Research Article

Using Extracellular miRNA Signatures to Identify Patients with LRRK2-Related Parkinson's Disease

Luca Jannik Braunger[1]*, Felix Knab[1]* and Thomas Gasser[1,2]

[1]Hertie Institute for Clinical Brain Research, Department of Neurodegeneration, University of Tuebingen, Tuebingen, Germany

[2]German Center for Neurodegenerative Diseases (DZNE), Tuebingen, Germany

*Authors contributed equally

Corresponding author: Felix Knab, felix.knab@med.uni-tuebingen.de

Conflict of interest: The authors declare that they have no conflict of interest.

Abstract

Background: Mutations and polymorphisms in the Leucine Rich Repeat Kinase 2 gene were found to be highly relevant in both sporadic and familial cases of Parkinson's disease. Specific therapies are entering clinical trials and patient stratification is crucial but remains challenging. Dysregulated microRNA expression levels have been observed in cases of sporadic Parkinson's disease and were proposed as biomarker candidates. In this proof-of concept study we evaluated the potential of using the extracellular miRNA signature of patients to identify LRRK2-driven molecular patterns in Parkinson's disease.

Methods: We quantified the expression levels of 91 miRNAs via RT-qPCR in ten individuals with sporadic Parkinson's disease, ten patients carrying a *LRRK2* mutation and eleven healthy controls using both plasma and cerebrospinal fluid. We first compared the miRNA signatures of all individuals using heatmaps and t-tests. Next, we applied group sorting algorithms (Least Absoulte Shrinkage and Selection Operator, Principal Component Analysis, Random Forest) and tested the sensitivity and specificity of their group predictions. Finally, we extracted a collection of discriminatory miRNAs from each analysis.

Results: miR-29c-3p was differentially expressed between *LRRK2* mutation carriers and sporadic cases, with miR-425-5p trending towards significance. Individuals clustered in principal component analysis along mutation status. Group affiliation was predicted with high accuracy in the least absolute shrinkage and selection operator regression model (sensitivity of 89%, specificity of 70%) and the random forest model (sensitivity of 89%, specificity of 70%). miRs-128-3p, 29c-3p, 223-3p and 424-5p were identified as promising discriminators among all analyses.

Conclusions: Identifying LRRK2-driven molecular patterns in Parkinson's disease by characterizing the extracellular miRNA signatures and using group sorting algorithms potentially provides an easily scalable method for molecular subtyping of patients. Knowing that LRRK2 also plays a role in patients with sporadic forms of the disease, this approach could enable identification of patients that are likely to benefit from drugs targeting LRRK2. In light of small sample size, this approach needs to be validated in larger cohorts.

Keywords: [Parkinson's disease, LRRK2, microRNA, miRNA, Biomarker, prediction, group sorting]

Background

Parkinson's disease (PD) is the second most prevalent neurodegenerative disease, affecting approximately 1% of people aged 60 years and older [1]. One central aspect of the pathology is the degeneration of neurons in the substantia nigra, which leads to a dopamine deficiency. Prior to the identification of a point mutation in the SNCA gene, which results in a familial form of the disease, PD was thought to be purely sporadic [2]. Today, we know that a genetic cause can be found in 5% to 10% of PD cases [3], [4] and that a large number of genetic loci modulate disease risk in sporadic patients. The majority of familial PD (fPD) cases is caused by mutations in the Leucine Rich Repeat Kinase 2 (*LRRK2*) gene [5], [6], with the G2019S mutation being the most common [7]. Despite its apparent role in PD, LRRK2 expression levels in the CNS are relatively low compared to peripheral tissues such as the lung and the kidneys, as well as blood monocytes [8], [9]. Further, its function is not yet fully understood, but it was shown to be involved in membrane trafficking [10], [11], cytoskeletal dynamics [12], synaptic transmission, and endolysosomal function [13]. Increased kinase activity is thought to be the cause of the disturbed cellular homeostasis observable in cell models of PD [14]. Most importantly, LRRK2 was shown to play a role in sporadic cases, with non-coding variants in the *LRRK2* gene increasing the risk of PD [15]. Further, sporadic PD (sPD) and fPD cases show overlapping clinical features [16]. Studying fPD cases therefore offers the opportunity to better understand the molecular pathogenesis of both fPD and sPD cases and additionally allows for the collection of biomaterial and clinical data from family members at risk of developing of PD.

There are currently no drugs available that effectively modify the disease course, and therapies are purely symptomatic. Targeting multiple molecular pathways in selected patients with an individually formulated combination of drugs could be a potential

strategy for disease modification [2]. However, this strategy necessitates not only the discovery of particular druggable pathways, but also the careful selection of patients who are most likely to benefit from a given treatment, ideally at an early or even asymptomatic stage in the course of their disease. While aberrant LRRK2 activity represents such a molecular target and LRRK2 inhibitors are approaching clinical testing [17], [18], both pre-selecting patients and monitoring target engagement and drug efficacy remain challenging. Currently, the diagnosis of PD is primarily based on clinical criteria [19] and there are no established biomarkers available that can aid in the early detection of disease onset before significant neurodegeneration occurs. Recently, microRNAs (miRNAs), small non-coding RNA molecules that are involved in the posttranscriptional regulation of gene expression [20], have been intensively studied as potential biomarkers for neurodegenerative diseases, including PD [21], [22]. Extracellular miRNAs are easily accessible in biofluids such as plasma or cerebrospinal fluid (CSF) and are potentially involved in the pathogenesis of PD [23], [24]. This makes them promising biomarker candidates since they might be used to track pathogenic PD processes and disease progression on a molecular level both before and after treatment.

So far, most PD biomarker studies have focused on the differentiation of sPD patients from healthy controls (HC), while in the present proof-of-concept study we aimed to test how well extracellular miRNAs can be used to reliably differentiate between sPD patients and *LRRK2* mutation carriers ($LRRK2_{MC}$). Since the distinction between fPD and sPD is possible by using more stringent methods, such as sequencing, in the long run we aim to identify individuals with relevant involvement of LRRK2-dependent pathways among the sPD population. However, in a first step we decided to define what constitutes a molecular LRRK2 fingerprint by comparing the extracellular miRNA

signatures of fPD and sPD cases. Given that LRRK2 was described to change the cellular miRNAome [25], we hypothesized that mutations in the *LRRK2* gene also alter the extracellular miRNA signature in $LRRK2_{MC}$. Knowing that LRRK2 also plays a significant role in sPD, this approach could be used in future studies to identify patients with a LRRK2 specific miRNA signature among the general PD population and therefore select those, that are likely responsive to a LRRK2-specific therapy. Additionally, with miRNAs involved in the PD pathogenesis [23], [24], regular assessment of the miRNA signature of patients could support monitoring of treatment efficacy in therapies targeting LRRK2, where a shift of miRNA expression levels could be reflective of, e.g., target engagement. Lastly, identifying miRNAs that discriminate sPD from fPD could help to deepen our understanding of the role of LRRK2 in the underlying mechanisms of PD.

Materials and Methods

Experimental Design

In a first step, we quantified the expression levels of 91 extracellular miRNAs from a cohort of ten sPD, ten fPD and eleven HC in both plasma and CSF using RT-qPCR (Figure 1). After processing of raw Ct values and calculation of log2 fold change values (log2fc), as described in more detail in the respective method section, we performed t-tests to identify single miRNAs with the potential of discriminating between groups. We further applied group sorting algorithms such as principal component analysis (PCA), least absolute shrinkage and selection operator (LASSO) and random forest (RF) using the calibrated Ct values of all miRNAs that could be reliably detected throughout the cohort. The ability to differentiate between sPD patients and $LRRK2_{MC}$ with and without PD was assessed for each sorting algorithm by evaluating their sensitivity and specificity. Finally, we selected those miRNAs that had a significant impact on the performance of each model.

Cohort Design

For each $LRRK2_{MC}$, one sPD and one HC was added to the study cohort. To obtain homogenous groups, individuals were selected based on age, gender, and disease duration. As the histopathological complexity of PD likely increases over time and patients with long disease durations might no longer display pathologies specific to mutation status rather than showing features common to all PD patients, disease durations were kept as short as possible. The group of $LRRK2_{MC}$ initially included a PD patient carrying two *LRRK2* variants (N1437S, S1647). Given that these polymorphisms were not reported to be pathogenic, we finally decided to exclude this patient from the analysis. The final cohort therefore comprised three groups (Table 1):

Ten $LRRK2_{MC}$ (fPD patients: n=6 (*LRRK2* G2019S: n = 3; *LRRK2* G2019S/G1819: n=1, *LRRK2* R1441C: n=1; *LRRK2* I2020T: n=1), asymptomatic G2019S mutation carriers: n=4), ten sPD patients and eleven HCs. After performing PCA as described in more detail in the respective method section, the data set of one of the asymptomatic $LRRK2_{MC}$ appeared as a technical or biological outlier and was removed from the plasma dataset before further analysis (Fig. S1). Clinical features are reported in Table 1. The researcher performing the experiments was blinded to the group annotation.

Table 1: Overview of clinical data.

	***LRRK2* mutated**		**Sporadic PD (n = 10)**	**Healthy Control (n = 11)**
Disease status	**PD (n = 6)**	**Asymptomatic (n = 4)**		
Age (years)	66.8 ± 8.7	52.3 ± 19.6	64.5 ± 10.9	60.1 ± 14.5
Gender (m:f)	1:5	0:4	2:8	2:9
DD (years)	8.8 ± 3.6	-	7 ± 4.1	-
MoCA	25 ± 4.1 (5/6)	28.5 ± 1.5	25.3 ± 3.7 (9/10)	28.3 ± 1.5 (4/11)
UPDRS-III	21.7 ± 8.2	1.0 ± 0.8 (3/4)	17.4 ± 10.1 (8/10)	0.7 ± 0.9 (3/11)
LEDD (mg)	612.8 ± 295.6	0.0 ± 0.0	357.5 ± 380.0	0.0 ± 0.0

DD: Disease duration, MoCA: Montreal-Cognitive-Assessment, UPDRS-III: Unified Parkinson's Disease Rating Scale, LEDD: L-Dopa equivalent daily dose. Age, Disease duration, MoCA, UPDRS-III and LEDD stated as mean value with standard deviation. MoCA and UPDRS-III were not accessible for some patients and numbers in brackets give fraction of accessible data.

miRNA selection and qPCR panel design

We scanned the literature for miRNAs that were previously reported to be dysregulated in the context of *LRRK2* mutations, functionally relevant to LRRK2 or generally dysregulated in the context of PD. Next, we selected a total of 91 miRNAs to be

included in our customized qPCR panel (Table S1). Selection was based on a set of criteria, including the number of relevant studies reporting dysregulation and a preference for miRNAs assessed in CSF and plasma samples over miRNAs reported in brain tissue or animal models. Additionally, miRNAs associated with LRRK2-linked PD were given priority over those associated with sporadic PD or other neurodegenerative diseases. For quality control purposes, we included a set of spike-in controls. Those included UniSpike 2 and 4 from the QIAGEN RNA Spike-In Kit (Cat. No.: 339390, QIAGEN, Hilden, Germany), which were used to monitor RNA isolation efficacy, UniSpike 6 (included in the QIAGEN miRCURY LNA RT Kit) to verify the efficacy of reverse transcription (RT) and UniSpike 3 (pre-applied primers on the custom qPCR plates) for inter-plate calibration. Additionally, a blank spot containing no primer was included and functioned as a negative control.

Collection and Storage of Biofluids from Study Participants

Blood and CSF samples were collected from individuals according to standardized protocols. Within 90 minutes after collection from the individual, the collected material was brought from the medical facilities to the Neuro-Biobank Tuebingen and immediately processed. CSF and EDTA plasma were centrifuged at 2000g for 10 minutes. Samples were then transferred to a 15 ml Falcon tube, mixed thoroughly by vortexing, and aliquoted to cryotubes. Finally, both CSF and plasma cryotubes were frozen and stored at -80°C in the biobank. Before RNA isolation, samples were slowly thawed on ice.

RNA isolation

Extracellular RNA was isolated from plasma and CSF using the QIAGEN miRNeasy Serum/Plasma Advanced Kit (Cat. No: 217204, QIAGEN) and following the vendor's

instructions with slight modifications. Briefly, 200 µl of plasma or 400 µl of CSF were used as a starting volume and subsequent reaction volumes were adjusted accordingly. After addition of RPL buffer and incubation for three minutes, we added 1 µl of the spike-in mix provided in the QIAGEN RNA Spike-In Kit (Cat. No.: 339390, QIAGEN) to the tube, in order to monitor the efficacy of the RNA isolation. Finally, RNA elution volume was reduced to 7 µl of RNase-free water.

Reverse Transcription

The RT reaction was performed using the QIAGEN miRCURY LNA RT Kit (Cat. No.: 339340, QIAGEN). The reaction mix included 2 µl of 5x miRCURY SYBR® Green RT Reaction Buffer, 1 µl of 10x miRCURY RT Enzyme Mix, 4.5 µl of RNase-free water, 0.5 µl of UniSpike 6 RNA spike-in and 2 µl of template RNA per sample. After brief centrifugation, tubes were incubated at 42 °C for 60 minutes, followed by 5 minutes of incubation at 95 °C, which inactivated the RT reaction. Finally, samples were diluted 1:40 using RNase-free water.

qPCR

qPCR was performed using QIAGEN miRCURY LNA miRNA Custom PCR Panels (Catalog No.: 339330, QIAGEN) along with the QIAGEN miRCURY LNA SYBR Green PCR Kit (Cat. No.: 339347, QIAGEN). The reaction mix included 515 µl of 2x miRCURY SYBR® Green Master Mix, 115 µl of RNase-free water and 400 µl of the cDNA template per sample. After pipetting 10 µl into each reaction-well, which already contained the dried primers, the plates were sealed and vortexed for one minute. After brief centrifugation, plates were incubated for ten minutes at room temperature to allow primers to resolve. Finally, plates were vortexed for one minute and briefly centrifuged.

Amplification and quantification were performed using a LightCycler® 480 instrument (Roche, Basel, Switzerland). The cycling program included a two-minute heat activation step at 95 °C, followed by 45 cycles of denaturation at 95 °C for ten seconds and annealing at 56 °C for 60 seconds.

Processing of raw data

Before applying statistical analysis, raw data were processed and Ct and fold-change values (fc) were calculated (Fig. S2): First, raw data from the qPCR were converted using the LC480Conversion software to make them readable for LinRegPCR (Version 11.0, [26]), which was used for calculating Ct values. Further processing and analysis of data was performed in R (Version 4.2.2) using RStudio (Version 2023.03.0+386) [27]. To account for variations between plates, the Ct values were normalized using an interplate calibration factor. The calibration factor for each plate was calculated by subtracting the mean Ct value of UniSpike 3 reactions from all plates from the mean Ct value of all UniSpike 3 reactions in a given plate. All Ct values of the respective plate were then calibrated by subtracting the respective calibration factor from all Ct values in that plate. After calibration, Ct values > 40 were considered unspecific and excluded from the dataset. If a miRNA was not specifically amplified in more than one individual, it was removed completely from the analysis, after confirming that missing values occurred equally over all groups (Fig. S3). This resulted in 58 remaining miRNAs for the plasma dataset and eleven miRNAs for the CSF dataset.

Next, the fc and log2fc values were calculated using the $2^{-\Delta\Delta Ct}$ method (Fig. S4) [28]. In a first step, for each individual we independently calculated the mean Ct by adding all Ct values of every quantified miRNA and dividing it by the number of quantified miRNAs, which resulted in a $mean_{Ct}$ for each individual. We then computed ΔCt for

each miRNA and individual by subtracting the $mean_{Ct}$ of the respective individual from every miRNA quantified in that individual. This resulted in multiple $\Delta Ct_{miR\text{-}n}$ per individual, where *n* represents the specific miRNA. The entire group of healthy controls was used as a reference group. Subsequently, for each miRNA we calculated the mean of $\Delta Ct_{miR\text{-}n}$, only using $\Delta Ct_{miR\text{-}n}$ from the HCs ($mean\ \Delta Ct_{miR\text{-}n\text{-}healthy}$). Next, ΔΔCt for each miRNA and individual was determined by subtracting the $mean\ \Delta Ct_{miR\text{-}n\text{-}healthy}$ from the patient's $\Delta Ct_{miR\text{-}n}$. Finally, the fold change was obtained using the following formula: fold change = $2^{-\Delta\Delta Ct}$. To achieve a linear scale with symmetry around zero, log2fc was used for the heatmaps and t-tests. Data from the HCs were used only as a reference for calculating the fold change values. Since the focus of this study was put on the identification of a LRRK2-driven molecular patterns and the discrimination between fPD and sPD cases, in the subsequent analyses, data from the HCs were not analyzed further.

Data Visualization

Heatmaps visualizing either the log2fc or calibrated Ct values were created using the *pheatmap* package in R [29]. Log2fc and Ct values were scaled row-wise using the *scale* function while ignoring missing values. After confirming normality of the data via assessing QQ-Plots, an unpaired t-test using the log2fc values was performed for each miRNA, comparing the sPD group with the $LRRK2_{MC}$. The p-value threshold to determine significance was corrected for multiple testing by dividing by the number of tests. The distribution of p-values was examined and visualized via histograms. PCA was performed on calibrated Ct values using the built-in *prcomp* function of R. As the function does not accept missing values, they were imputed by setting them to the mean

Ct value from all individuals. A two-dimensional plot was generated by graphing PC1 and PC2 values.

Least Absolute Shrinkage and Selection Operator regression models

The *caret* package [30] was used for building the LASSO regression models. For missing values, the mean of the Ct values of all patients was used. We selected sensitivity and specificity in differentiating between the sPD and $LRRK2_{MC}$ group, determined by Leave-One-Out-Cross-Validation (LOOCV) [31], as our experimental read-outs.

Random forest Prediction Model

For building a RF model that predicts group affiliation to either the sPD or the $LRRK2_{MC}$ group, the *randomForest* package [32] was used. The model consisted of 200 single decision trees. Missing values were computed using the *rfImpute* function. Sensitivity and specificity of the classification was tested on out-of-bag samples, along with proximity heatmaps describing similarity of patient pairs based on the number of trees that classify those two patients in the same terminal node. The Gini coefficient's mean decrease, which describes each decision node in terms of classification accuracy, was used to evaluate the importance of each variable. High purity of a node results in a low Gini coefficient. Consequently, variables that are more crucial for distinguishing between groups exhibit a greater mean decrease of the Gini coefficient compared to others.

Integration of CSF and Plasma data sets

To integrate the information from both the plasma and CSF datasets, combined variables were computed. First, Pearson's correlation analysis was performed between the Ct values of the eleven miRNAs reliably detectable in CSF (Ct_{CSF}) and the Ct values of the corresponding miRNAs detected in plasma (Ct_{plasma}). A two-tailed p-value cut-off was applied, and alpha was set to 0.05. All significant combinations were used to form new variables by multiplying the corresponding Ct values ($Ct_{multiplied}$). This generated 20 new variables, which were subsequently used to create a heatmap, perform t-tests and PCA and generate LASSO and RF models.

Identification of discriminatory miRNAs

In a final step, we wanted to select those miRNAs that most accurately discriminated between groups. This feature selection was performed for each of the analyses using the following criteria: 1) p-values resulting from the unpaired t-test were sorted in ascending order and the top five miRNAs with the smallest p-value were selected; 2) the loading scores of PC1 were assessed and the top five miRNAs were selected; 3) the five most influential miRNAs from the RF model were selected based on the mean decrease of the Gini coefficient; 4) the LASSO regression model automatically selects discriminatory features through shrinkage and elimination of less relevant variables by introducing penalties; in this case, three miRNAs were selected. As the CSF data set alone had proven to not efficiently discriminate between groups, these selection steps were only applied to the Ct_{plasma} and $Ct_{multiplied}$ data sets.

Results

Detection of miRNA Expression Patterns in Plasma and CSF

In plasma, a total of 58 miRNAs passed our selection criteria. 23 miRNAs showed stable expression in 29/30 individuals while the remaining 35 miRNAs were detected in all 30 individuals. When visualizing log2fc values using a heatmap, it appeared that the group of $LRRK2_{MC}$ showed similar expression patterns and could be distinguished from the sPD group (Fig. 2A). When using the Ct_{plasma} data, the expression patterns became less apparent but still noticeable (Fig. S5A). In contrast, in CSF only eleven miRNAs survived our selection process, with five miRNAs detected in all 31 individuals and six miRNAs in 30/31 individuals. Expression levels in CSF were generally lower compared to plasma, indicating that the decreased number of detected miRNAs in CSF is due to the sensitivity of the qPCR. In CSF, no clear clustering was observable (Fig. 2B). Heatmaps based on Ct_{CSF} also did not show any clear observable group-specific clusters (Fig. S5B).

Identification of single discriminatory miRNAs through t-test

After performing unpaired t-test on log2fc values in plasma, a total of 17 miRNAs were below a p-value threshold of 0.05 (Table S2). After correction for multiple comparisons by adjusting the p-value threshold accordingly to 0.0009, miR-29c-3p showed a significant difference ($LRRK2_{MC}$: -0.39 (SD: ±0.74); sPD: 0.62 (SD: ±0.54), $t(56) = 4.21$, $p = 0.0007$). In CSF, only miR-223-3p showed a p-value below the threshold before, but not after correction for multiple testing when setting the p-value threshold to 0.0045 ($LRRK2_{MC}$: -0.49 (SD: ±1.05); sPD: 0.62 (SD: ±1.20), $t(9) = -2.2$, $p = 0.041$). P-values in the plasma data set were not equally distributed but concentrated at the lower end of the scale (Figure 2C), while the p-values in the CSF data set appeared to

be randomly distributed (Figure 2D). This indicates that the miRNA expression levels measured in CSF alone seem to contain little information on group discrimination, while in plasma a subset of miRNAs can be used to identify $LRRK2_{MC}$. As a consequence, in the following analyses we primarily focused on the plasma dataset. We then sorted p-values from the plasma data set in ascending order and identified the top 5 miRNAs with the lowest p-value (Figure 2E). These included miR-128-3p (log2fc in $LRRK2_{MC}$: -1.01 (SD: ±0.66); sPD: 0.03 (SD: ±0.83), $t(56) = 3.06$, $p = 0.0071$), miR-148a-3p ($LRRK2_{MC}$: -0.60 (SD: ±0.56); sPD: 0.38 (SD: ±0.73), $t(56) = 3.31$, $p = 0.0042$), miR-153-3p ($LRRK2_{MC}$: -0.97 (SD: ±1.19); sPD: 0.65 (SD: ±0.69), $t(56) = 3.58$, $p = 0.0035$), miR-29c-3p ($LRRK2_{MC}$: -0.39 (SD: ±0.74); sPD: 0.62 (SD: ±0.54), $t(56) = 4.21$, $p = 0.0007$) and miR-425-5p ($LRRK2_{MC}$: -0.33 (SD: ±0.38); sPD: 0.54 (SD: ±0.55), $t(56) = 4.01$, $p = 0.0010$).

Detecting patient clusters using plasma miRNA expression levels and PCA

PCA explained 82% of the overall variance in the plasma dataset. PC1 (77.4% of variance) seemed to explain the majority of the variance and appeared to discriminate groups, while PC2 only explained 4.6% of the overall variance (Fig. 3A). Interestingly, all $LRRK2_{MC}$ clustered together and the various *LRRK2* mutations could not be distinguished (Fig. 3A). The PCA of the CSF data explained 84.2% of the overall variance, with both PC1 (75.4% of variance) and PC2 (8.8% of variance) not able to differentiate between the groups (Fig. 3B).

miRNA expression levels in CSF correlate to Plasma levels

A total of 20 miRNA combinations from Ct_{plasma} and Ct_{CSF} values showed significant correlation (Fig. 3C) with R-values ranging from 0.36 to 0.51 (Table S3). Significantly correlating combinations were used to obtain new variables by multiplication.

Multiplied Ct values discriminate groups in t-test and PCA

When creating a heatmap based on the $Ct_{multiplied}$ data, no clear group-specific clusters were noticeable (Fig. S6A). In the t-test ten miRNA-combinations reached a p-value below 0.05 (Table S4). Only the combination of miR-29c-3p from plasma and miR-27a-3p from CSF passed the adjusted p-value threshold of 0.0025 ($LRRK2_{MC}$: 930.8 (SD: ±61.6); sPD: 813.6 (SD: ±55.9), $t(18) = 4.2$, $p = 0.0007$). PCA for $Ct_{multiplied}$ discriminated the groups while explaining 90.2% of the total variance (PC1: 83.2%, PC2: 7%) (Fig. S6B).

LASSO Regression and Random Forest Models predict group membership

LASSO regression and the RF model were performed using the Ct_{plasma} and the $Ct_{multiplied}$ dataset (Fig. 4A). Sensitivity of the LASSO model was 88.8% (8/9) for both Ct_{plasma} and $Ct_{multiplied}$. Specificity was 70.0% (7/10) (Ct_{plasma}) and 80.0% (8/10) ($Ct_{multiplied}$), respectively (Fig. 4B). Predictions from the RF model using Ct_{plasma} had a sensitivity of 88.8 % (8/9), while the model using $Ct_{multiplied}$ showed a sensitivity of 77.7% (7/9). Specificity was 70.0% (7/10) when using Ct_{plasma} and 80.0% (8/10) when using $Ct_{multiplied}$ (Fig. 4B). Classifications performed by the RF model were further analyzed by looking into proximity scores of patient pairs. The resulting heatmap for the Ct_{plasma} dataset indicated that patients within the same group exhibit greater proximity to each other than patients from different groups (Fig. 4C). This becomes

even more apparent in the model using $Ct_{multiplied}$ (Fig. 4D). Next, mean decrease of Gini scores were calculated for both of these RF models (Figure 4E and 4F) to assess variable influence on classification accuracy. Interestingly, in both RF models, miR-223-3p had the highest score, indicating a great impact on the model performance.

Selection of discriminatory miRNAs identified by multiple tests or models

Finally, for each test or model, we identified those miRNAs, that showed the greatest potential for group separation or prediction (Table 2). From the RF models, PCA and the t-tests, we extracted five miRNAs each, while the LASSO model identified three miRNAs. Interestingly, of those miRNAs, three were identified by more than one method; 1) miR-29c-3p was identified in t-tests and LASSO, 2) miR-128-3p in t-test and RF, and 3) miR-424-5p in LASSO and RF (Fig. 4G).

Table 2: miRNAs selected as most influential by the different models.

t-test selected	PCA selected	LASSO selected	Random forest selected
miR-29c-3p	miR-19b-3p	**miR-29c-3p**	miR-26a-5p
miR-128-3p	miR-27a-3p	**miR-424-5p**	miR-106a-5p
miR-148a-3p	miR-28-5p	miR-532-5p	**miR-128-3p**
miR-153-3p	miR-30b-5p		*miR-223-3p*
miR-425-5p	miR-185-5p		**miR-424-5p**

The table displays miRNAs that were identified as relevant from each model or test. miRNAs highlighted in bold were identified in more than one analysis. Of note, miR-223-3p (italic) was extracted from both the Ct_{plasma} and the $Ct_{multiplied}$ RF models.

Discussion

In this proof-of-concept study, we investigated the potential of extracellular miRNA signatures in both plasma and CSF to identify LRRK2-driven molecular patterns in Parkinson's disease by bioinformatic analyses. We found that the expression levels of a subset of miRNAs measured in plasma as well as computed variables created through multiplication of correlated miRNA expression levels in plasma and CSF can distinguish between sPD and $LRRK2_{MC}$. While miRNAs have been intensively studied in the context of PD before, to the best of our knowledge, this is the first study that applies advanced machine learning based methods to a broad data set of miRNA expression levels in order to discriminate between genetic and sporadic forms of PD and identify molecular patterns driven by *LRRK2* mutation status.

PCA separated $LRRK2_{MC}$ from most individuals with sPD without differentiating between the various *LRRK2* mutations. This indicates that LRRK2 gene status has an effect on the extracellular miRNA signature in PD patients. Some of the individuals with sPD were positioned more closely to the cluster of $LRRK2_{MC}$ than others. We hypothesize that in these individuals LRRK2 might be more of a driver of disease pathophysiology than in the other sPD patients included in this study. This hypothesis needs to be tested in larger cohorts but our results underline the potential of the approach we have taken in this proof-of-concept study.

We extracted discriminatory miRNAs from each analysis and found that some miRNAs were overlapping between models. Overall, miR-29c-3p, miR-128-3p, miR-424-5p and miR-223-3p were found to be the most promising miRNAs for the differentiation between $LRRK2_{MC}$ and sPD patients.

miR-29c-3p, which was identified in t-test and LASSO, has been reported to play a role in disease modulation by mediating neuroinflammation through transcription factor

SP1 and has been found to be involved in apoptotic processes [33]. It has further been proposed as a PD specific marker in several studies, but contradicting findings on expression levels compared to controls exist [34], [35] [36]. In our cohort, sPD patients showed upregulated miR-29c-3p levels, while it was downregulated in *LRRK2*$_{MC}$ compared to HCs. Additionally, miR-223-3p, which we identified in the RF models of both Ct_{plasma} and $Ct_{multiplied}$, has been reported to be upregulated in midbrain dopamine neurons [37] and serum of PD patients [38]. It has been shown to be involved in the modulation of the NLRP3 inflammasome activity [39], again indicating towards shifts in neuroinflammatory processes. Interestingly, LRRK2 has long been suspected to play a role in inflammation and e.g., was shown to affect microglial activation and pro-inflammatory cytokine production [40]. This involvement in inflammatory processes could explain why in our study cohort the levels of miR-29c-3p and miR-223-3p differ in individuals with or without a LRRK2 mutation.

miR-128-3p, identified in t-test and RF, has previously been identified as a potential treatment target due to its ability to protect neurons from apoptosis. Specifically, it plays a crucial role in the Wnt/β-catenin pathway by targeting and mediating AXIN1, along with its interaction with HIF-1α [41], [42]. The upregulation in the sPD group could therefore be reflective of a compensatory mechanism, that might not be relevant in PD patients carrying a *LRRK2* mutation, given the reduced expression levels in $LRRK2_{MC}$.

miR-424-5p, identified in LASSO and RF, has been shown to be increased in the white matter of Alzheimer's disease patients [43] and the forebrain of PD patients, while being associated with FOXO1 [44]. Excitingly, FOXO1 activity is induced by LRRK2 [45] and consequentially, altered miR-424-5p levels could be a result of differential LRRK2 activity.

Due to the low concentration of RNA in CSF compared to plasma, reliable detection of miRNAs was challenging, resulting in the exclusion of many miRNAs from the analyses performed in CSF. This artificially reduced the complexity of the miRNA signature we were able to assess in CSF and might explain the relatively poor performance of CSF-based analyses. An alternative explanation could be that LRRK2 protein expression levels are known to be low in the CNS [8], [9] despite their apparent relevance in the pathophysiology of PD. In contrast, peripheral organs such as the lung and the kidney display higher expression levels, which could also explain the increased accuracy of models based on Ct_{plasma} in identifying $LRRK2_{MC}$. Together with the less invasive nature of a blood withdrawal compared to a lumbar puncture, plasma-based read-outs might therefore be preferable for the approach we have taken in the present study.

We have further analyzed the relation of miRNA expression levels in CSF and plasma and found that for most miRNAs, these two biofluids seem to display very different signatures. However, through correlation analysis we have identified a subset of miRNAs whose expression levels in the CNS and the periphery could be related. In addition, using $Ct_{multiplied}$ increased the performance of our prediction models. This on the one hand highlighted the potential of miR-223-3p, which seemed to be highly indicative of mutation status both in plasma and CSF. On the other hand, it also underlines the potential of biomarker studies that do not seek to identify singular read-outs but instead try to build multi-layered and composited read-outs identified using machine learning algorithms.

Although this proof-of-concept study revealed promising findings and deepens our understanding of LRRK2 associated PD, it clearly has limitations that we would like to address. The number of individuals included in the present study is relatively small.

When using classification by group sorting algorithms, larger sample sizes will always increase the robustness and replicability of the findings. Further, due to the relatively small sample size, our study is prone to technical outliers as shown in the PCA analysis. We believe that we have presented in detail the potential that lies in trying to identify LRRK2 signatures using miRNA expression levels in plasma; however, the logical next step clearly has to be to increase the cohort sizes. Further, we selected miRNAs to be included in our qPCR panel based on literature and therefore only considered miRNAs that have already been reported as dysregulated in PD or other neurodegenerative diseases. It might be helpful to adopt a broader approach, e.g., generating small-RNA libraries using RNA isolated from biofluids of multiple sPD and fPD patients, to identify differentially expressed or discriminatory miRNAs in the two groups.

Conclusion

In conclusion, in this proof-of-concept study we showed that there is great potential in using miRNA signatures along with group sorting algorithms to identify molecular patterns driven by *LRRK2* mutation status. This indicates that LRRK2 impacts the extracellular miRNAome in the periphery and our method could be used for molecular subtyping PD patients using plasma. Our approach needs to be validated in larger cohort sizes and could then be used in studies testing LRRK2 specific therapies.

List of Abbreviations

CSF: Cerebrospinal fluid

Ct_{CSF}: Ct values from CSF dataset

$Ct_{multiplied}$: multiplied Ct values from CSF and plasma dataset

Ct_{plasma}: Ct values from plasma dataset

Fc: Fold change

HC: Healthy control

LASSO: least absolute shrinkage and selection operator

LEDD: levodopa equivalent daily dose

Log2fc: log_2(Fold change)

LOOCV: Leave-One-Out Cross-Validation

LRRK2: Leucine-Rich Repeat Kinase 2

$LRRK2_{MC}$: *LRRK2* mutation carrier

miRNA: microRNA

MoCA: Montreal Cognitive Assessment

PCA: Principal component analysis

PD: Parkinson's disease

qPCR: real-time quantitative PCR

RF: Random Forest

RT: Reverse transcription

sPD: sporadic PD

UPDRS: Unified Parkinson Disease Rating Scale

Declarations

Ethics approval and consent to participate

The study was approved by the Hertie Institute for Clinical Brain Research Biobank and the ethics committee of the medical faculty of the University of Tübingen and University Clinic Tübingen (project ID: 199/2011B01). All participants gave written, informed consent.

Consent for publication

Not applicable

Availability of data and materials

The generated and analyzed data are available from the corresponding author upon reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

The project was supported and funded by the Johannes-Dichgans-scholarship, the Deutsche Forschungsgemeinschaft (Projectnumber D27.15350), the Hertie Stiftung and the Eva Theers Stiftung.

Authors' Contributions

FK conceptualized the study. FK and TG supervised the study. LJB performed the experiments and the statistical analyses. LJB generated all figures. LJB and FK drafted

a first version of the manuscript. All authors read, edited and approved the final manuscript.

Acknowledgements

Samples were obtained from the Neuro-Biobank of the University of Tuebingen, Germany (https://www.hih-tuebingen.de/en/about-us/core-facilities/biobank/). This biobank is supported by the local University, the Hertie Institute and the DZNE.

References

[1] O.-B. Tysnes and A. Storstein, "Epidemiology of Parkinson's disease," *J. Neural Transm.*, vol. 124, no. 8, pp. 901–905, Aug. 2017, doi: 10.1007/s00702-017-1686-y.

[2] L. V Kalia and A. E. Lang, "Parkinson's disease.," *Lancet (London, England)*, vol. 386, no. 9996, pp. 896–912, Aug. 2015, doi: 10.1016/S0140-6736(14)61393-3.

[3] B. Thomas and M. F. Beal, "Parkinson's disease," *Hum. Mol. Genet.*, vol. 16, no. R2, pp. R183–R194, Jul. 2007, doi: 10.1093/hmg/ddm159.

[4] T. Pringsheim, N. Jette, A. Frolkis, and T. D. L. Steeves, "The prevalence of Parkinson's disease: A systematic review and meta-analysis," *Mov. Disord.*, vol. 29, no. 13, pp. 1583–1590, Nov. 2014, doi: 10.1002/mds.25945.

[5] A. Zimprich *et al.*, "Mutations in LRRK2 Cause Autosomal-Dominant Parkinsonism with Pleomorphic Pathology," *Neuron*, vol. 44, no. 4, pp. 601–607, Nov. 2004, doi: 10.1016/j.neuron.2004.11.005.

[6] C. Paisán-Ruíz *et al.*, "Cloning of the Gene Containing Mutations that Cause PARK8-Linked Parkinson's Disease," *Neuron*, vol. 44, no. 4, pp. 595–600, Nov. 2004, doi: 10.1016/j.neuron.2004.10.023.

[7] L. Correia Guedes, J. J. Ferreira, M. M. Rosa, M. Coelho, V. Bonifati, and C. Sampaio, "Worldwide frequency of G2019S LRRK2 mutation in Parkinson's disease: A systematic review," *Parkinsonism Relat. Disord.*, vol. 16, no. 4, pp. 237–242, May 2010, doi: 10.1016/j.parkreldis.2009.11.004.

[8] M. Uhlén *et al.*, "Tissue-based map of the human proteome," *Science (80-.).*, vol. 347, no. 6220, Jan. 2015, doi: 10.1126/science.1260419.

[9] "The Human Protein Atlas." https://www.proteinatlas.org/ (accessed Sep. 12, 2023).

[10] C. J. Gloeckner *et al.*, "The Parkinson disease causing LRRK2 mutation I2020T is associated with increased kinase activity," *Hum. Mol. Genet.*, vol. 15, no. 2, pp. 223–232, Jan. 2006, doi: 10.1093/hmg/ddi439.

[11] T. Hatano *et al.*, "Leucine-rich repeat kinase 2 associates with lipid rafts," *Hum. Mol. Genet.*, vol. 16, no. 6, pp. 678–690, Mar. 2007, doi: 10.1093/hmg/ddm013.

[12] P. N. Gandhi, X. Wang, X. Zhu, S. G. Chen, and A. L. Wilson-Delfosse, "The Roc domain of leucine-rich repeat kinase 2 is sufficient for interaction with microtubules," *J. Neurosci. Res.*, vol. 86, no. 8, pp. 1711–1720, Jun. 2008, doi: 10.1002/jnr.21622.

[13] G. R. Jeong and B. D. Lee, "Pathological Functions of LRRK2 in Parkinson's Disease," *Cells*, vol. 9, no. 12, p. 2565, Nov. 2020, doi: 10.3390/cells9122565.

[14] M. Jaleel *et al.*, "LRRK2 phosphorylates moesin at threonine-558: characterization of how Parkinson's disease mutants affect kinase activity," *Biochem. J.*, vol. 405, no. 2, pp. 307–317, Jul. 2007, doi: 10.1042/BJ20070209.

[15] J. Simón-Sánchez *et al.*, "Genome-wide association study reveals genetic risk underlying Parkinson's disease," *Nat. Genet.*, vol. 41, no. 12, pp. 1308–1312, Dec. 2009, doi: 10.1038/ng.487.

[16] C. Schiesling, N. Kieper, K. Seidel, and R. Krüger, "Review: Familial Parkinson's disease – genetics, clinical phenotype and neuropathology in relation to the common sporadic form of the disease," *Neuropathol. Appl. Neurobiol.*, vol. 34, no. 3, pp. 255–271, Jun. 2008, doi: 10.1111/j.1365-2990.2008.00952.x.

[17] S. Azeggagh and D. C. Berwick, "The development of inhibitors of leucine-rich repeat kinase 2 (LRRK2) as a therapeutic strategy for Parkinson's disease: the current state of play," *Br. J. Pharmacol.*, vol. 179, no. 8, pp. 1478–1495, Apr. 2022, doi: 10.1111/bph.15575.

[18] D. Jennings *et al.*, "Preclinical and clinical evaluation of the LRRK2 inhibitor DNL201 for Parkinson's disease," *Sci. Transl. Med.*, vol. 14, no. 648, Jun. 2022, doi: 10.1126/scitranslmed.abj2658.

[19] R. B. Postuma *et al.*, "MDS clinical diagnostic criteria for Parkinson's disease," *Mov. Disord.*, vol. 30, no. 12, pp. 1591–1601, Oct. 2015, doi: 10.1002/mds.26424.

[20] D. P. Bartel, "MicroRNAs: genomics, biogenesis, mechanism, and function," *Cell*, vol. 116, no. 2, pp. 281–297, Jan. 2004, doi: 10.1016/S0092-8674(04)00045-5.

[21] K. Burgos *et al.*, "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*, vol. 9, no. 5, p. e94839, May 2014, doi: 10.1371/journal.pone.0094839.

[22] Y. Gui, H. Liu, L. Zhang, W. Lv, and X. Hu, "Altered microRNA profiles in cerebrospinal fluid exosome in Parkinson disease and Alzheimer disease," *Oncotarget*, vol. 6, no. 35, pp. 37043–37053, 2015, doi: 10.18632/oncotarget.6158.

[23] A. Singh and D. Sen, "MicroRNAs in Parkinson's disease," *Exp. brain Res.*, vol. 235, no. 8, pp. 2359–2374, 2017, doi: 10.1007/s00221-017-4989-1.

[24] Y. H. Nies, N. H. Mohamad Najib, W. L. Lim, M. A. Kamaruzzaman, M. F. Yahaya, and S. L. Teoh, "MicroRNA Dysregulation in Parkinson's Disease: A Narrative Review," *Front. Neurosci.*, vol. 15, p. 660379, 2021, doi: 10.3389/fnins.2021.660379.

[25] S. Gehrke, Y. Imai, N. Sokol, and B. Lu, "Pathogenic LRRK2 negatively regulates microRNA-mediated translational repression," *Nature*, vol. 466, no. 7306, pp. 637–641, Jul. 2010, doi: 10.1038/nature09191.

[26] C. Ramakers, J. M. Ruijter, R. H. L. Deprez, and A. F. . Moorman, "Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data," *Neurosci. Lett.*, vol. 339, no. 1, pp. 62–66, Mar. 2003, doi: 10.1016/S0304-3940(02)01423-4.

[27] R Core Team, "A language and environment for statistical computing." R Foundation for Statistical Computing, Vienna, Austria, 2022.

[28] K. J. Livak and T. D. Schmittgen, "Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2−ΔΔCT Method," *Methods*, vol. 25, no. 4, pp. 402–408, Dec. 2001, doi: 10.1006/meth.2001.1262.

[29] R. Kolde, “pheatmap: Pretty Heatmaps.” 2019.

[30] M. Kuhn, “Building Predictive Models in R Using the caret Package,” *J. Stat. Softw.*, vol. 28, no. 5, 2008, doi: 10.18637/jss.v028.i05.

[31] R. Kohavi, “A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection,” vol. 14, 2001.

[32] A. Liaw and M. Wiener, “Classification and Regression by randomForest,” *R News*, vol. 2, no. 3, pp. 18–22, 2002.

[33] R. Wang, Y. Yang, H. Wang, Y. He, and C. Li, “MiR-29c protects against inflammation and apoptosis in Parkinson’s disease model in vivo and in vitro by targeting SP1,” *Clin. Exp. Pharmacol. Physiol.*, vol. 47, no. 3, pp. 372–382, Mar. 2020, doi: 10.1111/1440-1681.13212.

[34] T. Botta-Orfila *et al.*, “Identification of blood serum micro-RNAs associated with idiopathic and LRRK2 Parkinson’s disease,” *J. Neurosci. Res.*, vol. 92, no. 8, pp. 1071–1077, 2014, doi: 10.1002/jnr.23377.

[35] X. Bai *et al.*, “Downregulation of blood serum microRNA 29 family in patients with Parkinson’s disease,” *Sci. Rep.*, vol. 7, no. 1, p. 5411, Jul. 2017, doi: 10.1038/s41598-017-03887-3.

[36] B. Ozdilek and B. Demircan, “Serum microRNA expression levels in Turkish patients with Parkinson’s disease,” *Int. J. Neurosci.*, vol. 131, no. 12, pp. 1181–1189, Dec. 2021, doi: 10.1080/00207454.2020.1784165.

[37] C. E. Briggs, Y. Wang, B. Kong, T.-U. W. Woo, L. K. Iyer, and K. C. Sonntag, “Midbrain dopamine neurons in Parkinson′s disease exhibit a dysregulated miRNA and target-gene network,” *Brain Res.*, vol. 1618, pp. 111–121, Aug. 2015, doi: 10.1016/j.brainres.2015.05.021.

[38] R. Mancuso, S. Agostini, A. Hernis, M. Zanzottera, A. Bianchi, and M. Clerici, “Circulatory miR-223-3p Discriminates Between Parkinson’s and Alzheimer’s Patients,” *Sci. Rep.*, vol. 9, no. 1, p. 9393, Jun. 2019, doi: 10.1038/s41598-019-45687-x.

[39] F. Bauernfeind, A. Rieger, F. A. Schildberg, P. A. Knolle, J. L. Schmid-Burgk, and V. Hornung, “NLRP3 Inflammasome Activity Is Negatively Controlled by miR-223,” *J. Immunol.*, vol. 189, no. 8, pp. 4175–4181, Oct. 2012, doi: 10.4049/jimmunol.1201516.

[40] M. S. Moehle *et al.*, “LRRK2 Inhibition Attenuates Microglial Inflammatory Responses,” *J. Neurosci.*, vol. 32, no. 5, pp. 1602–1611, Feb. 2012, doi: 10.1523/JNEUROSCI.5601-11.2012.

[41] G. Zhang *et al.*, “HIF-1α/microRNA-128-3p axis protects hippocampal neurons from apoptosis via the *Axin1* -mediated Wnt/β-catenin signaling pathway in Parkinson’s disease models,” *Aging (Albany. NY).*, vol. 12, no. 5, pp. 4067–4081, Mar. 2020, doi: 10.18632/aging.102636.

[42] L. Zhou *et al.*, “MicroRNA-128 Protects Dopamine Neurons from Apoptosis and Upregulates the Expression of Excitatory Amino Acid Transporter 4 in Parkinson’s Disease by Binding to AXIN1,” *Cell. Physiol. Biochem.*, vol. 51, no. 5, pp. 2275–2289, 2018, doi: 10.1159/000495872.

[43] W.-X. Wang, Q. Huang, Y. Hu, A. J. Stromberg, and P. T. Nelson, “Patterns of

microRNA expression in normal and early Alzheimer's disease human temporal cortex: white matter versus gray matter," *Acta Neuropathol.*, vol. 121, no. 2, pp. 193–205, Feb. 2011, doi: 10.1007/s00401-010-0756-0.

[44] R. R. Thomas, P. M. Keeney, and J. P. Bennett, "Impaired Complex-I Mitochondrial Biogenesis in Parkinson Disease Frontal Cortex," *J. Parkinsons. Dis.*, vol. 2, no. 1, pp. 67–76, 2012, doi: 10.3233/JPD-2012-11074.

[45] T. Kanao, K. Venderova, D. S. Park, T. Unterman, B. Lu, and Y. Imai, "Activation of FoxO by LRRK2 induces expression of proapoptotic proteins and alters survival of postmitotic dopaminergic neuron in Drosophila," *Hum. Mol. Genet.*, vol. 19, no. 19, pp. 3747–3758, Oct. 2010, doi: 10.1093/hmg/ddq289.

Figures

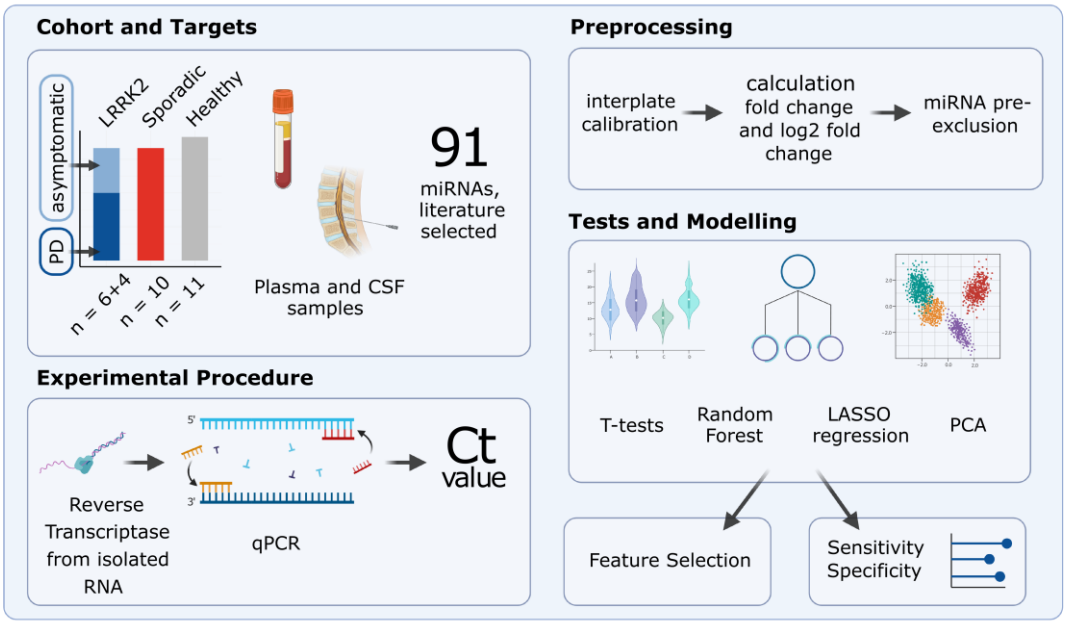

Figure 1: Study Design Overview. The experimental cohort compromised 31 patients categorized into three groups: individuals carrying a *LRRK2* mutation, sporadic PD patients and healthy controls. A total of 91 miRNAs, selected by reviewing relevant literature, were quantified in CSF and plasma samples using reverse transcription and qPCR. Subsequently, the obtained raw Ct values were calibrated and fc and log2fc values were calculated. miRNAs that were expressed only in subgroups were excluded. Group differences were then analyzed using t-tests, PCA, Random Forest and LASSO regression. Created with BioRender.com.

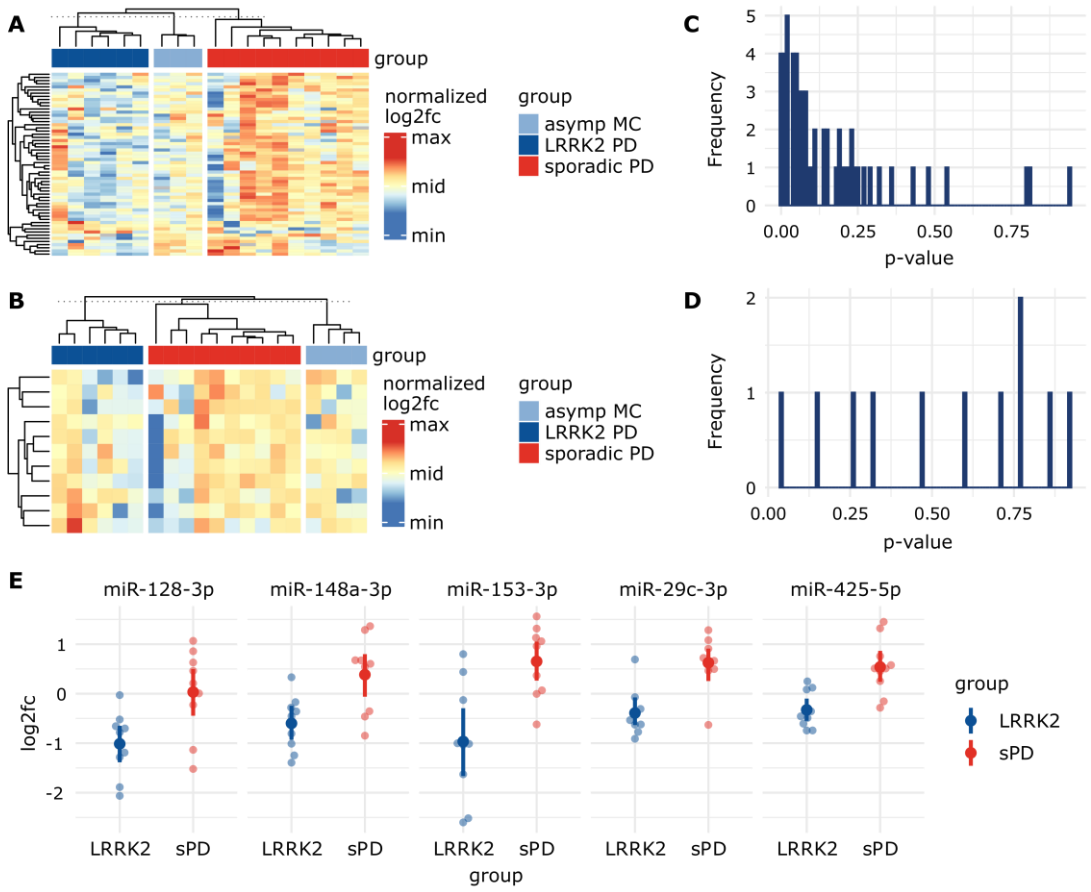

Figure 2: Visualizing miRNA signatures and group differences. A: Heatmap based on log2fc values from the plasma dataset, normalized per row. Rows represent different miRNAs while columns represent patients. Grey cells represent missing values. Column sorting function was inactivated to not interfere with group arrangement. The similarity indicated by the column dendrograms is not proportional between groups. **B:** Heatmap based on log2fc values from CSF dataset, normalized per row. **C:** Histogram showing the distribution of p-values obtained from the plasma and the **D:** CSF datasets after performing unpaired t-tests using log2fc values. **E:** Scatterplots display log2fc values from sPD patients and $LRRK2_{MC}$. Five miRNAs with the lowest p-value in the t-tests were selected. Dark blue or red circle indicate the mean, vertical lines indicate standard deviation.

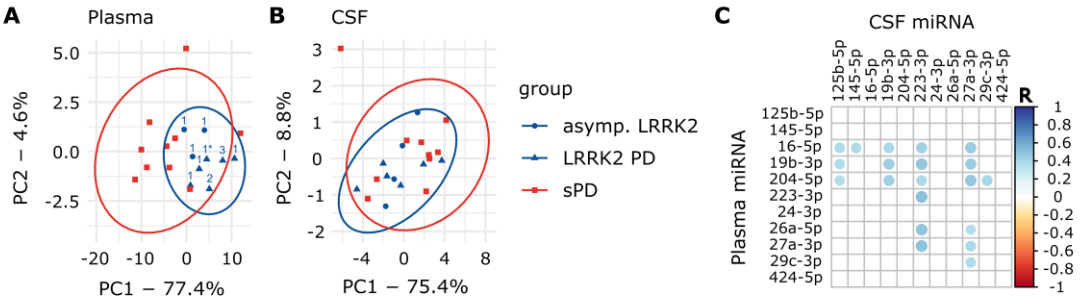

Figure 3: PCA graphs and correlation of plasma and CSF data. A: PCA graph from plasma and **B:** CSF Ct values. Ellipses indicate 95% confidence interval. Numbers indicate *LRRK2* mutation (1: G2019S, 1*: G2019S + G1819, 2: R1441C, 3: I2020T). In the plasma PCA graph, group ellipses overlap while the blue $LRRK2_{MC}$ ellipse trends to only include individuals of the $LRRK2_{MC}$ group. Based on PCA, the $LRRK2_{MC}$ group could be interpreted as a subpopulation of all PD patients. **C:** Correlation matrix presenting statistically significant ($p < 0.05$) correlations between Ct_{plasma} and Ct_{CSF} values of the eleven miRNAs included in the CSF dataset.

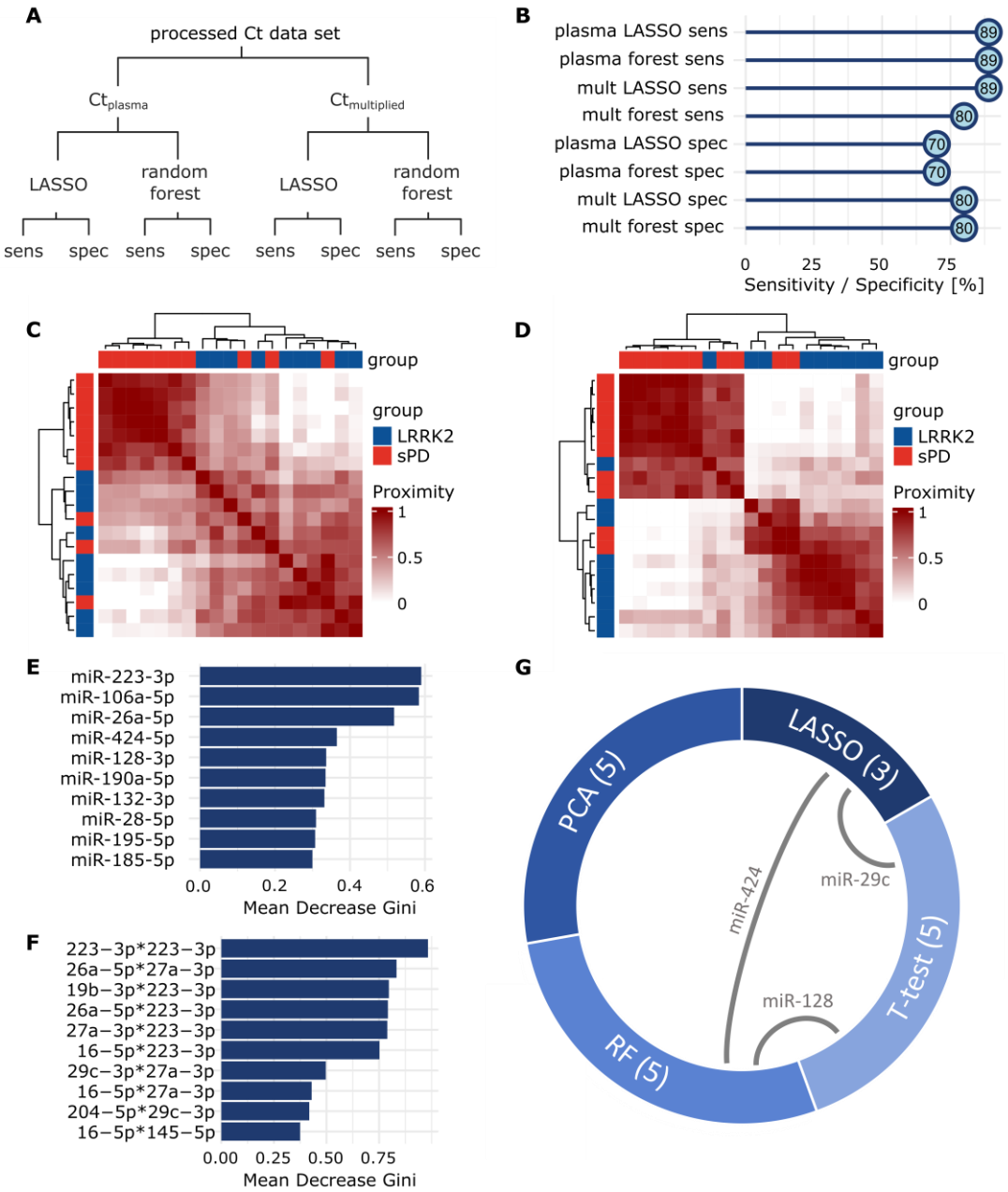

Figure 4: Group prediction with LASSO and Random Forest. A: Overview of performed analyses and respective read-outs. **B:** Sensitivity and specificity values acquired by LASSO regression and Random Forest, using Ct_{plasma} (Plasma) and $Ct_{multiplied}$ (mult). **C:** Heatmaps display the proximity scores calculated in RF for all sPD patients (red) and $LRRK2_{MC}$ (blue) using Ct_{plasma} or **D:** $Ct_{multiplied}$. Groups clearly separate along mutation status. **E:** Mean Decrease Gini scores were calculated after building the respective RF models for both the Ct_{plasma} and **F:** the $Ct_{multiplied}$ data set. Values were sorted in descending order and the top ten miRNAs are displayed. High

decrease of the Gini coefficient translates to a high impact on the performance of the respective RF model. **G:** From each analysis, most influential or discriminatory miRNAs were extracted. Chord diagram displays the overlap of miRNAs selected from the different analyses. Numbers in brackets display set size and grey connecting lines indicate miRNA overlaps.