data randomization schemes (see Methods). Thus, the tRF that emerged as important for successfully classifying samples using PLS-DA represent statistically sound choices and represent high-priority candidates for future validation experiments.

In addition to identifying DA tRF between control and PD samples, we further tested whether tRF were DA when we partitioned the PD group into those without dementia (PND) and those with dementia (PDD). Notably, we identified a considerable number of tRF that are DA specifically in one of control vs. PND, control vs. PDD, or PND vs. PDD. It is conceivable that these tRF capture changes that are specific to disease progression. The mutation background of these patients is not known (beyond GBA and LRRK2 for prefrontal cortex patients), so these tRF could also change as a result of individual-specific molecular alterations.

Using the PLS-DA approach, we are able to sensitively and specifically classify samples as either control or PD. Specifically, we classify samples with high sensitivity ("sens.") and specificity ("spec.") in male prefrontal cortex samples (98% sens. / 86% spec.), in female CSF samples (90% sens. / 98% spec.), in male CSF samples (91% sens. / 79% spec.), in female serum samples (100% sens. / 81% spec.), and in male serum samples (89% sens. / 83% spec.). We note that several samples from each of the five categories could not be assigned reliably to their true category across the N^2 simulations (Fig. 4A). Such heterogeneity in the case of the biofluid comparisons could be the result of cross-talk from having included tRF that were previously observed in other tissue contexts unrelated to PD, such as those catalogued in MINTbase [23,29]. Alternatively, differences in patient characteristics such as mutation landscape inciting disease or other clinicopathologic factors could be behind this classification difficulty.

Looking at the tRF forming the smallest possible discriminative tRF signatures, we noted the inclusion of a handful of 5′-tRNA halves (5′-tRHs) with known mechanisms of action in the brain. Several research groups have explored the function of some tRHs in the brain, and found that 5′-guanine rich 5′-tRHs from tRNA Ala and tRNA Cys may serve a protective role in human motor neurons, by forming G quadruplex structures [7]. The G-quadruplex structures promote stress granule formation. ALS-associated C9orf72 RNA repeats seem to interfere with this function, thereby promoting motor neuron death. Such properties may explain the fact that we see tRHs from tRNA Ala in the signature for prefrontal cortex and CSF samples: these tRHs may play a role in PD pathophysiology, similar to the observed role in ALS. We note that 5′-tRHs are known to be induced generally by stress conditions in a wide variety of cell types [1,2,31,32].

In addition to tRHs implicated in ALS, we observe 5′-tRHs from tRNA Gly and tRNA Val in our signatures. These 5′-tRHs were previously shown to be up-regulated acutely after ischemic stroke events [8]. Though the downstream purpose of this mechanism remains unclear, these tRHs apparently play a role in response to a traumatic event in the brain. We find these 5′-tRHs in the candidate biomarker signature for serum samples, indicating that they may be active in the vasculature in PD contexts. We note that mouse studies demonstrated that overexpression of these same 5′-tRHs inhibits proliferation of human umbilical vein endothelial cells *in vitro*.

In summary, in this presentation, we explored the role of a novel species of non-coding RNA in the context of PD, in actual brain samples and biofluids from PD patients and controls. We note that this study represents an exploration of the possibility that tRFs may serve as reliable biomarkers in the context of PD. Further validation and confirmatory efforts are both warranted and required. We find that tRF exhibit differential behavior in the disease state. Through additional scrutiny of these molecules, we identified a large collection of DA tRF not previously encountered in other tissues and/or diseases. The DA tRF that we identified in serum represent good candidates for the development of non-invasive biomarkers for PD. We stress, however, that this analysis only serves to demonstrate the potential of these molecules as biomarkers. A number of next steps will be necessary to corroborate or refute the relevance of these molecules as biomarkers for the disease.

In our effort to gauge the stringency of our biomarker selection, we also carried out a comparison with Alzheimer's Disease (AD) deep sequencing data (Supp. Material and Supp. Table 7). We found that the VIP biomarker tRF that are contained in the signatures we discussed above are indeed specific to PD. However, we note that these AD datasets come from a different cortical area and the samples were processed by a different laboratory.

Future work will need to validate the PD-specific nature of the tRF that emerged from our work (Fig. 4) by, e.g., comparing data from two or more independent PD cohorts, run and processed using the same methods in the same laboratory. If successful, the analyses should be extended to encompass larger patient cohorts, to solidify the practical utility of these tRF signatures as non-invasive diagnostic biomarkers of PD. Future work should also seek to regenerate our tRF signatures *de novo* through independent sequencing of independent sample collections from prefrontal cortex, CSF, and serum. Currently, deep sequencing represents the ideal choice for this task. Indeed, in prior work, we showed that popular commercially-available qPCR methods are prone to quantification errors because they lack the ability to guarantee the identity of both 5′ and 3′ endpoints of short molecules such as tRF and microRNAs [33]. Lastly, future experiments should also seek to recapitulate the known functions of some tRF in our signatures in PD-specific contexts, to determine whether these molecules are involved in similarly important mechanisms in PD.

Financial acknowledgements and disclosures

R.M.: Supported by Thomas Jefferson University funds.

E.L.: Supported by Thomas Jefferson University funds.

I.R.: Supported by Thomas Jefferson University funds and the W. M. Keck Foundation.

Financial disclosure/conflict of interest statement

None of the authors have competing interests – financial or non-financial – to disclose.

Acknowledgements

We thank Drs. Diane Merry and Yohei Kirino for their insightful comments on the interpretation of our results. We also thank Tess Cherlin for her insightful comments on the interpretation of results and presentation of findings.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.parkreldis.2019.05.035.

References

[1] H. Fu, J. Feng, Q. Liu, F. Sun, Y. Tie, J. Zhu, R. Xing, Z. Sun, X. Zheng, Stress induces tRNA cleavage by angiogenin in mammalian cells, FEBS Lett. 583 (2) (2009) 437–442.
[2] D.M. Thompson, C. Lu, P.J. Green, R. Parker, tRNA cleavage is a conserved response to oxidative stress in eukaryotes, RNA 14 (10) (2008) 2095–2103.
[3] S. Honda, P. Loher, M. Shigematsu, J.P. Palazzo, R. Suzuki, I. Imoto, I. Rigoutsos, Y. Kirino, Sex hormone-dependent tRNA halves enhance cell proliferation in breast and prostate cancers, Proc. Natl. Acad. Sci. U. S. A. 112 (29) (2015) E3816–E3825.
[4] P. Ivanov, M.M. Emara, J. Villen, S.P. Gygi, P. Anderson, Angiogenin-induced tRNA fragments inhibit translation initiation, Mol. Cell 43 (4) (2011) 613–623.
[5] P. Kumar, C. Kuscu, A. Dutta, Biogenesis and function of transfer RNA-related fragments (tRF), Trends Biochem. Sci. 41 (8) (2016) 679–689.
[6] A.G. Telonis, P. Loher, S. Honda, Y. Jing, J. Palazzo, Y. Kirino, I. Rigoutsos, Dissecting tRNA-derived fragment complexities using personalized transcriptomes reveals novel fragment classes and unexpected dependencies, Oncotarget 6 (28) (2015) 24797–24822.
[7] P. Ivanov, E. O'Day, M.M. Emara, G. Wagner, J. Lieberman, P. Anderson, G-quadruplex structures contribute to the neuroprotective effects of angiogenin-induced tRNA fragments, Proc. Natl. Acad. Sci. U. S. A. 111 (51) (2014) 18201–18206.
[8] Q. Li, B. Hu, G.W. Hu, C.Y. Chen, X. Niu, J. Liu, S.M. Zhou, C.Q. Zhang, Y. Wang, Z.F. Deng, tRNA-derived small non-coding RNAs in response to ischemia inhibit angiogenesis, Sci. Rep. 6 (2016) 20850.
[9] A.E. Schaffer, V.R. Eggens, A.O. Caglayan, M.S. Reuter, E. Scott, N.G. Coufal, J.L. Silhavy, Y. Xue, H. Kayserili, K. Yasuno, R.O. Rosti, M. Abdellateef, C. Caglar, P.R. Kasher, J.L. Cazemier, M.A. Weterman, V. Cantagrel, N. Cai, C. Zweier, U. Altunoglu, N.B. Satkin, F. Aktar, B. Tuysuz, C. Yalcinkaya, H. Caksen, K. Bilguvar, X.D. Fu, C.R. Trotta, S. Gabriel, A. Reis, M. Gunel, F. Baas, J.G. Gleeson, CLP1 founder mutation links tRNA splicing and maturation to cerebellar development and neurodegeneration, Cell 157 (3) (2014) 651–663.
[10] W.J. Bradshaw, S. Rehman, T.T. Pham, N. Thiyagarajan, R.L. Lee, V. Subramanian, K.R. Acharya, Structural insights into human angiogenin variants implicated in Parkinson's disease and Amyotrophic Lateral Sclerosis, Sci. Rep. 7 (2017) 41996.
[11] T.U. Steidinger, S.R. Slone, H. Ding, D.G. Standaert, T.A. Yacoubian, Angiogenin in Parkinson disease models: role of Akt phosphorylation and evaluation of AAV-mediated angiogenin expression in MPTP treated mice, PLoS One 8 (2) (2013) e56092.
[12] T.U. Steidinger, D.G. Standaert, T.A. Yacoubian, A neuroprotective role for angiogenin in models of Parkinson's disease, J. Neurochem. 116 (3) (2011) 334–341.
[13] L.M. Chahine, M.B. Stern, A. Chen-Plotkin, Blood-based biomarkers for Parkinson's disease, Park. Relat. Disord. 20 (Suppl 1) (2014) S99–S103.
[14] L.V. Kalia, A.E. Lang, Parkinson's disease, Lancet 386 (9996) (2015) 896–912.
[15] J.A. Santiago, J.A. Potashkin, Evaluation of RNA blood biomarkers in individuals at risk of Parkinson's disease, J. Parkinson's Dis. 7 (4) (2017) 653–660.
[16] M.M. Mielke, W. Maetzler, A 'bird's eye' view on the current status and potential benefits of blood biomarkers for Parkinson's disease, Biomark. Med. 8 (2) (2014) 225–227.
[17] C. Klein, M.G. Schlossmacher, The genetics of Parkinson disease: implications for neurological care, Nat. Clin. Pract. Neurol. 2 (3) (2006) 136–146.
[18] B. Thomas, M.F. Beal, Parkinson's disease, Hum Mol Genet 16 Spec No. 2 (2007) R183–R194.
[19] J.A. Obeso, M.C. Rodriguez-Oroz, C.G. Goetz, C. Marin, J.H. Kordower, M. Rodriguez, E.C. Hirsch, M. Farrer, A.H. Schapira, G. Halliday, Missing pieces in the Parkinson's disease puzzle, Nat. Med. 16 (6) (2010) 653–661.
[20] A. Gupta, V.L. Dawson, T.M. Dawson, What causes cell death in Parkinson's disease? Ann. Neurol. 64 (Suppl 2) (2008) S3–S15.
[21] P. Loher, A.G. Telonis, I. Rigoutsos, MINTmap: fast and exhaustive profiling of nuclear and mitochondrial tRNA fragments from short RNA-seq data, Sci. Rep. 7 (2017) 41184.
[22] R. Magee, P. Loher, E. Londin, I. Rigoutsos, Threshold-seq: a tool for determining the threshold in short RNA-seq datasets, Bioinformatics 33 (13) (2017) 2034–2036.
[23] V. Pliatsika, P. Loher, A.G. Telonis, I. Rigoutsos, MINTbase: a framework for the interactive exploration of mitochondrial and nuclear tRNA fragments, Bioinformatics 32 (16) (2016) 2481–2489.
[24] A.G. Hoss, A. Labadorf, T.G. Beach, J.C. Latourelle, R.H. Myers, microRNA profiles in Parkinson's disease prefrontal cortex, Front. Aging Neurosci. 8 (2016) 36.
[25] A.G. Hoss, A. Labadorf, J.C. Latourelle, V.K. Kartha, T.C. Hadzi, J.F. Gusella, M.E. MacDonald, J.F. Chen, S. Akbarian, Z. Weng, J.P. Vonsattel, R.H. Myers, miR-10b-5p expression in Huntington's disease brain relates to age of onset and the extent of striatal involvement, BMC Med. Genomics 8 (2015) 10.
[26] K. Burgos, I. Malenica, R. Metpally, A. Courtright, B. Rakela, T. Beach, H. Shill, C. Adler, M. Sabbagh, S. Villa, W. Tembe, D. Craig, K. Van Keuren-Jensen, Profiles of extracellular miRNA in cerebrospinal fluid and serum from patients with Alzheimer's and Parkinson's diseases correlate with disease status and features of pathology, PLoS One 9 (5) (2014) e94839.
[27] M. Martin, Cutadapt Removes Adapter Sequences from High-Throughput Sequencing Reads vol. 17, (2011) 2011, (1).
[28] P. Loher, A.G. Telonis, I. Rigoutsos, Accurate profiling and quantification of tRNA fragments from RNA-seq data: a vade mecum for MINTmap, Methods Mol. Biol. 1680 (2018) 237–255.
[29] V. Pliatsika, P. Loher, R. Magee, A.G. Telonis, E. Londin, M. Shigematsu, Y. Kirino, I. Rigoutsos, MINTbase v2.0: a comprehensive database for tRNA-derived fragments that includes nuclear and mitochondrial fragments from all the Cancer Genome Atlas projects, Nucleic Acids Res. 46 (D1) (2018 Jan 4) D152–D159.
[30] C.K. Stein, P. Qu, J. Epstein, A. Buros, A. Rosenthal, J. Crowley, G. Morgan, B. Barlogie, Removing batch effects from purified plasma cell gene expression microarrays with modified ComBat, BMC Bioinf. 16 (2015) 63.
[31] M. Saikia, R. Jobava, M. Parisien, A. Putnam, D. Krokowski, X.H. Gao, B.J. Guan, Y. Yuan, E. Jankowsky, Z. Feng, G.F. Hu, M. Pusztai-Carey, M. Gorla, N.B. Sepuri, T. Pan, M. Hatzoglou, Angiogenin-cleaved tRNA halves interact with cytochrome c, protecting cells from apoptosis during osmotic stress, Mol. Cell. Biol. 34 (13) (2014) 2450–2463.
[32] S. Yamasaki, P. Ivanov, G.F. Hu, P. Anderson, Angiogenin cleaves tRNA and promotes stress-induced translational repression, J. Cell Biol. 185 (1) (2009) 35–42.
[33] R. Magee, A.G. Telonis, T. Cherlin, I. Rigoutsos, E. Londin, Assessment of isomiR discrimination using commercial qPCR methods, Noncoding RNA 3 (2) (2017).