AWARD NUMBER: W81XWH-20-1-0849

TITLE: Plasma Cell-Free RNA as Noninvasive Biomarker for Parkinson's Disease

PRINCIPAL INVESTIGATOR: Laura Ibanez

CONTRACTING ORGANIZATION: Washington University, Saint Louis, MO

REPORT DATE: January 2024

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Development Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
January 2024	Final	15Sep2020-14Sep2023

4. TITLE AND SUBTITLE
Plasma Cell-Free RNA as Noninvasive Biomarker for Parkinson's Disease

5a. CONTRACT NUMBER
W81XWH-20-1-0849

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Laura Ibanez

E-Mail:ibanezl@wustl.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Washington University School of Medicine
660 S. Euclid Ave. St. Louis
63110, Missouri - US

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Development Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Parkinson disease (PD) is the most common neurodegenerative disorder, after Alzheimer disease (AD). Many attempts have been made to find a good biomarker, including alpha-synuclein protein levels in the cerebrospinal fluid (CSF). Cell-free nucleic acids-based diagnostic tests have revolutionized prenatal screening. They have also been investigated in cancer and fetal development among other traits, including neurodegenerative diseases. We have successfully developed a preliminary predictive model for AD using cell-free plasma RNA sequencing (cfRNASeq) and machine learning techniques. We used an exploratory dataset (10 AD cases and 10 controls) to train a predictive model. We obtained an area under the ROC (AUC) of 0.84 in an independent replication dataset (10 independent AD cases and 10 controls). Moreover, this model provided similar accuracy (AUC=0.86) when tested in four preclinical AD. Using state-of-art deep neural network approaches, the accuracy increased up to 0.94. Overall, these results indicate that we can identify individuals that will progress to dementia. We think this technique can be applied to PD to generate disease-specific predictive model.

We ***hypothesize*** that there are detectable changes in the plasma free nucleic acid composition due to PD pathogenesis, even in early stages. We will use bioinformatics tools to construct a predictive model for PD, leveraging longitudinal plasma data that will allow the modeling of plasma cfRNA composition changes over the curse of the disease, thus maximizing the power of selecting informative transcripts to construct the predictive model.

We will firstly **accurately predict preclinical PD using cell-free nucleic species** in three steps: *A. Create prediction models for PD using cfRNA* generating cfRNASeq data from 200 plasma samples (50 PD individuals at 3 time-points – early pre-clinical (5-10 years before symptoms), pre-clinical (2-5 years before symptoms) and symptomatic (5-8 years after diagnostic) and 50 controls). We will use multiple analytical approaches including digital deconvolution and machine learning, feature selection and deep neural networks (similar to what we have used to generate our model for AD) to build a robust predictive model that includes the optimal number of transcripts. *B. Replication:* We will quantify the transcripts selected in A to be part of the predictive model using a more scalable and cost-effective technology such as Nanostring, Sequenom or custom transcript array to replicate the predictive model in an independent dataset (50 preclinical PD cases and 50 controls). We will also include subjects from African American and Latin ancestries (20 cases and 20 controls from each) and carriers of PD-causing mutations in *PARK1* and *PARK2* (n=20) to test the performance of the model in non-European ethnicities and in mutation carriers. *C. Specificity:* We will quantify the transcripts selected in A and replicated in B using the same scalable and cost-effective technology in 80 cases of other neurodegenerative diseases (AD, Lewy body dementia, progressive supranuclear palsy, amyotrophic lateral sclerosis and frontotemporal dementia) and additional 40 controls to test whether the predictive model is specific for PD or neurodegeneration. We expect the predictive model to be specific for PD; however, some overlap is expected due the commonalities of neurodegenerative diseases. We are currently generating longitudinal cfRNASeq data on AD individuals. With the data generated in the first aim, we will perform **integrative analyses of AD and PD** to describe biological differences and commonalities across the two most common neurodegenerative diseases. This will allow the description of biological mechanisms such as differences in genes or pathways across diseases, differences in the timeline of the disease for common genes/pathways and the improvement of the differential diagnosis.

If successful, this method could improve the cost-effectiveness of the currently available tools to diagnose and monitor PD, and provide a scalable blood-based early diagnostic screening tool.

Dr. Ibanez research interest is focused on using genetics to improve the management of individuals that suffer from neurodegenerative diseases, specially the early management by using high-throughput technologies and bioinformatics. Currently there is a great potential on using multi-omic approaches that integrate all levels of biological information. By using powerful bioinformatic tools we can combine genetic variance data with the RNA translation and the final protein production. This can be linked to diseases and be used to create predictive models that can easily be used in clinical settings to improve the management and quality of life of patients. In the last year and a half, Dr. Ibanez efforts have been focused on creating predictive models for Alzheimer Disease using different approaches. Even this model is on preliminary stages, she has proven that she can successfully create predictive models using cell-free RNA. This proposal will allow her to leverage the biology from the data generated, which could add some biological understanding to the biology of these neurodegenerative diseases.

This proposal has the potential of leading to a biomarker for early diagnosis and prognosis of PD. Moreover, the transcripts included in the prediction model will probably have biological relevance which, together with the integrative analyses with AD, might lead to potential drug targets in future studies.

15. SUBJECT TERMS

None listed.

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
					USAMRDC
a. REPORT	b. ABSTRACT	c. THIS PAGE	Unclassified	9	19b. TELEPHONE NUMBER *(include area code)*
Unclassified	Unclassified	Unclassified			

TABLE OF CONTENTS

Introduction

Parkinson's disease (PD) is the most common neurodegenerative disorder, after Alzheimer disease (AD). Many attempts have been made to find a good biomarker, including alpha-synuclein protein levels in the cerebrospinal fluid (CSF). Cell-free nucleic acids-based diagnostic tests have revolutionized prenatal screening. They have also been investigated in cancer and fetal development among other traits, including neurodegenerative diseases. We will use bioinformatics tools to construct a predictive model for PD, leveraging longitudinal plasma data that will allow the modeling of plasma cfRNA composition changes over the curse of the disease, thus maximizing the power of selecting informative transcripts to construct the predictive model.

Keywords

Parkinson's Disease, Biomarkers, cell-free RNA, Machine Learning

Accomplishments

What are the major goals of the project?

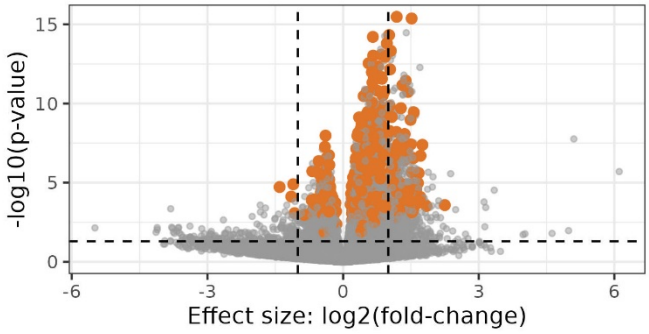
This project has four major goals:

1. Create a Predictive Model for Parkinson's Disease.
2. Replicate the Predictive Model in an independent Dataset.
3. Test the Specificity & Sensitivity of the Predictive Model.

What was accomplished under these goals?

1) Major activities: Overall, during the duration of this award, we have extracted and sequenced RNA from 189 plasma samples from Washington University in Saint Louis (WUSTL), successfully performed differential expression analyses, and generated preliminary predictive models. Then we have accessed 516 independent plasma samples from an international collaborator (Barcelona), that have been processed to replicate the results, with 206 of them passing QC parameters and being included in the replication. The two populations, WUSTL and Barcelona had key differences regarding clinical data availability. The main difference was the lack of medication records for the Barcelona population. To account for this difference, we removed from analyses transcripts that associated with medication usage, selected as those differentially epressed between treated and untreated PD participants in the WUSTL population ($p<0.05$). After stringent quality control, we found 600 transcripts differentially accumulated in plasma of PD participants compared to healthy controls after correcting by se and age when meta-analyzing both populations (Figure 1). Then, we leveraged the 14 transcripts that were significant in both populations (Table 1) to build a predictive model to aid in the diagnosis of PD. We observed an ROC curve in the replication dataset (Barcelona) of 0.78, with high specificity for PD (AUC for Alzheimer disease 0.56).

Figure 1. Volcano plot showing the results from the discovery phase (WUSTL). We identified 600 transcripts differentially accumulated in plasma of PD participants compared to healthy controls after adjusting by sex and age.



2) Specific objectives: The goal for the final year of the award were to replicate the results in an independent dataset (both differential expression and predictive model) and prepare the publication. To date, we are finalizing the analysis and preparing the manuscript.

3) Significant results or key outcomes:

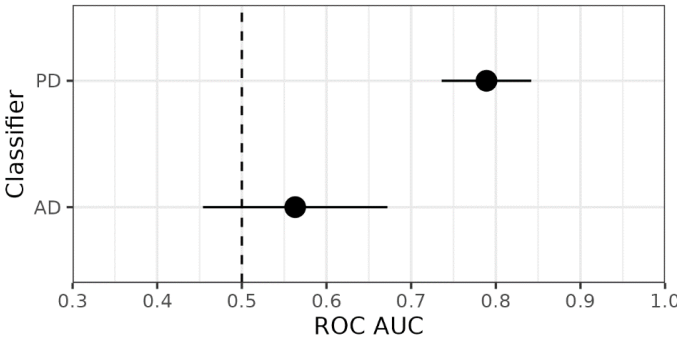
Differential Expression Analyses: Given the absence of medication record for the replication dataset, we had to approach this analysis differently that we did in the last progress report. After stringent quality control (as described in previous reports), we identified those genes associated with the presence of medication in the WUSTL dataset, and remove them from the analyses. Then we re-analyzed both datasets adjusting by age and se. We have identified 698 down-regulated transcripts and 1,253 up-regulated transcripts that pass multiple test correction when comparing all PD cases to controls in the discovery dataset. Of those, 14 were also significant in the replication dataset (Table 1), and 600 transcripts replicated in the meta-analyses between the WUSTL and the dataset from Spain. Pathway analyses are undergoing.

Table 1. Top differentially abundant transcripts in plasma when comparing PD cases to Controls

Gene	Discovery		Replication		Meta-Analysisi	
	Fold Change	P-Value	Fold Change	P-Value	P-Value	Adjusted P-Value
ENSG00000006744.19	-0.327	0.007	-1.17	0.031	0.002	0.024
ENSG00000053371.12	0.710	1.65×10^{-10}	0.921	0.043	1.87×10^{-10}	4.40×10^{-08}
ENSG00000058453.17	-0.402	0.019	-1.460	0.029	0.004	0.044
ENSG00000059691.12	-0.467	0.017	-1.836	0.021	0.003	0.033
ENSG00000099954.18	0.806	0.014	2.771	0.049	0.006	0.051
ENSG00000121058.5	-0.411	0.013	-1.455	0.039	0.004	0.041
ENSG00000122965.11	-0.373	0.004	-1.545	0.014	6.87×10^{-04}	0.010
ENSG00000131943.17	-0.299	0.018	-1.085	0.032	0.005	0.045
ENSG00000155438.12	-0.329	0.005	-1.444	0.024	0.001	0.017
ENSG00000168010.11	-0.418	0.004	-1.075	0.019	8.66×10^{-04}	0.011
ENSG00000188215.10	-0.322	0.022	-1.917	0.010	0.002	0.024
ENSG00000197858.11	0.307	0.003	1.240	0.027	0.001	0.014
ENSG00000198799.12	0.400	0.012	2.040	0.030	0.003	0.033
ENSG00000204231.10	-0.286	0.006	-1.042	0.035	0.002	0.023
ENSG00000204252.14	-0.426	6.58×10^{-04}	-1.174	0.020	1.65×10^{-04}	0.003

Creation of the Predictive Model: After stringent quality control (described in the prior progress report), we have used the normalized counts for the 14 differentially expressed transcripts and Ridge regression to build a predictive model. After normalization and scaling the two independent datasets using Z-Scores, we used the discovery dataset (WUSTL) to train the model, or in other words, calculate the weights of each transcript, and then the replication (Barcelona) to test the model, or evaluate performance. The area under the ROC curve (AUC) values for the Barcelona dataset was 0.789 (0.736-0.842 - Figure 2). We also had a dataset with Alzheimer's Disease (AD) plasma cfRNA samples, thus we leveraged to evaluate if the model was specific to PD. After normalization, we obtained an AUC of 0.563 (0.454-0.672 – Figure 2) suggesting that the model is specific to PD since it predicts AD randomly (close to 0.50).

Figure 2. Wisker plot with the Area Under the Curve (AUC) values for the predictive model in the replication dataset (Barcelona). The weight of the model were calculated using the discovery dataset (WUSTL). The bottom wisker represents the AUC for Alzheimer's Disease.



4) Other achievements. Include a discussion of stated goals not met.

Other Achievements: Via our collaboration, we have not only obtained plasma samples from PD individuals, but also a significant number of samples from individuals with Parkinson's Disease Dementia, and Dementia with Lewy Bodies to develop more specific predictive models. This is the main aim of the R00 obtained by Dr. Ibanez using the results from this career development award as preliminary data. We are working on these analyses in parallel with the preparation of this

publication. Additionally, the analyses performed in AD are finalized and the paper has been accepted for publication in iScience from the Cell group.

What opportunities for training and professional development has the project provided?

Over the last three years Dr. Ibanez has successfully taken advantage of this developmental grant to successfully:

- secure a tenure track position at Washington University in Saint Louis.
- expand her research to other RNA types (circular,non-coding, and small RNAs).
- apply to three R01, all of them discussed but below the pay line.
- expand the size of her lab to currently be supporting 15 individuals.
- appointed the Dominantly Inherited Alzheimer's Disease Network (DIAN) Fluid Biomarker Core Leader for both, the Observational study and the Clinical Trials Unit.
- increase her national and international recognition with more than ten invited seminars and three oral communications in international meetings in the last three years.
- secure funding from private foundations to support her team financially and scientifically.
- publish the use of cell-free RNA to predict Alzheimer's Disease that includes a subset of the data generated with the present developmental award as sensitivity analyses. *Cell-free RNA signatures predict Alzheimer's disease. Cisterna-Garcia A. iScience Dec. 2023*

How were the results disseminated to communities of interest?

The results described in this report were reported in poster format in the 2023 AD/PD meeting (Gothenburg – March 28th to April 1st) and the 2023 AAIC meeting (Amsterdam – July 16th to 20th). We do not plan to submit an abstract to the 2023 AAIC meeting (Washington DC – July 28th – August 1st) to prioritize the publication of this results.

What do you plan to do during the next reporting period to accomplish the goals?

This is the final report – We are currently finalizing analysis and planning for a publication

Impact

What was the impact of the development of the principal discipline of the project?

Nothing to Report

What was the impact on other disciplines?

Nothing to Report

What was the impact on technology transfer?

We are in discussions with the Office of Transfer Materials to disclose this model as an invention, or keep it under the already protected methods developed in the AD cell-free RNA research.

What was the impact on society beyond science and technology?

Nothing to Report

Challenges / Problems

Changes in approach and reasons for change?

As stated in previous progress reports, reviewers and experts suggested to find an independent dataset to replicate our predictive models instead of finding an alternative technology for clinical translation. We successfully secured a collaboration, obtained the samples, processed them, and generated the data. The new plasma samples (from our collaborator in Spain) led to unexpectedly low-quality data. However, we have performed quality control and are finalizing the analyses to publish a manuscript.

Actual or anticipated problems or delays and actions or plans to resolve them

Nothing to Report

Changes that had a significant impact on expenditures

Nothing to Report

Significant changes in use or care of human subjects, vertebrate animals, biohazards, and/or select agents

Nothing to Report

Significant changes in use or care of human subjects

Nothing to Report

Significant changes in use or care of vertebrate animals.

Nothing to Report

Significant changes in use of biohazards and/or select agents

Nothing to Report

Products

Nothing to Report

Participants and other collaborating organizations

What individuals have worked on the project?

Name:	*Laura Ibanez*
Project Role:	*Principal Investigator*
Researcher Identifier (e.g. ORCID ID):	0000-0003-2381-7059
Nearest person month worked:	1.1
Contribution to Project:	*Dr. Ibanez has designed and supervised all the steps for the correct development of the project. She performed half of the ribodepletion protocols, generated all the libraries and performed the quality control. She planned for the pooling and was in contact with the*

	sequencing center during the sequencing project. She has performed the reported analyses and is preparing the manuscript.
Funding Support:	*National Institute of Aging* *Michael J. Fox Foundation* *Bright Focus Foundation* *Alzheimer's Association*

Has there been a change in the active other support of the PD/PI(s) or senior/key personnel since the last reporting period?

- Not from last report.

What other organizations were involved as partners?

- Not from last report.

Special Reporting Requirements

N/A

Appendices